HAL
archives-ouvertes.fr

Is It Harmful? Re-examining Privacy Concerns

Agnieszka Kitkowska, Erik Wästlund, Joachim Meyer, Leonardo Martucci

▶ **To cite this version:**

Agnieszka Kitkowska, Erik Wästlund, Joachim Meyer, Leonardo Martucci. Is It Harmful? Re-examining Privacy Concerns. Marit Hansen; Eleni Kosta; Igor Nai-Fovino; Simone Fischer-Hübner. Privacy and Identity Management. The Smart Revolution : 12th IFIP WG 9.2, 9.5, 9.6/11.7, 11.6/SIG 9.2.2 International Summer School, Ispra, Italy, September 4-8, 2017, Revised Selected Papers, AICT-526, Springer International Publishing, pp.59-75, 2018, IFIP Advances in Information and Communication Technology, 978-3-319-92924-8. 10.1007/978-3-319-92925-5_5 . hal-01883632

HAL Id: hal-01883632
https://hal.inria.fr/hal-01883632

Submitted on 28 Sep 2018

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

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

Distributed under a Creative Commons Attribution| 4.0 International License

Is It Harmful? Re-examining privacy concerns

Agnieszka Kitkowska[1], Erik Wästlund[1], Joachim Meyer[2], and Leonardo A. Martucci[1]

[1] Karlstad University, Karlstad, Sweden.
agnieszka.kitkowska@kau.se
[2] Tel Aviv University, Tel Aviv, Israel.

Abstract. The increased popularity of interconnected devices, which we rely on when performing day-to-day activities expose people to various privacy harms. This paper presents findings from the empirical investigation of privacy concerns. The study revealed that people, regardless of their diversity, perceive privacy harms as generic and simplified models, not individually as suggested in Solove's framework. Additionally, the results identified differences in privacy concerns related to information disclosure, protection behavior, and demographics. The findings may benefit privacy and system designers, ensuring that policies and digital systems match people's privacy expectations, decreasing risks and harms.

Keywords: Privacy, Human Factors, Attitudes, Decision Making.

1 Introduction

The widespread Internet availability and access to various devices, from PCs, through mobile to smart devices, enabled the establishment of an ecosystem of interconnected applications. People adapt these technologies and feed them with a large amount of data. Such applications assist people with performing most of their daily activities, including socializing, healthcare, financial transactions, work and more. People voluntarily, and sometimes unknowingly contribute data to Internet-based applications, and that may expose them to privacy risks, violations, and harms.

Due to the increasing amount of security breaches, digital privacy became a subject of public debate. The news about data leakages and their potential effects frequently appear in media, informing the audience about the potential privacy risks. Since privacy violations are in the center of interest, governments and policymakers introduced legal guidelines and regulations aiming to protect personal data, such as the General Data Protection Regulation (GDPR) in Europe[42] or FTC requirements in USA [49]. Simultaneously, the academic research resulted in multiple studies about online privacy, demonstrating that people are concerned about their data, nevertheless, they trade them for potential benefits arising from applications [4, 55, 16]. Despite the efforts of researchers and policymakers, as well as increased privacy awareness raised in media, people's attitudes and behaviors remain unchanged. Regardless of their concerns, people provide personal

information to online companies to use their services, ensure social interactions, improve well-being and more.

The aim of this study is to investigate privacy perceptions and to re-examine some of the privacy behaviors. The primary contribution of this research is a novel instrument to measure privacy attitudes, Privacy Harms Concerns (PHC) scale. Following the recommendation of the past research [28], we used *privacy harms* identified by Daniel Solove as a foundation for the scale's development [48] (Table 1). The goal of this research was to identify how people perceive privacy concerns relevant to harms (to ensure consistency labeled privacy concerns throughout the article). The results confirmed that people, in spite of their diversity, tend to have rather comprehensive and simplified view of privacy concerns, perceiving their severity and importance in a similar manner. Regardless of this general tendencies, we identified differences in privacy perceptions, information disclosure, and protection behaviors. Additionally, the findings demonstrate a potential for demographic differences in privacy concerns. Overall, the results contribute to further understanding of people's privacy attitudes and behaviors.

Table 1. Typology of privacy harms according to the Solove's framework.

Information collection	**Information processing**	**Information dissemination**	**Invasions**
Surveillance	Aggregation	Breach of confidentiality	Intrusion
Interrogation	Identification	Disclosure	Decisional interference
	Insecurity	Exposure	
	Secondary Use	Increased Accessibility	
	Exclusion	Blackmail	
		Appropriation	
		Distortion	

2 Related work

2.1 Privacy attitudes: concerns and harms

According to Westin, privacy concern is the intention to protect personal information from others [10]. Thus it carries a negative weight and should result in preventive or protective actions. As defined by Campbell, the information privacy concern is a subjective notation concentrated around the *input, use and control of data* [32]. Therefore, the information concern is related to the flow of data between the user and involved data processors. The online privacy research recognized various antecedents of privacy concerns such as trust, risk perception, previous privacy experience, privacy awareness, personality traits and demographic differences [31, 46, 8, 52, 21, 26]. Some of this research investigated the influence of concerns on privacy behaviors but the results are inconsistent. Some studies show that despite concerns, people disclose information, however,

it is a natural consequence of being a part of community [31]. On the other hand, there is a large volume of research illustrating, that regardless of privacy concerns, people tend to share their information, and their decisions are based on cost and benefit trade-off [44, 2, 18]. This so-called *privacy paradox* is frequently explained by factors such as information asymmetry [1, 2, 5] or psychological biases and heuristics [9, 20, 29, 12, 25].

To the best of our knowledge, privacy concerns have not been investigated from the perspective of privacy harms. Similarly to the notion of privacy itself, there is no clear definition of privacy harm. However, scholars from legislative sector tried to provide a coherent explanation of the term. For instance, Solove identified a privacy problem as a result of harm, claiming that harms do not have to be physical or emotional, they *can occur by chilling socially beneficial behavior (for example, free speech and association) or by leading to power imbalances that adversely affect social structure (for example, excessive executive power)* [50]. Similarly, Calo defines harm as a conceptualized negative consequence of privacy violation [11]. Nevertheless, the most comprehensive definition of privacy harms we found was provided by researchers investigating smart grids privacy, De & Métayer. They defined harm as *the negative impact on a data subject, or a group of data subjects, or the society as a whole, from the standpoint of physical, mental, or financial well-being or reputation, dignity, freedom, acceptance in society, self-actualization, domestic life, freedom of expression, or any fundamental right, resulting from one or more feared events* [15]. In this research, we follow this definition and consider harms as a multidimensional notion.

The previous research resulted in multiple scales measuring privacy concerns. Such measuring scales are constructed in various ways, for example by asking people directly about their concerns, treating concerns as latent variables or as moderators [38]. For instance, Smith et al. [47] developed Concerns for Information Privacy (CFIP) scale aiming to explore the concerns' structure. The study identified four dimensions of privacy concerns: improper access, unauthorized secondary use, error, and collection. Malhotra et al. developed Internet Users Information Privacy Concern (IUIPC) scale identifying three dimensions: collection, control, and awareness of privacy practices [32]. According to their research consumers perceive as the most important awareness and control over their data stored by online companies. The IUIPC scale can be applied to privacy research in various contexts. Regardless of the coherent nature of this scale, it seems to be an organization- and consumer-oriented, as authors put it, IUIPC is *representation of online consumers' concerns for information privacy.* Buchanan et al. developed another privacy concerns scale, measuring individual privacy issues, asking directly about concerns, for instance regarding personality theft, access to medical records etc[10].

Considering the definitions of privacy harms and the past research, we want to improve understanding of attitudes and re-examine dimensionality of privacy concerns. Hence, the first research question:

RQ1 How do people perceive privacy harms concerns?

RQ1.1 What are the main dimensions of privacy concerns?

RQ1.2 Are some concerns perceived as more severe than others?

2.2 Privacy behaviors

In order to cross-validate findings of privacy concerns, some research examined their relations with other attitudinal or behavioral factors, such as information disclosure or protection behavior.

According to the research, the information disclosure behavior varies, depending on psychological states [40], risk perceptions [14, 56], trust and more. Several studies explored the relationship between privacy concerns and information disclosure. For example, research showed significant effects of privacy concerns on information disclosure, influenced by psychological biases, such as optimism bias, over-disclosure and others [29, 52, 39]. Similarly, the researchers found evidence of irrational behavior, when people tend to disclose data knowingly about the potential risks[17].

In this research, we will not examine factors influencing privacy concerns or the direction of the relationship between attitude and behavior. Instead, we focus on a variance of privacy concerns among people who disclose (or not) sensitive or non-sensitive information. Hence, our next research question:

RQ2 Is there a relationship between privacy concerns and privacy behavior?

RQ2.1 Do privacy concerns vary among people disclosing and not-disclosing non-sensitive information?

RQ2.2 Do privacy concerns vary among people disclosing and not-disclosing sensitive information?

The past privacy research identified control as an important factor influencing privacy behaviors [19, 5]. To achieve control over online information disclosure people apply different protection measures. Some use technical protections, such as anti-malware or anti-virus software, add blockers, or other privacy enhancing technologies. Others may be more careful about their physical privacy (hiding PIN, shredding documentation), limit information provided to social networks (such as reduction of the posts' audience, limited profile visibilities etc.), decrease number of online profiles or even entirely resign from the online presence. The relationship between privacy concerns and protection is unclear. There is some research claiming, that such relationship exists, however, the correlations are low and people less concerned about privacy use more of protective measures [3, 37].

Considering the past research demonstrating that the relationship between concerns and behavior exists, we ask following questions:

RQ2.3 Is there a relationship between people's privacy concerns and general privacy caution?

RQ2.4 Is there a relationship between people's privacy concerns and technical privacy protection?

2.3 Demographics

To assess individual differences in privacy concerns some of the researchers used demographics, such as geographic/cultural differences, age, education or gender [41, 13]. However, the results of studies investigating demographic dependencies are inconclusive. For instance, there are studies claiming that gender impacts privacy perceptions and females are more concerned about their data than males. However, some of these findings show that the impact of gender on privacy attitudes and behaviors is indirect or insignificant [23, 6, 36]. Regarding the age, there seems to be a general tendency that older generations are more concerned about their privacy than the younger ones [51, 34]. Nevertheless, it does not mean that younger people ignore it. In contrary, the research demonstrates that younger people use technical protection measures to better manage their privacy [33].

The previous research associated privacy concerns with geographic/cultural background [54, 7]. The geographic divide was confirmed in the qualitative study of seven European countries, identifying main privacy concerns influencing information disclosure and a variety of privacy fears among different nationalities [34]. Similarly, other studies showed differences among respondents from North America and Europe[45], and France and Hong Kong [22]. Such differences were accredited to cultural dimensions, for instance, assertiveness or gender egalitarianism [41, 53].

Considering the previous research's findings, we aim to examine whether there are any significant demographic differences in privacy concerns and behaviors among the participants of our study. Hence, our last research question:

RQ3 Do privacy concerns differ depending on the demographic background?

3 Method

The online survey was created to answer the research questions. It contained 80 questions, divided into thematic sections, such as participants' demographics, opinions related to data collection and processing, security, identity, and personal questions. To measure the responses, we used mixed design, including questions collecting responses on the scale ranging from 0 to 100 (strongly disagree/strongly agree; never/always) and multiple choice questions.

Before participating in the survey, respondents were presented with informed consent, explaining what type of information will be requested during the survey, what is the study purpose and who should be contacted in case of any questions. Each participant had to agree to the informed consent and confirm that he/she is over 18 years old.

3.1 Instrument

The online survey consisted of three major sections: the new scale to measure privacy concerns, and two scales acquired from the past research, measuring

privacy behaviors. Due to the thematic division of the survey and to ensure the instrument's consistency, some of the questions from the PHC were mixed with questions from the scale measuring protection behavior.

To create the new scale, we applied the privacy harms framework defined by Solove [48]. We developed the 48 items scale derived from Solove's 16 privacy harms. Solove categorized 16 privacy harms into four groups, which are presented in Table 1. Solove's work addresses privacy harms from the legal perspective, however, in the past it was used in the information privacy research [27]. Additionally, we believe that privacy harms may be recognizable and meaningful, since the framework origins from court cases and real-life examples. Originally we aimed to measure each individual privacy harm, hence we used three items for each of them. The instrument collected continuous data, scores ranging from 0 to 100 (strongly disagree/strongly agree). After all data were collected, some of the items were modified, to ensure scores' consistency.

The scale measuring information disclosure was acquired from Joinson et al. [24]. It consisted of 11 items, asking respondents questions of personal nature. To ensure consistency, the information disclosure scale was modified and did not include two questions requiring respondents to type answers in the text boxes. The scale aimed to measure disclosure of sensitive and non-sensitive information. The sensitive items were measured by asking intimate questions, such as 'How many different sexual partners have you had?'. The non-sensitive items contained less invasive questions, for instance 'Are you right or left handed?'. The disclosure level was measured by providing respondents with option 'I prefer not to say', which if chosen was coded as 1 (don't disclose). All other responses were coded 0. In a result participants who do not disclose scored 5 per sensitive and 4 per non-sensitive items. All other participants were treated as disclose group. This resulted in division of respondents to two groups: disclosing sensitive information ($N = 273$) and non-disclosing sensitive information ($N = 109$), and disclosing non-sensitive ($N = 325$) and non-disclosing non-sensitive information ($N = 57$).

The second scale acquired from the previous research aimed to measure protection behavior [10]. It consisted of 12 items, 6 measuring a general privacy caution and 6 measuring technical protection [10]. To ensure consistency we modified the scale, and instead of Likert scale, we applied range scores. In a result, we collected continuous data with scores ranging from 0 to 100 (never/always).

3.2 Data collection

The online survey was distributed on two platforms, Microworkers and CallForParticipants (CFP). Participation in the survey was voluntary. Microworkers' participants received financial compensation \$1-\$1.50 per response, while CFP respondents did not receive any compensation. The total number of participants reached 437 (375 from Microworkers, 62 from CallForParticipants), however, only 382 responses were valid. On Microworkers the response validity was checked automatically. Additionally, all responses were monitored manually, one by one. Furthermore, any surveys completed in less than five minutes or longer than four

hours were removed. Participants had to respond to all questions and in a result, there was no missing data; the survey allowed respondents to backtrack and amend responses. Each respondent could participate in the survey only once.

Furthermore, to decrease the possibility of statistical bias, the data set was scanned for outliers. As recommended in the literature, instead of using a standard method for detecting extreme cases, such as the mean plus/minus two or three standard deviations [30], we applied $3x$ Inter-quartile Range. All responses that contained outliers were removed from the analysis, which left the sample of 382 responses.

To assess the desired demographics, we used a geographic cluster sampling, with cluster sizes aiming to reach 100 respondents each. Choice of geographic areas was based on the results from the Data Protection Eurobarometer [35]. We focused on four geographic areas: UK, USA, Italy and Nordic countries (Sweden, Norway, Finland, Denmark, and Germany). Among the respondents 57.9% ($N = 221$) were males and 42.1% ($N = 161$) females; the average age was 32 years ($Min = 18; Max = 70$). The full demographics beak-down is presented in Table 2.

Table 2. Participants demographics

Demographic	N	**Percent**
Country		
Italy	91	23.8
Nordic countries	76	19.9
UK	113	29.6
USA	102	26.7
Gender		
Male	221	57.9
Female	161	42.1
Education		
High school	70	18.3
Higher education	203	53.1
Still studying	109	28.5
Age		
18-24	98	25.7
25-34	153	40.1
35-44	76	19.9
over 44	55	14.4
Total	382	

4 Results

4.1 Dimensions of privacy concerns

To assess the answer to the RQ1 we commenced with investigating its sub-question: *What are the dimensions of privacy concerns?* (RQ1.1). We created the PHC and used the Exploratory Factor Analysis (EFA) to assess dimensions of privacy concerns.

The EFA was used because it allows to ascertain factors that may explain correlations between variables, but it does not require underlying theoretical structure [43]. The Kaiser-Meyer-Olkin measure (.903) and Bartlett test for sphericity (significant at the level $p < .001$) confirmed EFA's suitability. We used orthogonal rotation, varimax presuming that the correlations between the variables are weak.

To extract factors, we used the principal axis factoring (PAF) allowing to measure the latent structure of variables and their relationships [43]. From the original 48 items 30 items remained, after removing factors with communalities $< .3$, item loadings $< .3$ and factors consisting of less than three loaded items.

After applying the solution and scree plot analysis, we extracted seven factors, identifying people's perceptions of privacy concerns: *unauthorized access*, *misuse of data*, *secondary use of data*, *insecurity*, *exposure*, *interrogation*, *distortion*. When computing the internal consistency for the scale based on the factors, the Cronbach alpha scores for the identified factors were all above .7 (Table 3).

Additionally we computed the means for each dimension of the privacy concerns as demonstrated in Table 4. We used the means in further analysis, to assess the relationship with behavior and investigate demographics.

Table 3. The results of Exploratory Factor Analysis; $N = 382$.

Extracted factors	**Cronbach alpha**
Factor 1: *unauthorized access to data*	.865
Factor 2: *misuse of data*	.836
Factor 3: *secondary use of data*	.811
Factor 4: *insecurity*	.736
Factor 5: *exposure*	.745
Factor 6: *interrogation*	.721
Factor 7: *distortion*	.735

4.2 Information disclosure

To asses the differences in concern between respondents who disclose sensitive/non-sensitive items (RQ2.1 and RQ2.2) we performed the independent-sample t-Test. We checked the outcomes of Levene's test that were significant at level $> .05$, hence we report the results for equal variances not assumed.

Table 4. Means of the privacy concerns dimensions, N=382.

Dimension	M	SD
Insecurity	90.02	10.4
Exposure	77.82	17.7
Unauthorized access	72.75	17.2
Secondary use of data	72.42	20.0
Misuse of data	71.23	16.1
Distortion	63.75	21.5
Interrogation	45.89	21.2

We found a significant difference among respondents that disclose (M=70.5, SD=20.6) and do not disclose (M=77.1, SD=17.5) sensitive information about the *secondary use of data*, $t(380) = -2.9$, $p = .002$; and *interrogation* (M=42.9, SD=21.4; M=53.3, SD=18.8 respectively), $t(380) = -4.4$, $p < .001$.

We identified the same type of concerns among participants disclosing non-sensitive information. The respondents who did not disclose information (M=77.6, SD=19) were significantly more concerned about the *secondary use of data* than those who disclose it (M=71.5, SD=20.1), $t(380) = -2.1$, $p = .029$; the same behavior was observed regarding *interrogation* (M=52.9, SD=21.9; M=44.6, SD=20.9 respectively), $t(380) = -2.7$, $p = .010$.

4.3 Protection behavior

To determine the relationship between privacy concerns and protection behaviors (RQ2.3 and RQ2.4) we performed Pearson Correlation tests, and examined scatter plots for the correlated variables.

Table 5. Correlations between privacy concerns and protection behaviors; $N = 382$.

	General caution	Technical Protection
Unauthorized access	.318**	.290**
Misuse of data	.404**	.357**
Secondary use	.024	.184**
Insecurity	.201**	.346**
Interrogation	-.243**	-.027
Exposure	.215**	.233**
Distortion	.358**	.246**

** Correlation is significant at the .001 level (2-tailed)

We identified significant correlations between general caution and technical protection behavior, and privacy concerns, ranging between $r = .184$ and $r = .404$ (Table 6). The results demonstrate positive correlations for general caution and concerns about *unauthorized access, misuse of data, insecurity, exposure* and *distortions*, and a negative correlation for *interrogation*. Similarly,

positive correlations were found for technical protection behavior and *unauthorized access, misuse of data, secondary use of data, insecurity, exposure* and *distortions*. However, we did not identify a relationship between general caution and *secondary use*, as well as between technical protection and *interrogation*.

4.4 Demographics

We conducted One-Way Analysis of Variance (ANOVA), t-Tests and Chi-Square to analyze whether there are significant differences in privacy concerns among people from various demographics (RQ3).

First, we analyzed responses of participants from different geographic locations (Table 7). There were significant effects for *secondary use of data* ($F(3, 381) = 5.010; p = .002$), *interrogation* ($F(3, 381) = 3.241; p = .022$) and *distortion* ($F(3, 381) = 2.885; p = .036$). The post-hoc Tukey test results confirmed significant differences ($p = .001$) between Italy (M=77.2; SD=19.9) and the UK (M = 77.2; SD = 17.9) regarding the *secondary use of data*. Similarly, there was a significant difference ($p = .038$) between Italy (M =40.3; SD = 20.5) and the Nordic Countries (M = 49.8; SD = 19.7), and Italy and the UK (M = 48.3; SD = 20.7), ($p = .034$) in concerns related to *interrogation*. Additionally, we found a significant difference ($p = .017$) between the USA (M = 68.4; SD = 20.1) and Nordic Countries (M = 60.6; SD = 19.7), and the USA and Italy (M = 60.4; SD = 23.0) about *distortion* ($p = .010$).

Table 6. Differences in privacy concerns among participants from different geographic areas; $N = 382$, $p < .05$ (One-Way ANOVA).

Source	SS	df	MS	F	p
Secondary use					
Between	5857.7	3	1952.5	5.0	.002
Within	147327.7	378	389.7		
Total	153185.4	381			
Interrogation					
Between	4311.6	3	1437.2	3.2	.022
Within	167642.2	378	443.4		
Total	171953.8	381			
Distortion					
Between	3974.4	3	1324.8	2.8	.036
Within	173576.6	378	459.1		
Total	177551.0	381			

We performed One-way ANOVA and the post-hoc Tukey test to asses whether there are potential differences in privacy concerns, protection behavior and information disclosure among participants from different age groups. For this purpose we divided our sample to four age groups: $18 - 24, 25 - 34, 35 - 44$ and

over 45 years old. We found a significant effect of age on concerns about the *unauthorized access* ($F(3, 378) = 4.860, p = .002$), *misuse of data* ($F(3, 378) = 3.094, p = .027$), *secondary use of data* ($F(3, 378) = 3.162, p = .013$), *insecurity* ($F(3, 378) = 4.710, p = .003$) and *exposure* ($F(3, 378) = 3.759, p = .011$). The participants belonging to $35-44$ and $18-24$ years old groups differed in perception about *unauthorized access* and *misuse of data*; *over* 45 and $18-24$ differed in perceptions of *exposure*; over 45 differed from $18-24$ and $25-34$ years old in perception of *secondary use* of data. Lastly, participants belonging to $35-44$ and *over* 45 years old differed from the $18-24$ years old in concerns about *insecurity*. We did not find any significant differences among participants from different age groups in relation to protection behavior and information disclosure.

Lastly, we used the independent t-Test to see whether there are significant gender differences about privacy concerns, but we did not find any $p < .05$. Similarly, we did not identify any gender dependencies in regards to both general caution and technical protection. Furthermore, we used Chi-Square test to determine whether the sensitive and non-sensitive information disclosure differed among males and females, however, once again the results were insignificant.

5 Discussion

To improve understanding of privacy perceptions we investigated privacy harms by creating the new scale measuring privacy concerns (RQ1). As we wanted to achieve a greater understanding of people's attitudes, we used the legal framework as a basis for the study design. The results demonstrated, that privacy perceptions vary from those identified by Solove. However, there are some resemblances. While Solove proposed to consider harms at the individual level, the results showed that people express privacy concerns differently. They tend to perceive concerns as comprehensive and simplified models. Possibly, such perception is related to the cognitive information processing, intending to decrease the cognitive effort and use affect heuristics.

We identified seven dimensions of privacy concerns: *insecurity, exposure, unauthorized access, secondary use of data, misuse of data, distortion* and *interrogation*. The analysis of the means suggests that people express high concerns about *security*. They want to be informed about data security breaches and in general, they expect that online services will guarantee safety. According to the findings, people worry about *exposure*, which may suggest that they care about online presence and information visibility. They want to be in control of personal information, ensuring that none of it is used without their knowledge or permission. The findings show general worries about the *secondary use of data*, such as selling or sharing data with external organizations, and about *misuse of data*, such as blackmail or malicious use of information by strangers to reach their own goals. *Distortion* seems to be less important, and *interrogation* is perceived as the least severe. Considering *interrogation*, paradoxically, respondents expressing concerns about secondary use or misuse of information did not find the information probing important. Overall, the new dimensions show similari-

ties to Solove's findings. The results show that almost all of the harms defined by Solove are subject to concern, however, not at the individual level and not accordingly with the process of information flow. Additionally, it seems that invasions are the one group of harms which is perceived as less severe than others.

The identified dimensions of privacy concerns relate to findings from the past research. For instance an improper access and secondary use of data, the two of four dimensions defined by CFIP [47]. Similarly, our findings relate to the factors identified by IUIPC: collection (*interrogation* and *insecurity*) and control (*exposure*, *distortion*) [32]. The seven dimensions of PHC add to the previous scales by identification of wider range of concerns. Our findings origin from participants with broad demographics, while CFIP was based on students and professionals from business environment, IUIPC was customer oriented. Furthermore, the PHC uncovers issues related to the *self* (me as a person and as a part of the society), such as *distortion* or *exposure*, showing that personal image, online reputation, fear of the damages, which could be caused by disclosed data are important factors causing privacy concerns.

Additionally, we investigated whether privacy perceptions differ among people who disclose sensitive and non-sensitive information (RQ2). The findings demonstrate that privacy concerns of participants who do not disclose both sensitive and non-sensitive information differ from those who disclose information. Respondents who do not disclose information expressed concerns about their data being sold to third parties and about providing feedback related their online activities. This result suggests that people concerned about their data ownership use preventive methods, such as non-disclosure, to ensure that none of their information, whether it has sensitive or non-sensitive nature, is provided to the online companies. Additionally, the results found that people's privacy concerns are the same among those who disclose/non-disclose sensitive and non-sensitive information. Presumably, if one worries about the privacy, he/she will behave in the same way regardless of information sensitivity.

Further, the study identified relationships between protection behaviors and privacy concerns (RQ2). Despite the low correlations between protection behaviors and privacy concerns, scatter-plots' analysis confirmed the relationships. Respondents with higher technical protection behavior seemed to have high concerns about the *unauthorized access*, *misuse* and *secondary use of data*, *insecurity*, *exposure* and *distortions*. The same applies to general caution, except there is no correlation with *secondary use of data*. Instead, the higher general caution, the higher *interrogation* concerns. Interestingly, our results did not find any correlation between technical protection and *interrogation*. This may suggest, that people using different technical protections may feel confident that data will not be sold or transferred to unknown organizations, because of users' preventive measures. On the other hand, it may be related to the fact, that people do not perceive *interrogation* as a very severe concern.

The demographic results indicate possible differences in privacy perceptions among respondents from different geographic locations, education and age groups (RQ3). We identified differences between respondents from different countries.

This could imply the role of cultural diversity in shaping people's concerns. However, due to the small sample size, our findings are only an indication of possible cultural dependencies, which require further studies.

Considering other demographics, our results show that people from older generations express more concerns about privacy than the younger generations, confirming findings from the previous research [51]. The age divide may be explained by the fact that older people have more experience, awareness, and knowledge related to privacy violations. Also, the younger population may use internet as a tool for communications, to develop social relationships or as a source of leisure activities, while older people may use it to cope with day-to-day activities, such as work, financial transactions, information source. For that reason, older generation may add more value to their online information, and in a result express stronger privacy concerns. On the other hand, as demonstrated in the past research, the younger generation may express fewer concerns due to their protection behaviors.

Limitations There is a number of limitations in this study. The method: self-reported survey, may decrease validity and reliability of the results. However, as the study was designed to reach international respondents within a short time, this method was the most effective. Similarly, the enlarged sample size could improve the results, especially the demographic assumptions. Furthermore, the research explored general privacy concerns and did not investigate whether they would change considering specific context, for instance, different technologies. The collected data did not allow to model causal relationships between concerns and behaviors. The investigation of causal relations could provide a better overview on the role of privacy concerns in the decision making.

6 Conclusion

This study contributes a new measurement instrument for privacy concerns. To differentiate it from the existing privacy scales, we aimed to shift the focus of privacy concerns to privacy harms, based on the framework developed by Solove. We demonstrated that identified privacy concerns vary among individuals, by analyzing self-reported behavior and demographics. The new instrument can be used in future studies assessing privacy attitudes.

Additionally, the results suggest that there are some general tendencies in privacy concerns. The findings show that people create simplified models of privacy harms, such as worries about security, unlawful use of data, disclosure or exposure. All of these concerns can be addressed by developers and designers to ensure privacy. Due to the similarities among people from different demographics, we can assume that there is a potential to build systems with 'privacy for all' or 'privacy with no borders'.

Future work Our privacy scale requires further validation in qualitative and quantitative studies. For instance, to improve the scale it is recommended to

implement it in experiments of the actual privacy behavior, using the PHC as pre- and/or post-questionnaire. Our results will be fundamental to develop models for instruments influencing peoples' behavior, nudging people's privacy choices and improving their privacy risk awareness. Similarly, further studies of PHC could result in the set of guidelines for developers and designers of privacy enhancing technologies (PET). Such guidelines could enable easier assessment of people's privacy needs, improving usability of PETs and in a result increasing users' satisfaction.

Acknowledgment This work has received funding from the European Unions Horizon 2020 research and innovation programme under the Marie Skłodowska-Curie grant agreement No 675730.
To obtain more information about the study or to gain access to the original questionnaire, please contact the corresponding author.

References

[1] A. Acquisti, L. Brandimarte, and G. Loewenstein. Privacy and Human Behavior in the Age of Information. *Science*, 347(6221):509–514, 2015.

[2] A. Acquisti and J. Grossklags. Privacy attitudes and privacy behavior. *Economics of Information Security*, pages 1–15, 2004.

[3] A. Acquisti and J. Grossklags. Privacy and Rationality in Individual Decision Making. Challenges in Privacy Decision Making. The survey. *IEEE Security & Privacy*, 3(1):26–33, 2005.

[4] A. Acquisti and C. Taylor. The Economics of Privacy . *Journal of Economic Literature*, 52:1–64, 2016.

[5] I. Adjerid, A. Acquisti, L. Brandimarte, and G. Loewenstein. Sleights of Privacy : Framing , Disclosures , and the Limits of Transparency. *Symposium on Usable Privacy and Security (SOUPS)*, page 17, 2013.

[6] K. Bartel Sheehan. An investigation of gender differences in on-line privacy concerns and resultant behaviors. *Journal of Interactive Marketing*, 13(4):24–38, 1999.

[7] S. Bellman, E. J. Johnson, S. J. Kobrin, and G. L. Lohse. International differences in information privacy concerns: A global survey of consumers. *Information Society*, 20(5):313–324, 2004.

[8] A. Bergström. Online privacy concerns: A broad approach to understanding the concerns of different groups for different uses. *Computers in Human Behavior*, 53:419–426, 2015.

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

[10] T. Buchanan, C. Paine, A. N. Joinson, and U.-D. Reips. Development of measures of online privacy concern and protection for use on the internet. *Journal of the Association for Information Science and Technology*, 58(2):157–165, 2007.

[11] R. Calo. The boundaries of privacy harm. *Indiana Law Journal*, 86(3), 2011.

[12] L. J. Camp. Mental models of privacy and security. *IEEE Technology And Society Magazine*, 28(3):37–46, 2009.

[13] H. Cho, M. Rivera-Sánchez, and S. S. Lim. A multinational study on online privacy: global concerns and local responses. *New media & Society*, 11(3):395–416, 2009.

[14] L. Coventry, D. Jeske, and P. Briggs. Perceptions and actions: Combining privacy and risk perceptions to better understand user behaviour. In *Symposium on Usable Privacy and Security (SOUPS) 2014*, 2014.

[15] S. J. De and D. Le Métayer. Privacy harm analysis: a case study on smart grids. In *Security and Privacy Workshops (SPW), 2016 IEEE*, pages 58–65. IEEE, 2016.

[16] T. Dinev and P. Hart. Internet privacy concerns and their antecedents - measurement validity and a regression model. *Behaviour & Information Technology*, 23(6):413–422, 2004.

[17] S. Egelman. "my profile is my password, verify me!": the privacy/convenience tradeoff of facebook connect. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, pages 2369–2378. ACM, 2013.

[18] M. Fagan and M. M. H. Khan. "Why Do They Do What They Do?": A Study of What Motivates Users to (Not) Follow Computer Security Advice. *Proceedings of the Symposium On Usable Privacy and Security (SOUPS)*, 2016.

[19] J. Fogel and E. Nehmad. Internet social network communities: Risk taking, trust, and privacy concerns. *Computers in human behavior*, 25(1):153–160, 2009.

[20] A. Gambino, J. Kim, S. S. Sundar, J. Ge, and M. B. Rosson. User disbelief in privacy paradox: Heuristics that determine disclosure. In *Proceedings of the 2016 CHI Conference Extended Abstracts on Human Factors in Computing Systems*, pages 2837–2843. ACM, 2016.

[21] M. Harbach, M. Hettig, S. Weber, and M. Smith. Using personal examples to improve risk communication for security & privacy decisions. In *Proceedings of the 32nd annual ACM conference on Human factors in computing systems - CHI '14*, pages 2647–2656, 2014.

[22] K. T. Ho and C. Li. From privacy concern to uses of social network sites: A cultural comparison via user survey. *Proceedings - 2011 IEEE International Conference on Privacy, Security, Risk and Trust and IEEE International Conference on Social Computing, PASSAT/SocialCom 2011*, pages 457–464, 2011.

[23] M. G. Hoy and G. R. Milne. Gender Differences in Privacy-Related Measures for Young Adult Facebook Users. *Journal of Interactive Advertising*, 10(2):28–45, 2010.

[24] A. N. Joinson, C. Paine, T. Buchanan, and U. D. Reips. Measuring self-disclosure online: Blurring and non-response to sensitive items in web-based surveys. *Computers in Human Behavior*, 24(5):2158–2171, 2008.

[25] F. Kehr, D. Wentzel, and T. Kowatsch. Privacy Paradox Revised: Pre-Existing Attitudes, Psychological Ownership, and Actual Disclosure. In *IS Security and Privacy*, pages 1–12, 2014.

[26] F. Kehr, D. Wentzel, T. Kowatsch, and E. Fleisch. Rethinking privacy decisions: pre-existing attitudes, pre-existing emotional states, and a situational privacy calculus. *ECIS 2015 Completed Research Papers*, 2015.

[27] B. P. Knijnenburg and A. Kobsa. Making Decisions about Privacy: Information Disclosure in Context-Aware Recommender Systems. *ACM Transactions on Interactive Intelligent Systems*, 3(23), 2013.

[28] S. Kokolakis. Privacy attitudes and privacy behaviour: A review of current research on the privacy paradox phenomenon. *Computers & Security*, 7(2):1–29, 2015.

[29] H. Krasnova, E. Kolesnikova, and O. Guenther. "It Won't Happen To Me!": Self-Disclosure in Online Social Networks. *AMCIS 2009 Proceedings*, page 343, 2009.

[30] C. Leys, C. Ley, O. Klein, P. Bernard, and L. Licata. Detecting outliers: Do not use standard deviation around the mean, use absolute deviation around the median. *Journal of Experimental Social Psychology*, 49(4):764–766, 2013.

[31] C. Lutz and P. Strathoff. Privacy concerns and online behavior - Not so paradoxical after all? *Multinationale Unternehmen und Institutionen im Wandel Herausforderungen für Wirtschaft, Recht und Gesellschaft*, pages 81–99, 2013.

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

[33] A. E. Marwick and D. Boyd. Networked privacy: How teenagers negotiate context in social media. *New Media & Society*, 16(7):1051–1067, 2014.

[34] C. L. Miltgen and D. Peyrat-guillard. Cultural and generational influences on privacy concerns: a qualitative study in seven European countries. *European Journal of Information Systems*, 23(2):103–125, 2014.

[35] T. Opinion and Social. Special Eurobarometer 431 'Data Protection'. Technical report, European Union, 2015.

[36] Y. J. Park. Do men and women differ in privacy? Gendered privacy and (in)equality in the Internet. *Computers in Human Behavior*, 50:252–258, 2015.

[37] Y. J. Park, S. W. Campbell, and N. Kwak. Affect, cognition and reward: Predictors of privacy protection online. *Computers in Human Behavior*, 28(3):1019–1027, 2012.

[38] S. Preibusch. Guide to measuring privacy concern: Review of survey and observational instruments. *International Journal of Human Computer Studies*, 71(12):1133–1143, 2013.

[39] S. Preibusch, K. Krol, and A. R. Beresford. The privacy economics of voluntary over-disclosure in web forms. In *The Economics of Information Security and Privacy*, pages 183–209. Springer, 2013.

[40] A. Raij, A. Ghosh, S. Kumar, and M. Srivastava. Privacy risks emerging from the adoption of innocuous wearable sensors in the mobile environment. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, pages 11–20. ACM, 2011.

[41] P. J. Reed, E. S. Spiro, and C. T. Butts. Thumbs up for privacy?": Differences in online self-disclosure behavior across national cultures. *Social Science Research*, 2016.

[42] G. D. P. Regulation. Regulation (eu) 2016/679 of the european parliament and of the council of 27 april 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing directive 95/46. *Official Journal of the European Union (OJ)*, 59:1–88, 2016.

[43] T. G. Reio and B. Shuck. Exploratory Factor Analysis: Implications for Theory, Research, and Practice. *Advances in Developing Human Resources*, 17(1):12–25, 2015.

[44] D. Roback and R. L. Wakefield. Privacy risk versus socialness in the decision to use mobile location-based applications. *ACM SIGMIS Database*, 44(2):19, 2013.

[45] S. Sheth, G. Kaiser, and W. Maalej. Us and Them: A Study of Privacy Requirements Across North America, Asia, and Europe. *Proceedings of the 36th International Conference on Software Engineering*, pages 859–870, 2014.

[46] C. V. Slyke, J. T. Shim, R. Johnson, and J. Jiang. Concern for Information Privacy and Online Consumer Purchasing. *Journal of the Association for Information Systems*, 7(6):415–444, 2006.

[47] H. Smith, S. Milberg, and S. Burke. Information privacy: measuring individuals' concerns about organizational practices. *MIS Quarterly*, 20(2):167–196, 1996.

[48] D. Solove. A taxonomy of privacy. *University of Pennsylvania Law Review*, 154(477):477–560, 2006.

[49] D. Solove and W. Hartzog. The FTC and the New Common Law of Privacy. *Columbia Law Review*, 114(3):583–676, 2014.

[50] D. J. Solove. I've got nothing to hide and other misunderstandings of privacy. *San Diego L. Rev.*, 44:745, 2007.

[51] W. M. Steijn, A. P. Schouten, and A. H. Vedder. Why concern regarding privacy differs: The influence of age and (non-) participation on facebook. *Cyberpsychology: Journal of Psychosocial Research on Cyberspace*, 10(1), 2016.

[52] F. Stutzman, R. Capra, and J. Thompson. Factors mediating disclosure in social network sites. *Computers in Human Behavior*, 27(1):590–598, 2011.

[53] Y. Sun, N. Wang, X. L. Shen, and J. X. Zhang. Location information disclosure in location-based social network services: Privacy calculus, benefit structure, and gender differences. *Computers in Human Behavior*, 52:278–292, 2015.

[54] Sun Sun Lim, Hichang Cho, and M. Rivera-Sanchez. A multinational study on online privacy: global concerns and local responses. *New Media & Society*, 11(3):395–416, 2009.

[55] M. Taddicken. The 'Privacy Paradox' in the Social Web: The Impact of Privacy Concerns, Individual Characteristics, and the Perceived Social Relevance on Different Forms of Self-Disclosure. *Journal of Computer-Mediated Communication*, 19(2):248–273, 2014.

[56] S. Trepte, T. Dienlin, and L. Reinecke. Risky behaviors: How online experiences influence privacy behaviors. *Von Der Gutenberg-Galaxis Zur Google-Galaxis. From the Gutenberg Galaxy to the Google Galaxy*, 2014.