Data, data, everywhere: quantifying software developers' privacy attitudes

Dirk van der Linden[1], Irit Hadar[2], Matthew Edwards[1], and Awais Rashid[1]

[1] Bristol Cyber Security Group
University of Bristol, Bristol, UK
{dirk.vanderlinden,matthew.john.edwards,awais.rashid}@bristol.ac.uk
[2] Department of Information Systems
University of Haifa, Haifa, Israel
hadari@is.haifa.ac.il

Abstract. Understanding developers' attitudes towards handling personal data is vital in order to understand whether the software they create handles their users' privacy fairly. We present the results of a study adapting an existing user-focused privacy concern scale to a software development context and running it with a sample of 123 software developers, in order to validate it and develop a model for measuring the extent to which a software developer is (dis)favorable to ensuring their users' privacy. The developed scale exceeds thresholds for internal reliability (α>.8), composite reliability (CR>.8), and convergent validity (AVE>.6). Our findings identified a model consisting of three factors that allows for understanding of developers' attitudes, including: (1) informed consent, (2) data minimization, and (3) data monetization. Through analysis of results from the scale's deployment, we further discuss mismatches between developers' attitudes and their self-perceived extent of properly handling their users' privacy, and the importance of understanding developers' attitudes towards data monetization.

Keywords: developer · privacy · attitude · scale development.

1 Introduction

Understanding individual developers' attitudes towards privacy – more specifically, their *attitude* towards handling personal data, is an important precursor to understand what drives the formation of their privacy practices while developing software. Attitudes – mental constructs which reveal the extent of positive or negative feelings someone holds towards a particular thing [12] – are an important precursor to behavioral intention. The Theory of Reasoned Action (TRA) posits that behavioral intention via attitude combined with social norms are key predictors to whether someone will behave in a particular way [19]. A software developer's attitude towards the handling of personal data is thus an important aspect of understanding how they will handle personal data in reality.

Yet, little work exists to aid in the large-scale measurement of software developers' attitudes towards privacy, or, more practically, their handling of personal data. A scale proposed by Woon and Kankanhalli [29] allows for measurements of developers' attitudes towards incorporating security into application development. But, security is not privacy, and security mindsets have been shown to push developers towards understanding privacy as little more than a *technical* data security issue, discarding the much larger *socio-technical* considerations that properly handling privacy entails [6,15]. More encompassing scales to measure attitudes regarding privacy exist from a consumer's point of view (cf. [7,17,26]). A review of such scales [21] showed that, even in the context of a general population of users, privacy attitudes are elicited in an ad-hoc manner through questionnaires.

To that end, this paper aims to contribute by addressing the lack of systematically developed scales measuring privacy attitude of software developers. To do so, we adapted a widely used and validated scale for internet users' information privacy concerns (IUIPC) [17] to a software development context, empirically testing it among a sample of professional software developers (N=123), and constructing a software developer-specific model from factors arising out of the data. We make the following major contributions:

- *We present the Software Developers' Privacy Attitude (SDPA) scale for measuring the extent to which a software developer is (dis)favorable to ensuring their users' privacy.* Our analysis identified a three-factor model capturing software developers' attitudes towards handling personal data of users of their software: (1) *informed consent:* the extent to which they ensure their users are given the option and ability to provide informed consent, (2) *data minimization:* the extent to which they minimize the data they collect from users, and (3) *data monetization:* the extent to which they perceive monetizing data as impacting user privacy.
- *We discuss mismatches between developers' attitudes and their self-perceived extent of 'properly' handling their users' privacy.* In particular, our application of the scale showed that developers' attitudes towards data collection reveals they may collect more data than their self-perceived behavior would indicate. More such mismatches may exist depending on the development context and type of personal data handled, which requires more work to point out such potentially dangerous mismatches of attitude and self-perceived behavior.
- *We discuss the importance of understanding developers' attitudes towards the third identified factor of data monetization.* This factor arose out of the data as the most strongly loaded factor, contrasting original work from users' point of view, showing many developers do not look disfavorably on monetizing user data in marketing transactions (through, e.g., advertisement analytics). Because software development is increasingly the domain of solo or small scale developer teams [11,25] who make these monetization decisions, more research is necessary to understand how advertisement-reliant ecosystems such as mobile and web applications may push developers towards understating the impact their decisions have on their users' privacy.

The rest of this paper proceeds as follows. Section 2 discusses the background and related work. Construction of the scale and its implementation are shown in Section 3, and reflections on its further development and use are discussed in Section 4. Finally, we conclude in Section 5.

2 Background & Related Work

2.1 Understanding Privacy Attitudes

Senarath and Arachchilage [22] indicate that developers have practical challenges when attempting to embed privacy into their software, in particular relating privacy requirements into engineering techniques, and lack knowledge on privacy concepts. Specific concepts of privacy aware systems such as data minimization were similarly found to be difficult [23].

In order to achieve a fuller picture of software developers' privacy perceptions and overall mindset, Hadar et al. [15] conducted a qualitative investigation using in-depth semi-structured interviews. They found that developers largely approach privacy through a data security lens, focusing on technical aspects and security solutions. They further found that developers' work environment, and in particular the organizational privacy climate, plays an important role in shaping developers' privacy perceptions and attitudes. These findings were qualitatively substantiated; a quantitative scale is required to test correlations between environmental (or other) factors and developers' privacy attitudes.

Ayalon et al. [3] performed an online survey based on example scenarios, to assess software developers' privacy attitudes. They were able to demonstrate evidence possibly suggesting the effect of organizational climate on developers privacy attitudes and behavior. They further found that personal experience as end users affects developers' privacy practices. However, they did not perform a systematic scale development, and the items used to assess personal attitudes had low internal reliability. Our work goes beyond here by showing the systematic adaptation of a scale for quantification of software developers' privacy attitudes, available for other researchers to employ.

2.2 Quantifying Privacy Attitudes

In this paper, we adapt the Internet Users Information Privacy Concern (IUIPC) scale [17], which in itself is an adaption and extension of the Concern for Information Privacy (CFIP) scale [26]. As mentioned in the introduction, we use the more general psychological notion of 'attitude', which refers to mental constructs which reveal the extent of positive or negative feelings someone holds towards a particular thing [12]. The IUIPC focuses on consumers' privacy *concerns* – in effect focusing on eliciting specifically the negative feelings, or *attitudes*, that participants hold towards the presented items. We use attitude to allow the scale to clearly capture developers' positive and negative view of data practices, thereby permitting us to contrast how developers may view certain practices positively (e.g., making money by transferring data to third parties) while their users may view them negatively (i.e., by losing control over their data).

The IUIPC is based on the notion of fairness and justice, assuming that the essence of privacy lies in fair and just handling of personal data, from which its three main dimensions flow: personal data is only handled fairly and justly if (1) it is collected appropriately (*collection*), (2) its data subjects are made aware of that collection (*awareness*)), and they are given control over it (*control*). These notions clearly align with the normative context set by regulation such as the GDPR and some of its key articles software has to abide by – including data minimization Art. 5(1)(c); lawfulness, fairness and transparency Art. 12–14; and control: Arts. 15–21.

Malhotra et al. note that with appropriate rewording the scale is expected to reasonably apply to other privacy contexts [17, p. 349] – such as software development. Ensuring an appropriate rewording to this context requires consideration of what just and fair handling of personal data means from a software developers' point of view.

However, developers' attitudes are likely to be more nuanced, reflecting their need to balance the data they collect and the control and awareness they give users over it with their own business demands (e.g., having to incorporate advertisement analytics SDKs to achieve monetization) and legal compliance (e.g., having to comply with relevant data protection acts). Not all developers will have the luxury of delegating business decisions to others, and many have to deal with decisions about how to handle personal data. The extent to which handling of personal data is just, and more importantly, *fair* to developers, will likely incorporate trade-off considerations between what benefits the user, and what benefits the developer or their organization's bottom line. As a result, there may be different attitudes towards, and relations between, the items from the IUIPC. Thus, our paper represents a starting point to explore how developers' privacy attitudes may quantitatively differ from users. To the best of our knowledge, this work is the first to do so in a systematic fashion.

3 Development of the SDPA scale and model

This section will detail the adaption of the items from the IUIPC to a software development context, through its deployment with a sample of software developers, and the statistical analysis performed to construct the final model.

3.1 Adapting the IUIPC scale to software development

Table 1 summarizes the key aspects of the original consumer-focused scale, and how we adapt it to a software development context. We adapted the consumer-focused items developed or adapted by Malhotra et al. [17] to a software development context as per Table 1 (e.g., online companies ⇒ the software I develop; consumer ⇒ my user). Full details are shown in Appendix B. This resulted in items (a)–(i):

Table 1. Comparison between IUIPC and SDPA.

	IUIPC	**SDPA**
Purpose	To reflect Internet users' concerns about information privacy	To reflect Software developers' attitudes towards handling users' personal data and ensuring privacy
Focus	Individuals' perceptions of fairness/justice in the context of information privacy	Software developers' perception of fairness/justice towards their users in context of personal data handling
Context	Mostly online environment	Software development
Communication	Mostly two-way communication	*ibid*
Dimensions	Collection, control, awareness of privacy practices	informed consent, data minimization, data monetization
Representation	Second-order factor	*ibid*

(a) It usually bothers me when the software I develop asks my users for personal information.
(b) I sometimes think twice before asking my users for personal information with the software I develop.
(c) It bothers me to collect personal information from so many users with the software I develop.
(d) I'm concerned that the software I develop is collecting too much personal information about my users.
(e) My users privacy is really a matter of their right to exercise control and autonomy over decisions about how their information is collected, used, and shared by the software I develop.
(f) My users control of personal information collected by the software I develop lies at the heart of user privacy.
(g) I believe that my users' privacy is invaded when control is lost or unwillingly reduced as a result of a marketing transaction.
(h) The software I develop should disclose the way the data are collected, processed, and used.
(i) A good privacy policy for the software I develop should have a clear and conspicuous disclosure.
(j) It is very important to me that my users are aware and knowledgeable about how their personal information will be used.

Based on the three main factors identified by Malhotra et al. [17] as predictive of behavioral intent for privacy concern, we adapted three questions to assess how software developers perceive themselves to fairly/justly handle personal data in software development. This resulted in items (i)–(iii):

(i) I properly deal with the extent to which my software collects data of its users

(ii) I properly deal with the extent to which my software gives users control over their data

(iii) I properly deal with the extent to which my software informs its users how their data is used

The full questionnaire developed with these items (shown in randomized order to participants) is shown in Appendix A.

3.2 Deploying the adapted scale

To test the adapted scale, we performed a questionnaire-based study measuring the attitude of software developers towards their users' privacy using the newly adapted SDPA scale. Each item was followed by a 7pt scale anchored with "strongly disagree" and "strongly agree". We obtained approval from our Institutional Review Board (IRB) before any empirical work began. No personal information was captured from any participants.

We used Prolific [1] to recruit participants employed as software developers. In total 123 developers completed the study in the 3-day run-time. Each participant was paid £0.30 for completion of the study. A summary of relevant demographics is shown in Table 2. Note that most covariates (sex, age, etc.) were not explicitly elicited through the questionnaire as Prolific provided these data. Results of the deployment for all items are shown in Figures 1– 2.

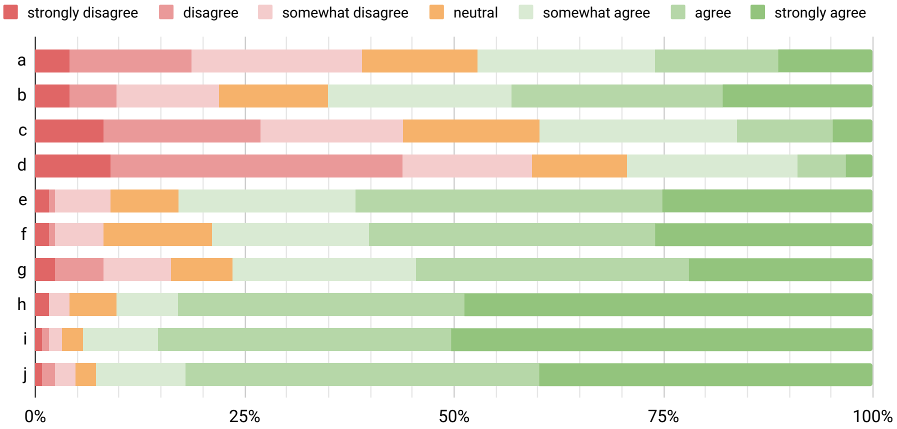

Fig. 1. Detailed results for all items (n=123). Factor 1 includes items e, f, h, i, j; Factor 2 includes items a, b, c, d; Factor 3 includes item g.

Table 2. Demographic data of developers (N=123) from the scale's first deployment.

		(N=123)
Age	18–24	16%
	25–34	59%
	35–44	13%
	> 44	14%
Sex	Male	79%
	Female	21%
Type of software	Web	85%
	Desktop	57%
	Mobile	51%
	Plugins	30%
	Embedded	15%
	OS	10%
Type of employment	Full-time	81%
	Part-time	15%
	Other	4%
Place of employment	Europe	59%
	North America	37%
	Australasia	4%
	South America	1%

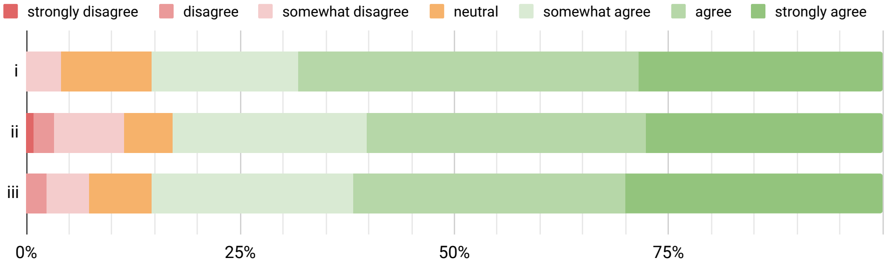

Fig. 2. Detailed results for self-perceived proper handling of personal data (n=123).

3.3 Constructing the SDPA model

Our initial approach to building the SDPA scale for developers adopted the IUIPC's mapping of the 10 items to 3 factors (Collection: a–d, Control: e–g, and Awareness: h–j) for consumers – as indicated by the table segments in Table 3.

To validate this measurement model, we conducted a confirmatory factor analysis (CFA) of goodness-of-fit for the consumer model on our data, evaluated in terms of comparative fit index (CFI), goodness-of-fit index (GFI) and root mean square error of approximation (RMSEA). A model is considered to be satisfactory if CFI > 0.95, GFI > 0.90 and RMSEA < 0.06 [17]. Our adaption of the IUIPC's scale following their consumer-focused model showed acceptable CFI (0.96) and GFI (0.93) but poor RMSEA (0.07).

As such, we explored a better factor loading of the items to account for software developers' context. To do this, we performed a principal component and factor loading analysis (see Table 3). The identified factors are based on item loadings above 0.6 indicating sufficient convergent validity [9]. Subsequently, we calculated a correlation matrix of all items (see Table 4).

Table 3. Results from component analysis. All factor loadings above .40 are listed below. Interim reliabilities (Cronbach's α): F1, .84; F2, .81; F3 n/a. For comparison, the original model Malhotra et al. defined included three factors: a–d, e–g, and h–j.

Item	**Factor 1**	**Factor 2**	**Factor 3**
a		.783	
b		.644	
c		.886	
d		.776	
e	.759		
f	.71		
g			.902
h	.786		
i	.809		
j	.789		

Table 4. Correlation matrix of items. Correlations significant at $p < .05$ are shown.

item	a	b	c	d	e	f	g	h	i	j
a	–									
b	0.46	–								
c	0.69	0.47	–							
d	0.45	0.25	0.58	–						
e		0.36			–					
f	0.23	0.28			0.41	–				
g	0.33	0.22				0.31	–			
h		0.25	0.22		0.45	0.48	0.32	–		
i		0.26			0.39	0.54		0.57	–	
j	0.24	0.27			0.37	0.59	0.24	0.54	0.51	–

Table 3 indicates a three-factor model captures developers' attitudes best. These factors measure developers' attitudes towards respectively (1) *informed consent*, (2) *data minimization*, and (3) *data monetization* A two-factor model would have discarded item (g), which loads strongly onto the third factor otherwise – even establishing convergent validity. While this represents a single scale measure, it is in line with other psychological work showing similar validity between single and multiple scale measures [28,8], likely due to its specific focus. Moreover, the importance of including this factor lies also in its unique developer-specific view on handling personal data, giving information not yet provided by other scales.

The revised SDPA model for developers passed all three cut-offs under CFA [CFI= 0.97, GFI= 0.93, RMSEA = 0.06]. The collection factor overlaps with the factor identified by Malhotra et al. for users, but developers' views towards control and awareness diverge, with items spread across two distinct, new factors. Table 5 summarizes the final factors and relevant descriptive statistics. All three factors exhibit convergent validity (CR and AVE above resp. .7 and .5 [4]), and discriminant validity beecause square root of AVE is larger than correlation co-efficients of the factors [14,10]. Additionally, a repeated measures T test confirmed all factors' responses differed significantly (F1–F2: t(122)=-14.87, p<0.01; F1–F3: t(122)=-4.58, p<0.01; F2–F3: t(122)=7.55, p<0.01).

Table 5. Factor summary statistics and estimated correlation matrix

							Rho (p<0.05)†		
Factor	**Mean**	**SD**	α	σ**2**	**CR***	**AVE***	*1*	*2*	*3*
1. Informed consent	5.91	1.25	.84	1.55	.88	.59	*.77*		
2. Data minimization	4.06	1.76	.80	3.10	.86	.60	.22	*.78*	
3. Data monetization	5.26	1.57	n/a	2.46	.81	.81	.29	.26	*.90*

* Internal reliability requires $\alpha > 0.7$, convergent validity requires CR> .7 and AVE > .5

† Data on diagonals indicates squared root of AVE.

3.4 Assessing model–variable correlations

We found no indication of an effect of contextual variables or demographic covariates on the SDPA score. Other studies using the original scale we adapted varied in finding some correlations for covariates like age and educational level (cf. [30]), to only finding correlations for educational level and income, with more nuanced differences between age groups (cf. [16]). The lack of correlations to covariates is likely due to high homogeneity in the sample such as most developers being male, similar age ranges, and most having developed web apps likely to elicit more privacy considerations. We did establish a significant correlation between score on the SDPA and self-perceived behavior (Spearman's Rho, r(121)=0.30, p<.05). Exploring this in more detail, relating our factors back to the self-perceived behavior's original factor model reveals only strong correlations between our *informed consent* factor and perceived behavior, as shown in Table 6. Most informative here is the lack of a key correlation: factor 2, *data minimization*, is not significantly correlated to the item measuring self-perceived behavior regarding data minimization (Spearman's Rho, r(121)=0.15, p>.05). The detailed results shown in Figures 1– 2 visualize this by showing that participants rated their own behavior towards data minimization as highly proper, while their attitude reveals a less fair & just approach.

Table 6. Correlation matrix of factors and self-perceived items. Correlations significant at p<.05 are shown.

item	**Factor 1**	**Factor 2**	**Factor 3**
pCOL	.53	!	.22
pCON	.48		
pAWA	.55		
pIUIPC	.57		
SDPA	.65	.83	.5

3.5 Constructing the final instrument

The final SDPA instrument as developed here contains three factors measuring developers' attitudes towards key aspects of handling personal data, as defined below. To measure the extent to which a software developer is (dis)favorable to these aspects, the items should be accompanied by a 7pt scale anchored with "strongly disagree" and "strongly agree".

the SDPA instrument

Factor 1, *informed consent*: the extent to which developers ensure their users are given the ability and option to provide informed consent.

- My users' privacy is really a matter of their right to exercise control and autonomy over decisions about how their information is collected, used, and shared by the software I develop.
- My users' control of personal information collected by the software I develop lies at the heart of user privacy.
- The software I develop should disclose the way the data are collected, processed, and used.
- A good privacy policy for the software I develop should have a clear and conspicuous disclosure.
- It is very important to me that my users are aware and knowledgeable about how their personal information will be used.

Factor 2, *data minimization*: the extent to which developers minimize the data they collect from their users.

- It usually bothers me when the software I develop asks my users for personal information.
- I sometimes think twice before asking my users for personal information with the software I develop.
- It bothers me to collect personal information from so many users with the software I develop.
- I'm concerned that the software I develop is collecting too much personal information about my users.

Factor 3, *data monetization*: the extent to which developers perceive transferring data to marketing parties as impacting users' privacy.

- I believe that my users' privacy is invaded when control is lost or unwillingly reduced as a result of a marketing transaction.

While the covariates investigated in this work did not yield any significant correlations, further work deploying the scale in specific software development situations may yield meaningful context-specific covariates, and allow for testing of other detailed variables.

3.6 Threats to validity

Internal validity. We adapted validated measures in the study to ensure accurately measuring the needed concepts. All identified factors displayed strong item loading, achieving internal and composite reliability, as well as convergent validity across their items. The data monetization factor so far consists of a single measure, which, while possible to extend to a multi-item measure in future work, is acceptable from a statistical (its convergent validity being well above established thresholds) and psychological (cf. [28,8]) point of view.

External validity. The generalization of these results is limited to some extent by the use of Prolific, which presents a Western bias – most developers, even if spread geographically, worked in Western countries and had English as their first language. However, given the focus of privacy research on this same domain, specifically with European regulations, we accept this as a workable constraint. The identified model presents three key factors identifying developers' attitudes towards handling personal data of their users, but should not be taken as presenting a complete picture of their attitude towards privacy. Further work may expand the model by identifying additional factors, and/or contexts in which factors' salience changes.

4 Discussion – Use Cases for the SDPA scale

This section will explore potential uses of the developed scale, and how the results elicited during our construction of the scale already hint towards interesting points of further application.

4.1 Identifying mismatches between attitude and (self-perceived) behavior

Figure 1 shows that software developers seem to be comfortable asking for a lot of data from their users. Many disagree with being bothered by collecting personal information from (many) users through the software they develop. Yet, when asked if they deal 'properly' with the extent to which their software collects data of its users, developers predominantly agreed (see Fig. 2, item (i)). This mismatch is further shown by the lack of expected correlation between the model's data collection factor (Spearman's rho, $\mathrm{r}{=}.15, \mathrm{p}{>}.1$), as compared to other correlations shown in Table 6). Thus, there seems to be a difference of opinion between what developers typically think is proper with regards to extent of personal data collection, and what principles like the Fair Information Practice Principles (FIPP) and legislation like the GDPR [20] set as proper data collection behavior. In particular, the FIPP's *collection limitation principle* notes there should be limits to collection of personal data, explicitly worked out in Art. 5(1)(c) of the GDPR:

> "Personal data shall be adequate, relevant and limited to what is necessary in relation to the purposes for which they are processed (data minimisation)"

Developers' attitudes as measured here show that they are not in line with this notion of data minimization. This is potentially dangerous from a compliance perspective – but more importantly, shows that the social norms surrounding developers in their professional work are not yet where privacy preserving work intends them to be. The application of this scale thus allows for identification of areas where developers need to be made more aware of how they can value and ensure users' privacy.

A similar mismatch exists also between developers' attitudes and how they perceive themselves to properly ensure giving users control over their data. Items e and f in Figure 1 show that nearly a quarter of developers disagree with the principle that users' right and ability to control their personal information is vital for their privacy. This is in stark contrast to the rights of data subjects (i.e., users) set out in Arts. 12–23 (GDPR), including e.g., right of access, rectification, erasure to data – none of which can happen without a user's adequate ability to control their personal data.

Our query of self-perceived behavior asked developers whether they 'properly' dealt with the various fair and just factors that constitute user concern, but did not specify whether this was to be interpreted morally or legally. Given the above mismatches between developer attitudes and legal specifications, this question could be explored in future work: do developers see themselves as operating to their own (sub-legal) standard, or do they believe they are operating in accordance with relevant laws and regulations which (they believe) are overly sensitive to user concerns?

Further work should also assess other mismatches such as, investigating whether specific types of software (e.g., those working on mobile software) or using specific mechanisms (e.g., those using monetization SDKs) lead to different types of mismatches between developer attitude and regulatory principles set forth in the GDPR.

4.2 Investigating monetization's effect on privacy attitude

In the development of the model, data monetization stood out clearly as a significant component of software developers' privacy attitudes. As Figure 1 shows in item (g), a quarter of developers do not look disfavorably towards monetizing user data in marketing transactions. Unless communicated clearly, and explicitly establishing a lawful basis for transfering such data (i.e., obtaining explicit consent from the user), such transfers would be not in line with the principle of purpose limitation set out in Art. 5(1)(b) (GDRP). However, often such transfers will happen with the consent of the user even though they are not aware of it as the 'consent' was given in a longer, confusing text, or users simply did not understand the potential impact on their privacy when agreeing.

In this particular case, it is not the legal minutia that is most interesting – it is understanding why software developers would do this. It most obviously links to a need for *monetization* – the increasing pressure for developers to achieve return-on-investment from software they write. Looking at one of the most represented software types, mobile apps, a recent European Parliament briefing shows that the EU app economy is highly successful, accounting for approximately one third of revenues in the global market [18].

To make money with such software, several new revenue models have become widespread over the past decade, such as advertisements, micro-purchases, and so on. But making money with mobile apps is hard, and many developers make very little money indeed [24]. Effectively, advertisements rule the world as a revenue model [27], being used in nearly 40% of all apps. The use of such advertisement libraries brings security and privacy challenges with them, as several malicious advertisement libraries such as Xavier [13] and [5] have been found to put users' privacy at risk by stealing their personal data. Careful selection of which advertisement library to trust is thus a matter of trade-offs between promised revenue, and perceived risk – not of the users' privacy being impacted, but of the developer being held liable for it.

The difficulty of monetizing software in this sector may offer an explanation for the lack of disfavorable attitudes towards monetizing user data in marketing transactions. The European Parliament briefing further noted that many developers have expressed concerns about privacy regulations and further proposals, claiming they would create a disproportionate burden on them [18] – impacting their ability to generate income through revenue models like these.

We would argue this matter needs insight into developers' privacy attitudes, but cannot be approached in isolation – the socio-economic context that shapes their very *need* to trade-off user privacy for achieving some revenue is a complex system of inter-woven personal, economic, and regulatory factors and requires its due attention in further work.

4.3 Theory development through combined application of the scale

Many other applications exist for the proposed scale in order to further develop theory of software developers' privacy attitudes. Some particular contexts we envision for further work include:

Determining risk and benefit trade-offs. The relationship between perceived risk and perceived benefit is well established in psychological literature [2], showing that this relationship is inverse. Further theory development could assess to what extent decisions of developers that may be beneficial to them, such as using advertisement libraries that pose a potential risk for their users' privacy vs. perceived low likelihood of being fined under extant data protection legislation.

Determining the link between security and privacy in development. In order to establish software that safeguards its users' privacy, security must be designed into it from the start as well. The extent to which developers' attitudes towards handling personal data and their intention to practice secure application development (cf. [29]) can allow for further insight into when and where security and privacy mindsets are separate or complimentary.

Determining the impact of developer privacy attitudes on their software. A reliable quantitative scale for developers' privacy attitude gives us a measure which may be used to quantify the impact of developer attitude and attitudinal/culture interventions on the software they develop for their users. These could be established in terms of user concerns about particular software (i.e., correlation between some users' IUIPC scores for a piece of software, and the SDPA score of its developers) or more direct privacy outcomes such as breaches reported, data items collected, and user awareness.

5 Conclusion

In this paper we presented the development of a scale to measure software developers' attitudes towards how they handle personal data in the software they develop, conducted a study with 123 software developers, and discussed points of interest that arose for further research.

We showed that the scale achieved internal and composite reliability and convergent validity over its items. The model that emerged from the developed scale with high goodness-of-fit pinpointed three factors that help to understand software developers' attitudes: (1) informed consent, (2) data minimization, and (3) data monetization. Through analysis of the scale's first use we showed that there exist mismatches between developers' attitudes on the one hand, and their self-perceived behavior on the other hand. Monetization, in particular, presents such a mismatch where further study in the complex socio-economic reality of software development is needed to understand why developers may wittingly impact their users' privacy through the use of revenue models such as advertisements.

Finally, we proposed a number of further research directions to build out theory of software developers' privacy attitudes, including determining risk and benefit trade-offs, and links between secure development and privacy-minding development.

Acknowledgments. This work is partially supported by EPSRC grant EP /P011799/1, Why Johnny doesn't write secure software? Secure software development by the masses, and by the Center for Cyber Law & Policy (CCLP), established by the University of Haifa in collaboration with the Israeli National Cyber Bureau.

References

1. Prolific. https://www.prolific.ac (2019), online; last accessed 8 May 2019
2. Alhakami, A.S., Slovic, P.: A psychological study of the inverse relationship between perceived risk and perceived benefit. Risk Analysis **14**(6), 1085–1096 (1994)
3. Ayalon, O., Toch, E., Hadar, I., Birnhack, M.: How developers make design decisions about users' privacy: The place of professional communities and organizational climate. In: Companion of the 2017 ACM Conference on Computer Supported Cooperative Work and Social Computing. pp. 135–138. ACM (2017)
4. Bagozzi, R.P., Yi, Y.: On the evaluation of structural equation models. Journal of the academy of marketing science **16**(1), 74–94 (1988)
5. Bauer, A., Hebeisen, B.: Igexin advertising network put user privacy at risk. https://blog.lookout.com/igexin-malicious-sdk (2017), online; accessed 20 October 2018
6. Birnhack, M., Toch, E., Hadar, I.: Privacy mindset, technological mindset. Jurimetrics **55**, 55 (2014)
7. Buchanan, T., Paine, C., Joinson, A.N., Reips, U.D.: Development of measures of online privacy concern and protection for use on the internet. Journal of the American Society for Information Science and Technology **58**(2), 157–165 (2007)
8. Cheung, F., Lucas, R.E.: Assessing the validity of single-item life satisfaction measures: Results from three large samples. Quality of Life research **23**(10), 2809–2818 (2014)
9. Chin, W.W., Gopal, A., Salisbury, W.D.: Advancing the theory of adaptive structuration: The development of a scale to measure faithfulness of appropriation. Information systems research **8**(4), 342–367 (1997)
10. Chin, W.W., et al.: The partial least squares approach to structural equation modeling. Modern methods for business research **295**(2), 295–336 (1998)
11. Cravens, A.: A demographic and business model analysis of today's app developer. GigaOM Pro (2012)
12. Eagly, A., Chaiken, S.: Attitude structure and function, pp. 269–322. Oxford Univeristy Press, 4th edn. (1998)
13. Ecular Xu (Mobile Threat Response Engineer): Analyzing Xavier: An Information-Stealing Ad Library on Android. https://blog.trendmicro.com/trendlabs-security-intelligence/analyzing-xavier-information-stealing-ad-library-android/ (2017), online; accessed 20 October 2018
14. Fornell, C., Larcker, D.F.: Evaluating structural equation models with unobservable variables and measurement error. Journal of marketing research **18**(1), 39–50 (1981)

15. Hadar, I., Hasson, T., Ayalon, O., Toch, E., Birnhack, M., Sherman, S., Balissa, A.: Privacy by designers: software developers privacy mindset. Empirical Software Engineering **23**(1), 259–289 (2018)
16. Lee, H., Wong, S.F., Chang, Y.: Confirming the effect of demographic characteristics on information privacy concerns. In: PACIS 2016 Proceedings (2016)
17. Malhotra, N.K., Kim, S.S., Agarwal, J.: Internet users' information privacy concerns (iuipc): The construct, the scale, and a causal model. Information systems research **15**(4), 336–355 (2004)
18. Marcin Szczepaski: European app economy: State of play, challenges and EU policy. http://www.europarl.europa.eu/RegData/etudes/BRIE/2018/621894/EPRS_BRI(2018)621894_EN.pdf (2018), online; accessed 25 June 2019
19. Montano, D.E., Kasprzyk, D.: Theory of reasoned action, theory of planned behavior, and the integrated behavioral model. Health behavior: Theory, research and practice pp. 95–124 (2015)
20. Parliament, E.: 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. Official Journal of the European Union (2016), online; accessed 24 June 2019
21. Preibusch, S.: Guide to measuring privacy concern: Review of survey and observational instruments. International Journal of Human-Computer Studies **71**(12), 1133–1143 (2013)
22. Senarath, A., Arachchilage, N.A.: Why developers cannot embed privacy into software systems?: An empirical investigation. In: Proceedings of the 22nd International Conference on Evaluation and Assessment in Software Engineering 2018. pp. 211–216. ACM (2018)
23. Senarath, A., Arachchilage, N.A.G.: Understanding software developers' approach towards implementing data minimization. arXiv preprint arXiv:1808.01479 (2018)
24. SlashData Developer Economics: Developer Economics: State of the Developer Nation Q1 2015. https://www.developereconomics.com/reports/developer-economics-state-of-developer-nation-q1-2015 (2015), online; accessed 25 June 2019
25. SlashData Developer Economics: Developer Economics: State of the Developer Nation Q1 2016. https://www.developereconomics.com/reports/developer-economics-state-of-developer-nation-q1-2016 (2016), online; accessed 18 September 2017
26. Smith, H.J., Milberg, S.J., Burke, S.J.: Information privacy: measuring individuals' concerns about organizational practices. MIS quarterly pp. 167–196 (1996)
27. VisionMobile / SlashData Developer Economics: European App Economy 2015 – Creating Jobs & Driving Economic Growth in Europe. https://www.slashdata.co/free-resources (2015), online; accessed 25 March 2019
28. Wanous, J.P., Reichers, A.E., Hudy, M.J.: Overall job satisfaction: how good are single-item measures? Journal of applied Psychology **82**(2), 247 (1997)
29. Woon, I.M., Kankanhalli, A.: Investigation of is professionals intention to practise secure development of applications. International Journal of Human-Computer Studies **65**(1), 29–41 (2007)
30. Zukowski, T., Brown, I.: Examining the influence of demographic factors on internet users' information privacy concerns. In: Proceedings of the 2007 annual research conference of the South African institute of computer scientists and information technologists on IT research in developing countries. pp. 197–204. ACM (2007)

A Questionnaire

- Participant information sheet and informed consent.
  - ○ I consent to begin the study
- How many years of experience do you have as a software developer?
  - ○ less than 2 years
  - ○ 2 to 4 years
  - ○ 5 or more
- Think about software you have developed. Likely it captures some kind of personal data. This can be, for example, data like names, addresses, identification numbers, location data, usage statistics, or technical data like IP addresses. Here are some statements about personal data of people who use software you develop. From the standpoint of your *role as a software developer*, please indicate the extent to which you **agree** or **disagree** with each statement.

  (a) It usually bothers me when the software I develop asks my users for personal information.

  Strongly disagree ○ ○ ○ ○ ○ ○ ○ Strongly agree

  (b) I sometimes think twice before asking my users for personal information with the software I develop.

  Strongly disagree ○ ○ ○ ○ ○ ○ ○ Strongly agree

  (c) It bothers me to collect personal information from so many users with the software I develop.

  Strongly disagree ○ ○ ○ ○ ○ ○ ○ Strongly agree

  (d) I'm concerned that the software I develop is collecting too much personal information about my users.

  Strongly disagree ○ ○ ○ ○ ○ ○ ○ Strongly agree

  (e) My users privacy is really a matter of their right to exercise control and autonomy over decisions about how their information is collected, used, and shared by the software I develop.

  Strongly disagree ○ ○ ○ ○ ○ ○ ○ Strongly agree

  (f) My users control of personal information collected by the software I develop lies at the heart of user privacy.

  Strongly disagree ○ ○ ○ ○ ○ ○ ○ Strongly agree

  (g) The software I develop should disclose the way the data are collected, processed, and used.

  Strongly disagree ○ ○ ○ ○ ○ ○ ○ Strongly agree

  (h) A good privacy policy for the software I develop should have a clear and conspicuous disclosure.

  Strongly disagree ○ ○ ○ ○ ○ ○ ○ Strongly agree

(i) It is very important to me that my users are aware and knowledgeable about how their personal information will be used.

Strongly disagree ○ ○ ○ ○ ○ ○ ○ Strongly agree

- Finally, here are some statements about how you consider the extent to which you, as a developer, deal with different aspects of personal data of your software's users. Please indicate the extent to which you **agree** or **disagree** with each statement.

(i) I properly deal with the extent to which my software collects data of its users

Strongly disagree ○ ○ ○ ○ ○ ○ ○ Strongly agree

(ii) I properly deal with the extent to which my software gives users control over their data

Strongly disagree ○ ○ ○ ○ ○ ○ ○ Strongly agree

(iii) I properly deal with the extent to which my software informs its users how their data is used

Strongly disagree ○ ○ ○ ○ ○ ○ ○ Strongly agree

B Detailed item adaption

Item	IUIPC	SDPA
(a)	It usually bothers me when online companies ask me for personal information.	It usually bothers me when the software I develop asks my users for personal information.
(b)	When online companies ask me for personal information, I sometimes think twice before providing it.	I sometimes think twice before asking my users for personal information with the software I develop.
(c)	It bothers me to give personal information to so many online companies.	It bothers me to collect personal information from so many users with the software I develop.
(d)	I'm concerned that online companies are collecting too much personal information about me.	I'm concerned that the software I develop is collecting too much personal information about my users.
(e)	Consumer online privacy is really a matter of consumer right to exercise control and autonomy over decisions about how their information is collected, used, and shared.	My user's privacy is really a matter of their right to exercise control and autonomy over decisions about how their information is collected, used, and shared by the software I develop.
(f)	Consumer control of personal information lies at the heart of consumer privacy.	My users' control of personal information collected by the software I develop lies at the heart of user privacy.
(g)	I believe that online privacy is invaded when control is lost or unwillingly reduced as a result of a marketing transaction.	I believe that my users' privacy is invaded when control is lost or unwillingly reduced as a result of a marketing transaction.
(h)	Companies seeking information online should disclose the way the data are collected, processed, and used.	The software I develop should disclose the way the data are collected, processed, and used.
(i)	A good consumer online privacy policy should have a clear and conspicuous disclosure.	A good privacy policy for the software I develop should have a clear and conspicuous disclosure.
(j)	It is very important to me that I am aware and knowledgeable about how my personal information will be used.	It is very important to me that my users are aware and knowledgeable about how their personal information will be used.