Age Matters – Privacy Concerns of Pokémon Go Players in Germany (Extended Abstract)

David Harborth and Sebastian Pape

Chair of Mobile Business and Multilateral Security
Goethe-University Frankfurt am Main

This paper investigates privacy concerns about organizational information privacy practices of active Pokémon Go (PG) players in Germany. PG [8] is a location-based augmented reality (AR) smartphone game developed by Niantic, a former Google owned company. Azuma et al. define AR in a way that [...] an AR system [...] combines real and virtual objects in a real environment; runs interactively, and in real time; and registers (aligns) real and virtual objects with each other [1, p.34].

The massive success of PG [17] makes it possible, to investigate AR based on large scale user studies for the first time. This allows researchers to assess important aspects, as acceptance factors, use cases and privacy and security perceptions of the users with respect to AR technologies. Understanding these heterogenous perceptions is necessary since many experts predict that AR is going to become one of the next big technological innovations with a massive market potential [2, 7]. Privacy aspects are especially important for the case of AR since the technology tends to become more pervasive with the advancements of wearable AR technologies (e.g. head-mounted displays). This leads to a situation where the user is continuously provided with context-sensitive information aout her environment [3]. This in turn makes it necessary for the system to continuously gather and process all kinds of data.

Privacy violations can be suffered by users - due to the increasing collection of several different data types [7] - as well as their environment. The case of the social environment could be observed in the past for Google Glasses with several reports about angry civilians who had the feeling of being filmed by the wearer and bars which prohibited entry when wearing the glasses. This partly led to the failure of the device in the consumer market [16].

Based on this observation, two research questions arise. First, are privacy concerns a relevant issue for Pokémon Go players and do they differ in magnitude between different groups of players? Second, is there a relationship between the privacy concern dimensions and the actual use behavior? After a brief theoretical introduction, these questions are addressed by two statistical analyses based on empirical data of 683 active PG players in Germany.

The variables representing privacy concerns with regard to organizational information privacy practices are collection, errors, unauthorized secondary use and improper access. As these variables cannot be measured directly (latent variables), they have to be operationalized in order to quantify the concerns via a user study. We chose the privacy constructs by Smith et al. [14], as they are widely tested with regard to validity and reliability (cf. Stewart and Segars [15]).

Table 1: Descriptive Statistics (N=683)

	Mean	Median	Standard Deviation	Min.	Max.
Age	34.539	32	(11.531)	18.000	66.000
Gender	0.572	1	(0.495)	0.000	1.000
Educational Degree	3.977	4	(1.207)	1.000	7.000
Smartphone Experience	5.958	6	(2.359)	0.000	10.000
Collection	5.349	5.5	(1.172)	1.000	7.000
Errors	5.355	5.5	(1.198)	1.000	7.000
Unauthorized Secondary Use	6.015	6.5	(1.071)	1.000	7.000
Improper Access	5.971	6.333	(1.090)	1.000	7.000
Frequency of Actual Use	5.517	5	(1.619)	1.000	10.000

The constructs are built by calculating mean sum scores of the single items belonging to the respective construct (cf. Appendix A). *Collection* is defined as the concern of people that too much data about them is collected over time. *Errors* represent users' concerns about inaccurate or false personal data in databases. *Unauthorized secondary use* measures the concern that personal data is used for another purpose than initially disclosed without the user's authorization. *Improper access* captures concerns about unauthorized people having access to the user's personal data [14, p. 172].
Since the goal of this work is to assess privacy concerns of PG players, the sample only consists of active players of the game. In order to ensure high quality of the sample, a certified sample provider (certified with ISO 26362 norm) was employed to get access to their online panel for Germany. The survey itself was conducted by the authors with LimeSurvey (version 2.63.1) [13]. The panel provider distributed the survey's link to 9338 participants until the aimed sample size of active players was reached. Of 9338 approached participants, 683 active PG players remained, excluding several participants who dropped out due to wrong answers to test questions and age restrictions (data was only collected for participants older than or equal to 18 years). Table 1 presents the summary statistics for the data set.

The results show that the median age is 32 years and that there is a larger share of women than men in the sample. Furthmore, the secondary school leaving certificate (equals the German "Realschulabschluss") and the A levels degree (equals the German "Abitur") are the most common educational degrees. With regard to these demographics, it can be argued that this data set represents the German population to an acceptable degree. The smartphone experience has a median of 6 years. The privacy constructs have all at least a median value of 5.5, implying that most players agree to the statements made in the constructs' items. The actual use frequency has a median of 5 which stands for playing "several times a week".
The Shapiro-Wilk test for normality indicates that the distributions of the variables collection, errors, unauthorized secondary use and improper access are not normally distributed while actual use is normally distributed.

Table 2: Two-sample Wilcoxon rank-sum and two-sample t test (for actual use)

	Group Variables				
Variables	Age_{Median}	Age_{DN}	Gender	Education	Smartphone Exp.
Collection	z=-3.217**	z=-3.288**	n.s.	n.s.	n.s.
Errors	z=-6.644***	z=-5.852***	n.s.	z=2.766**	n.s.
Un. Sec. Use	n.s.	z=-3.020**	z=-4.019***	n.s.	n.s.
Imp. Access	z=-2.516**	z=-3.870***	z=-2.589***	n.s.	n.s.
Actual Use	n.s.	n.s.	n.s.	n.s.	t=-4.335***

t statistic for t test and z statistics for Wilcoxon rank-sum test
* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

In the first part of the empirical assessment, it is investigated whether users' privacy concerns and the actual use behavior (i.e. frequency of playing PG) differ with regard to age, gender, education and smartphone experience. Therefore, categorial variables for group comparisons are created by dividing the scale of continuous variables into two meaningful groups. The threshold for age is not clearly determinable because there are two rationales for dividing the data which are highly interesting to investigate in the context of privacy. The first approach is a median split, a commonly used technique for forming categorial variables in statistics [6]. This has the advantage of comparing two groups, similar in size, based on the actual median of the used data set. This results in the groups of participants aged 31 and younger and participants aged 32 and older. The second approach deals with the notion of "digital natives" (DN) versus "digital immigrants" (DI) [11]. Since there is a vivid discussion on whether the notion of DN is substantial [5], it could be interesting in the context of privacy concerns to apply this threshold and investigate, whether it is true that DI are rather privacy sensitive and more concerned than the younger generation. A commonly named threshold for the oldest year of birth of a DN is 1980 [10]. The resulting two groups contain 446 entries for DN (37 years old and younger) and 237 entries for DI.
For experience, only the median split approach is applied since this is the most meaningful approach, with groups with smartphone experience less than or equal to 5 years and greater than or equal to 6 years.
Education is divided into a group with participants without university degree (NU), comprising all participants whose highest educational degree is the German Abitur (N= 487) and a group with 196 participants holding at least a Bachelor's degree (U). Table 2 summarizes the results for the statistical assessment of whether the group differences in mean values are statistically significant or not.

For collection there are only significant differences for the two different types of age groups. Users' perceptions of the errors construct differs between younger and older participants as well as between participants without and with university degree. Interestingly, the group comparison of unauthorized secondary use is different for the age groups. For the median split, there is no statistically significant difference in the evaluation between younger and older players, whereas there is one for the case of DN versus DI. Furthermore, the evaluation for this construct

Table 3: Regression analysis of privacy concern variables and use behavior (N=683)

	(1)	(2)	(3)	(4)	(5)
	Dependent variable: actual use behavior				
Collection	0.0518				-0.00691
	(1.00)				(-0.11)
Errors		-0.0480			-0.170**
		(-0.91)			(-2.76)
Un. Sec. Use			0.124*		0.00368
			(2.05)		(0.03)
Imp. Access				0.149*	0.251*
				(2.53)	(2.19)
_cons	5.240***	5.774***	4.772***	4.630***	4.943***
	(18.70)	(20.00)	(12.85)	(13.01)	(12.78)

t statistics in parentheses
* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

differs significantly between women and men. For the improper access construct, statistically significant differences are prevalent for age as well as gender groups. The actual use variables is very homogenous across the different characteristics of PG players. The only significant difference can be observed between players with less and more smartphone experience. The differences between the mean values of the privacy concern dimensions are illustrated in Appendix B.

The question of whether privacy concerns influence the use behavior in general is addressed in the second part of the empirical analysis. Due to the breakdown into four variables, it is possible to assess which kind of privacy concern exerts what kind of influence on use behavior. This question is addressed with a two-stage process (cf. Table 3). First, each privacy concern variable is treated as the independent variable in a simple linear regression model, with actual use behavior as the dependent variable. Second, a multiple regression model, containing all independent variables, is calculated in order to assess the effect of the different dimensions of privacy concerns on use behavior simultaneously.

The results of the regression analysis indicate that privacy concerns have no significant impact on the actual use behavior. Although there are statistically significant relationships, the effect sizes are rather small and therefore not relevant. Thus, although privacy is perceived as important, it does not affect the use of PG. The game requires several more types of data compared to other smartphone applications due to its location-based and AR nature. In addition, not all processes with regard to data handling and processing are clearly stated in the privacy policies [4]. This contrary result could be a case of the privacy paradox [9], where people state that privacy is important for them but act in the opposite way.

This research on organizational information privacy practices with respect to the location-based AR smartphone game Pokémon Go shows that there are significant differences in the different dimensions of privacy concerns between younger and older players. Furthermore, gender matters for two of the four dimensions. An additional regression analysis with the actual use frequency revealed no clear

impact of privacy concerns on the former, indicating that players might well be aware of privacy dangers and are concerned about it, but do not act accordingly.

References

1. Ronald T. Azuma, Yohan Baillot, Steven Feiner, Simon Julier, Reinhold Behringer, and Blair Macintyre. Recent Advances in Augmented Reality. In *IEEE Computer Graphics And Applications*, number November/December, pages 34–47, 2001.
2. Heather Bellini, Wei Chen, Masaru Sugiyama, Marcus Shin, Shateel Alam, and Daiki Takayama. Virtual & Augmented Reality: Understanding the race for the next computing platform. Technical report, Goldman Sachs Equity Research, 2016.
3. Jens Grubert, Tobias Langlotz, Stefanie Zollmann, and Holger Regenbrecht. Towards Pervasive Augmented Reality: Context-Awareness in Augmented Reality. *IEEE Transactions on Visualization and Computer Graphics*, PP(99):1–20, 2016.
4. Josh Hafner. While you track Pokémon, Pokémon Go tracks you, 2016.
5. Ellen Johanna Helsper and Rebecca Eynon. Digital natives: where is the evidence? *British Educational Research Journal*, 36(3):503–520, 2010.
6. Jörg Henseler and Georg Fassot. Testing Moderating Effects in PLS Path Models: An Illustration of Available Procedures. In Vincenzo Esposito Vinzi, Wynne W. Chin, Jörg Henseler, and Huiwen Wang, editors, *Handbook of Partial Least Squares*, chapter 30, pages 713–735. Springer, 2010.
7. Paul Hyman. Augmented-Reality Glasses Bring Cloud Security Into Sharp Focus. *Comm. of the ACM*, 56(6):18–20, 2013.
8. Niantic Labs. Official website for the game pokémon go. `http://www.pokemongo.com/`.
9. Patricia A. Norberg, Daniel R. Horne, and David A. Horne. The Privacy Paradox: Personal Information Disclosure Intentions versus Behaviors. *Journal of Consumer Affairs*, 41(1):100–126, jun 2007.
10. John Palfrey and Urs Gasser. *Born Digital: Understanding the First Generation of Digital Natives*, volume 13. Basic Books, 2008.
11. Marc Prensky. Digital Natives, Digital Immigrants. *On the Horizon*, 9(5):1–6, 2001.
12. L.D. Rosen, K. Whaling, L.M. Carrier, N.A. Cheever, and J. Rokkum. The Media and Technology Usage and Attitudes Scale: An empirical investigation. *Comput Human Behav.*, 29(6):2501–2511, 2013.
13. Carsten Schmitz. LimeSurvey Project Team, 2015.
14. H. Jeff Smith, Sandra J. Milberg, and Sandra J. Burke. Information Privacy: Measuring Individuals' Concerns About Organizational Practices. *MIS Quarterly*, 20(2):167–196, 1996.
15. Kathy A. Stewart and Albert H. Segars. An Empirical Examination of the Concern for Information Privacy Instrument. *Information Systems Research*, 13(1):36–49, 2002.
16. Justin Burton Weidner. How & Why Google Glass Failed, 2015.
17. Howard Yu. http://fortune.com/2016/07/23/pokemon-go-augmented-reality/, 2016.

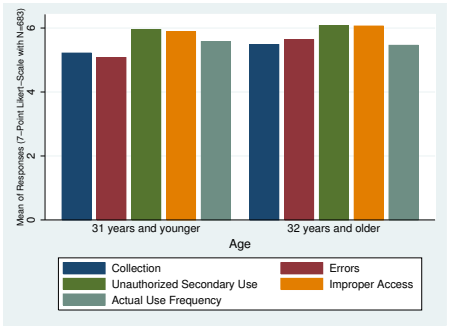
A Questionnaire

Collection
Coll1. It usually bothers me when companies ask me for personal information.
Coll2. When companies ask me for personal information, I sometimes think twice before providing it.
Coll3. It bothers me to give personal information to so many companies.
Coll4. I am concerned that companies are collecting too much personal information about me.
Errors
Err1. All the personal information in computer databases should be double-checked for accuracy – no matter how much this costs.
Err2. Companies should take more steps to make sure that the personal information in their files is accurate.
Err3. Companies should have better procedures to correct errors in personal information.
Err4. Companies should devote more time and effort to verifying the accuracy of the personal information in their databases.
Unauthorized Secondary Use
USU1. Companies should not use personal information for any purposes unless it has been authorized by the individuals who provided the information.
USU2. When people give personal information to a company for some reason, the company should never use the information for any other reason.
USU3. Companies should never sell the personal information in their computer databases to other companies.
USU4. Companies should never share personal information with other companies unless it has been authorized by the individuals who provided the information.
Improper Access
IA1. Companies should devote more time and effort to preventing unauthorized access to personal information.
IA2. Computer databases that contain personal information should be protected from unauthorized access – no matter how much it costs.
IA3. Companies should take more steps to make sure that unauthorized people cannot access personal information in their computers.
Use Behavior
Please choose your usage frequency for Pokémon Go:

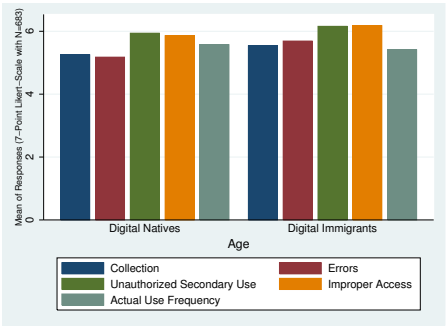
- Never
- Once a month
- Several times a month
- Once a week
- Several times a week
- Once a day
- Several times a day
- Once an hour
- Several times an hour
- All the time

The frequency scale is adapted from [12]. All other items are measured with a seven-point Likert scale, ranging from "strongly disagree" to "strongly agree".

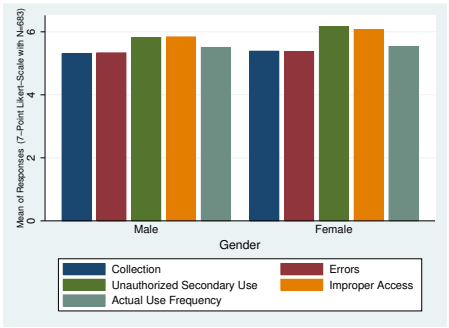
B Differences in Mean Values of Privacy Concerns Between Different Groups



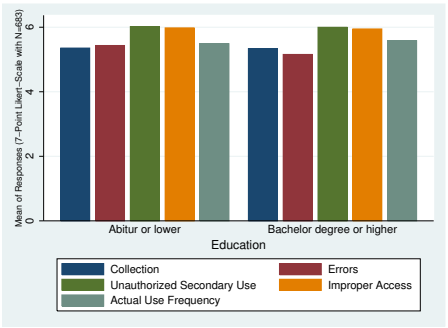
(a) Age_{Median}



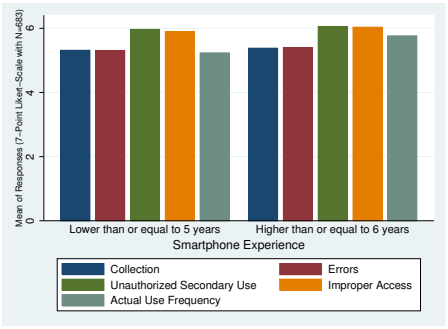
(b) Age_{DN}



(c) Gender



(d) Education



(e) Smartphone Experience

Fig. 1: Differences in mean values between the assigned groups of age, gender, education and smartphone experience for the privacy concern variables and actual use