Users' Willingness to Share Data on the Internet: Perceived Benefits and Caveats

Martina Ziefle, Julian Halbey and Sylvia Kowalewski
Human-Computer Interaction Center, RWTH Aachen University, Campus-Boulevard 57, 52074, Aachen, Germany

Keywords: Internet, Social Network Sites, Perceived Privacy, Willingness to Share Data, Trust, Gender, Conjoint Study.

Abstract: One of the major challenges of the ongoing digitalization and the ubiquitous usage of pervasive computing in all fields of our lives is to steer a sensible balance between benefits and drawbacks of using the Internet and to implement an appropriate data handling when using digital media. The broad availability of data, in line with the enormous velocity of information retrieval, is open to abuse and malpractice, with privacy threats as the most serious barrier. The consumers and their attitudes and behaviors when using the Internet play an important role in the discussion about privacy protection. The aim of the current study was to analyze Internet usage behaviors and users' willingness to share their data when using digital services and social network sites. In a two step empirical approach, we first explore users' perceptions of privacy in the context of Internet usage and social network sites by means of a focus group approach. In a second step, a quantitative study was carried out. Using a conjoint measurement approach, user scenarios were created from combinations of different levels of anonymization extent, data type, and benefits from sharing the data. The respondents' task was to decide under which conditions they would be willing to share their data. 80 volunteers (50,6% women) between 14 and 60 years of age participated in the conjoint study.

1 INTRODUCTION

Historically, never might there have been a bigger technological challenge for democratic societies than the sensible, responsible, and open-minded handling of Big Data that goes hand in hand with the tremendous chances and drawbacks of the Internet of Things (Karabey, 2012, Katal et al., 2013). In the IDC's sixth annual study of the digital universe (Gantz and Reinsel, 2012), a comprehensive overview of the Big Data challenge is outlined. Using a longitudinal analysis approach (comprising data collection between 2005 until 2020), the authors provide a detailed report about the reach of digital data, their quantity and the significant growth of data over the respective time. Quoting from the executive summary, analyses up to 2012 showed:

"From 2005 to 2020, the digital universe will grow by a factor of 300, from 130 exabytes to 40,000 exabytes, or 40 trillion gigabytes (more than 5,200 gigabytes for every man, woman, and child in 2020). From now until 2020, the digital universe will about double every two years. […] Between 2012 and 2020, emerging markets' share of the expanding digital universe will grow from 36% to 62%. A majority of the information in the digital universe, 68% in 2012, is created and consumed by consumers -watching digital TV, interacting with social media, sending camera phone images and videos between devices and around the Internet, and so on. Yet enterprises have liability or response-bility for nearly 80% of the information in the digital universe. They deal with issues of copyright, privacy, and compliance with regulations even when the data zipping through their networks and server farms is created and consumed by consumers. […] The amount of information individuals create themselves — writing documents, taking pictures, down-loading music, etc. — is far less than the amount of information being created about them in the digital universe" (ibid, pp. 1-2).

From this impressing and at the same time alarming data report, three major points should be noted:

(1) The ongoing digitalization in all fields of our life enables huge benefits for consumers on different levels – on a private and a professional as well as on a societal level.

(2) The availability of data and information opens up a significant knowledge gain in – for societies –

Ziefle, M., Halbey, J. and Kowalewski, S.
Users' Willingness to Share Data on the Internet: Perceived Benefits and Caveats.
DOI: 10.5220/0005897402550265
In *Proceedings of the International Conference on Internet of Things and Big Data (IoTBD 2016)*, pages 255-265
ISBN: 978-989-758-183-0

very important fields: medical and health issues, mobility and transport, production and business, learning and education, to just mention a few.

(3) Naturally, the broad availability of data, the measurability of information in important branches, and the enormous velocity of information gathering lends itself to abuse and malpractice. The damage is – though not exactly calculable – huge and can happen on different levels (privately, professionally, societally). Among the most prominent drawbacks, the danger of privacy loss by eaves-dropping as well as shoulder reading, technologies come to the fore. Consumer behaviors are tracked while using social network sites and while surfing the Internet (Takabi et al., 2010; Szongott et al., 2012).

For private consumers, but also for the professional context, it is thus important to find a sensible balance between the benefits and drawbacks of using the Internet and establish an appropriate data handling when using digital media. This balance can only be reached by a broad understanding of Internet usage behaviors and a joint approach from different disciplines (Dartmann et al., under revision). In order to achieve a far-reaching privacy preservation strategy, different research and policy approaches can be observed. While technical disciplines work mostly on the development of privacy enhancing technologies (k-anonymity or differential privacy technologies, e.g., Sweeny, 2002; Dwork, 2006, Dritsas et al., 2006), there is, naturally, a prominent research input from the field of law and legal regulations (Mayeda et al., 2016, Trestenjak, 2016), as well as approaches to support privacy protection behaviors in ecommerce and digital services from an economic and market perspective (Phelps et al., 2001; Matsusaki, 2016). Increasingly, normative studies dealing with the establishment of a digital etiquette and guidelines for safe and conscious digital behavior are pursued. Already in 2004, Ribble, Bailey and Ross (2004) claimed that across all of society there is a need to establish a mindset of technology appropriate behaviors in the context of technology education, referred to as Digital Citizenship.

“Digital citizenship can be defined as the norms of behavior with regard to technology use. As a way of understanding the complexity of digital citizenship and the issues of technology use, abuse, and misuse, we have identified nine general areas of behavior that make up digital citizenship” (Ribble et al., 2004, p. 7).

In their work, the authors claim different usage policy behaviors that should to be respected by teachers and students when interacting with technology (Ribble et al., 2004). Recent work strengthens this view (Hornung and Schnabel, 2009; Rouvroy and Poullet, 2009; Young and Quan-Haase, 2013). It is claimed that media competency as well as knowledge about potential dangers and caveats in the context of Internet usage and pervasive computing have to be globally developed. The responsibility, though, is not limited to policy, law, and education entities, but it is shifted to end users and consumers, referred to as the individual responsibility of control of own personal data and informational self-determination (Rouvroy and Poullet, 2009; Tene and Polonetsky, 2012). Then, individual consumers must take care of their self-protection when using ecommerce services and digital media, and – beyond technical, legal, and policy efforts – consumers have the duty to be responsible for their electronic actions and deeds.

Beyond this warranted claim for the establishment of such digital ethics and the broad public awareness that the responsibility for own digital behaviors is a necessity (Marx, 1998), still, from a social science point of view, there are some concerns, if not doubts, that this mere (normative) and rational claim is effective in the end (Lauffer and Wolff, 1977; Kalwar, 2008). The concerns are based on empirical findings which corroborate that users’ behaviors and their mental mindset or rationality diverge widely.

Privacy Paradox. Users – though being seriously concerned when asked about the fear of losing their privacy in the digital universe – nevertheless do mostly not protect themselves and are neither discreet nor careful with their personal data in the Internet (Lahlou 2008; Kowalewski et al., 2015; Boyd and Hargittai, 2010). This phenomenon is called privacy paradox (Awad et al., 2006; Norberg et al., 2007).

Perceived Privacy. The factual risk of privacy loss is not identical with the *perceived* risk of privacy loss and the perception of control (Spiekermann, 2007). Users might have not only a different understanding of dangers and caveats and a different perception of control but also a different appreciation of the temporary benefits of using social network sites. Thus, users might decide to take the risk of data sharing as they perceive to be in control. Likewise, users might decide to share data because the temporary benefit is higher for them than the potential risk.

Context. The perceived benefits, which might motivate users to share their data on the Internet, and the perceived risks, which might prevent users from sharing their data, may also be impacted by the respective usage context (Nissenbaum 2011; Kowalweski et al, 2015). Recent work showed that the tradeoff between benefits and barriers are quite different in a medical context (when it comes to intimate data concerning personal health and illness, Wilkowska and Ziefle, 2012) in comparison to the sports context in which users might be even keen to share data (van Heek et al., 2014). Also, the closer digital data are related to personal homes, the more observant and reluctant are consumers to allow others to use and see their personal data (Ziefle et al., 2011; van Heek et al., 2015; 2016).

Culture. Understanding digital behaviors and privacy concerns requires also the understanding of the respective situation and culture (Hargittai, 2007). Krasnova and Veltri (2010) examined privacy concerns of Facebook members in the USA and Germany. The findings revealed culture-specific differences: While German users expect more negative consequences (identity loss, damage, and privacy-related violations), American users – though also aware of cyber crime and privacy violations – still do report higher trust in the service provider and higher levels of perceived control. Thus, the perception of privacy varies between different cultures.

User Diversity. User diversity is another important factor with respect to Internet behaviors. Here, several cognitive and affective factors might be relevant: One is the users' (low) level of knowledge about factual risks and malpractice (Kowalewski et al., 2015). A second factor relates to different personalities and interacting styles (those who are willing to take risks vs. those who are more fearful, e.g., Karim et al., 2009). A third factor relates to the competency in using technical devices and digital media. The digital competency shapes the way users interact with the Internet (Akther, 2014) and is indirectly related to age effects (as persons with a different upbringing with technology do have another mental model of how technology works, e.g., Fogel and Nehmed, 2009; Freestone and Mitchell, 2004).

Gender. The usage of digital services and social network sites is gender-sensitive (Kennedy et al., 2003; Kim et al., 2007). The same applies for privacy concerns when using digital services which are more pronounced in women (Fogel and Nehmad, 200), Wilkowska et al., 2010). On the one hand, there is extensive research evidence that women – in contrast to men – report a lower self-efficacy when using digital devices which makes them much more careful in the interaction with technology (Durndell and Haag, 2002). In addition, frequent social network site users show greater risk taking attitudes (Fogal and Nehmad, 2009), and are usually male. In line with this, women report higher privacy concerns and fears of being victims of cyber crime (Halder and Jaishankar, 2011). On the other hand, women outperform males in terms of electronic communication with peers on social network sites. Women show a higher emotional involvement and social engagement in digital communication (Sun et al., 2016), accompanied by highly positive collective self-esteem and motivation to befriend peers and stay in contact with them (Barker, 2009; Thelwall et al., 2008; 2010).

2 QUESTIONS ADDRESSED AND LOGIC OF PROCEDURE

In this study, we take a social science perspective and argue that an effective policy of privacy protection behaviors cannot be effective and sustainable before we understand the behaviors of Internet users by means of empirical evidence. Therefore, a two-step empirical procedure was undertaken. In a first step, focus groups were carried out in which participants discussed the perceived benefits and caveats of sharing their data as well as their general experience with the Internet and the usage of social network sites. On the base of these argumentation lines, a quantitative conjoint study was carried out in order to study respondents' privacy preferences and their willingness to share their data.

Characteristically for the conjoint measurement approach is that selected attributes (probability of being identified (extent of anonymization), the data type and benefits from sharing the data) are combined to different usage scenarios for which participants had to decide if they would be willing to share their data under the respective conditions. This procedure allows us to empirically describe which of the factors is relevant to which extent and in what way might other conditions modulate the willingness to share their data.

As gender was revealed to be a decisive factor for Internet behavior, a comparison between female and male users was focused at.

The following questions guided this research.

(1) What are the most important factors for Internet usage and willingness to share data?
(2) What is the worst case scenario under which participants would not share their data at all and which is the best case scenario?

(3) Are there gender differences in the preferences and the decision scenarios?

3 METHODOLOGY

We followed a two step empirical procedure.

Focus Group Prior to the main study, focus groups were carried out in which participants (N = 8, 20-50 years of age) discussed the most important issues with respect to privacy and Internet usage. The participants' argumentation lines were analyzed in detail (not reported here). The factors that have been evaluated as most important (benefits and barriers) by participants were taken as basis for the selection of attributes used in the conjoint analysis.

Conjoint Analysis A choice-based conjoint analysis approach was selected, as it mimics the complex decision processes in real world scenarios in which users have to evaluate more than one attribute that influences the final decision (Luke and Tukey, 1964). Contrary to traditional surveys, in which participants answer single factors separately from each other, conjoint analyses simulate real-world user decisions in which users weigh potential benefits against perceived barriers. In the context here, the tradeoff between keeping one's own privacy vs. sharing data on the Internet was experimentally studied.

Methodologically, the given decision scenarios and tradeoffs consist of multiple attributes and differ from each other in the attribute levels. As a result, the relative importance of attributes deliver information about which attribute influences the respondents' choice the most. Part-worth utilities reflect which attribute level is valued the highest.

3.1 Questionnaire

First, the questionnaire structure is described as are the instructions which were given to respondents, followed by the selection and description of the attributes used for the decision scenarios.

3.1.1 Structure

The questionnaire was arranged in five sections.
The *first section* addressed demographic characteristics of the participants. Also, we asked if participants were familiar with, respectively aware of, the importance of privacy in the context of Internet usage. Answers could be given on a four-point scale, ranging from 1 = "the topic is quite novel to me" to 4 = "I am familiar with the topic."

In the *second part*, the experience of Internet usage was assessed, asking for the frequency of (1) shopping in the Internet, (2) using social networks sites, (3) searching for information for leisure activities, and (4) asking for price comparisons on the Internet. The frequency could be stated within the following graduations: several times daily, daily, several times a week, lesser than that, and never.

The *third part* introduced the topic of the study, followed by a detailed description of the single factors out of which the scenarios were formed. It was important that the respondents understood the reason for the study and the decisions that would be presented later on. Therefore, we gave respondents the following general introduction:

> *Internet users have a right to decide what is going to happen with their data. Principally, the society as a whole and every single individual can profit from the data mass generated on the Internet. What is important is an approach that satisfies the interests of all parties concerned. Here, privacy preserving technologies can step in as they anonymize data and thereby detach them from one's person. However, this procedure also reduces the usability of that data as one cannot, for example, link a gender to a person anymore. Thus, a complete anony-mization might not be reasonable in every case.*
>
> *The study aim is to find a solution that adheres to the interests of the data owners (i.e., you as internet user) and the ones utilizing said data.*

In the *fourth part* of the questionnaire, the single factors were introduced and explained in detail (see section 3.1.2.).

The *fifth and last part* finally presented twelve scenarios generated from combinations of the single factors. Data was collected in an online survey conducted in Germany in 2015. Completing the questionnaire took about 20 minutes.

3.1.2 Selection of Attributes

The selection of attributes was based on focus group study outcomes. Taken from the argumentation patterns, we assume that the preferences concerning data sharing on the Internet are influenced by different characteristics that have the highest utility for users.

For the conjoint study here, we selected the most important attributes raised by participants:

- the data type
- the benefit of sharing the data

- the extent to which anonymization is guaranteed.

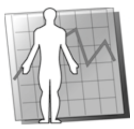
Type of Data	
	Shopping Preferences that are based on previous online purchases.
	Leisure-Time Interests: In this case, you would share hobbies and other interests – such as music, movies, books, or events – by agreeing that the contents of websites visited by you are analyzed.
	Location Data: In this case, you would permanently share your location when using your smartphone, cellphone, or computer.
	*Lifestyle Habits:* You would actively use apps meant to, e.g., record your level of fitness or driving habits. The apps would chronicle and analyze data that is important to you but also to health insurance funds or car insurance companies.

Figure 1: **"**Attribute: type of data.".

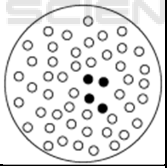
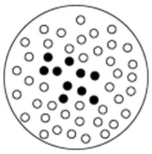
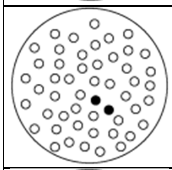
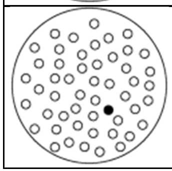
Anonymization	
	The same query (90-year-old man from a certain location) suddenly generates four equal results and individual persons are no longer clearly distinguishable. The 90-year-old is anonymous in a group of four, or, in other words, the *probability of identifying* him lies *at 25%.*
	There is also the possibility to adapt characteristics like age or gender in such a way that there are at least 10 persons with the same characteristics. This puts the *probability of identification at 10%.*
	In a group of two persons, the probability of clear *identification is 50%.*
	Without anonymization, the likelihood of *identification is 100%.*

Figure 2: Attribute "Extent of anonymization".

The first attribute that has been varied is <u>*the type of data*</u>. Respondents were instructed as follows:

You can decide on the type of data you are willing to share. This is information you can distribute in general on the Internet. We differentiate four types of data (that, to date, are already recorded without the users' knowledge:

Types of Benefits	
	Commendations: You periodically receive commendations via email or social networks. For example, (1) the cheapest offer of a product you have searched for or looked at on the internet several times in the recent past, (2) an insurance that is tailored to your lifestyle habits and offers better conditions than your current insurance package, or (3) events or leisure time activities in your proximity.
	Discounts: You periodically receive coupons via email. For example, for (1) online-shops you frequent, (2) leisure-time activities in your proximity (cinema, public swimming pool, etc.), or (3) insurance offers that are tailored to your personal lifestyle habits.
	Global benefits: Your shared data aid the public at large and not only singular big corporations. For example, researchers could use your data for research purposes that benefit society or medicine, cultural institutions or clubs could advertize in a more targeted manner, or software developers could utilize the user data to develop free software.

Figure 3: Attribute "Benefits".

The second factor was the <u>*extent of anonymization*</u>. The following instruction was given to participants in order to clarify what was meant by the single anonymization levels:

> *You can decide on the probability that your data can be linked to your person. Hypothetically, general data about people is collected. One could then assume this data is anonymous if a person's name is not recorded as well. How-ever, there are further characteristics such as age or zip code that enable an identification without access to one's*

name. Anonymization technologies prevent this by slightly modifying the data. Imagine, e.g., a village that has two 86-year-olds, one 87-year-old, and a 92-year-old. Instead of recording the precise age, only an age-group, e.g., aged 85 to 95, would be stored.

The third factor was the benefit from sharing data. It was of specific importance that benefits do reflect different persepctives (from local to global), as these dimensions were mentioned in the discussions during the focus groups.

The attribute "benefits" was introduced simply by the question: As an Internet user, how do I benefit from this?

3.1.3 Exemplary Decision Scenario

A combination of all corresponding levels would have led to 48 (4x4x3) possible combinations. As those decisions are quite taxing on the participants, we reduced the number of choice tasks to 12 random tasks. A test of design efficiency confirmed that the reduced test design was comparable to the hypothetical orthogonal design.

In each of the choice-based-conjoint decision task four sets of scenario configurations were presented. No restrictions were put on the level combinations, because all chosen attribute levels were combinable. Overall, participants had to evaluate twelve choice tasks, each consisting of three different combinations of the attributes *types of data*, *extent of anonymization*, and the types of *benefits*. Also, a "none" option was available in case that none of the scenarios seemed appropriate to respondents. Participants were instructed to select the senario they preferred the most. There was no possibility to skip tasks. In order to improve comprehensibility, attribute levels were presented by pictrogramms in addition to written information.
An exemplary decision scenario is given in Figure 4.

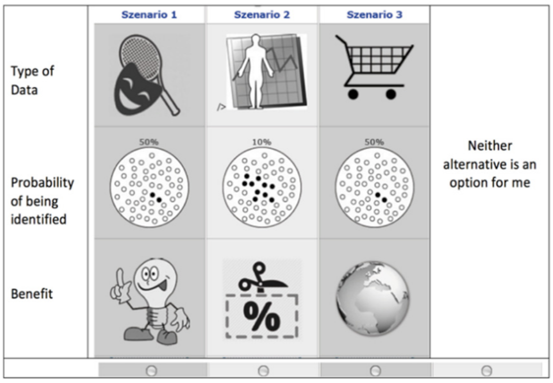

Szenario 1
Szenario 2
Szenario 3
Type of Data
Probability of being identified
Benefit
50%
10%
50%
%
Neither alternative is an option for me
○
○
○
○

Figure 4: Example decision scenario.

3.2 Sample

Overall, the data of 80 participants was analyzed. Age ranged between 14 years (youngest participant) and 60 years (oldest participant), with a mean age of 32.9. The sample was quite gender-balanced, with 50.6 % women.

Asked about prior knowledge about privacy and Internet usage, the sample reported to be quite aware of the topic (M=3.1/out of four points max). However, a significant gender difference was found: male respondents (M=3.5) reported to be more aware of the topic than females (M = 2.9).

With respect to the frequency of using the Internet, participants reported to use social network sites on a daily basis (M = (2.2) and seek information about leisure time activities several times a week (M = 3.2), whereas online shopping and price comparisons via Internet are accomplished less often (M = 3.4). Outcomes are depicted in Figure 5.

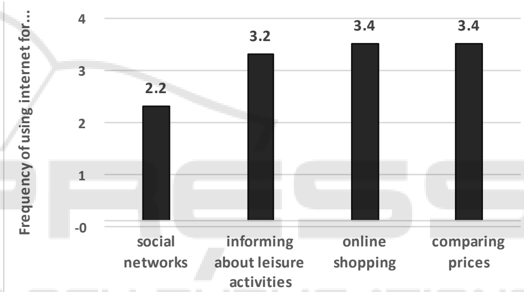


Figure 5: Relative importance of attributes.

4 RESULTS

The data analysis (estimation of part-worth utilities) was done with the Sawtooth Software (SSI Web, HB, SMRT). In order to identify the main impact factors on users' preferences to share their data on the Internet, we calculated the relative importance of each attribute. Then, part-worth utilities were analyzed (on the basis of Hierarchical Bayes) as to understand which of the three factors is the most relevant attribute across all decisions made and relative to all other attributes. When interpreting part-worth-utilities, it should be noted that these are data that are scaled to an arbitrary additive constant within each attribute. Thus, it is impossible to compare utility values between attributes. Comparisons of differences between attribute levels are possible if using zero-centered differentials (part-worth utilities that are scaled to sum to zero within each attribute).

4.1 Relative Importance

The relative importance scores of attributes on preferences to share data in the Internet are depicted in Figure 6.. As can be seen, the most important attribute is the probability of being identified (48.9%), followed by the type of data (30.8%). The attribute which was evaluated as least important on users' preferences was the benefit of sharing the data reaching a score of 20.3%.

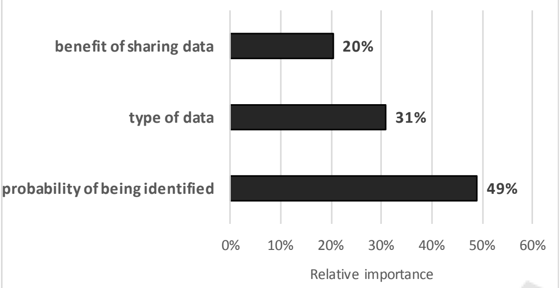


Figure 6: Relative importance of attributes.

The relative importance scores show that users evaluate the probability of being identified as the most important attribute whenever they think about sharing their data. Also, the type of data is relevant for them. Interestingly, the benefits they could gain from sharing the data is evaluated as least important, compared to the other attributes under study. At this point, it may be noteworthy that the ranking of importance and even the single scores are quite homogeneous across participants, neither impacted by gender nor by the self-reported familiarity with the topic ("how familiar are you with the topic privacy in the Internet").

4.2 Part-worth Utilities

The average zero-centered diff part-worth utilities for all attribute levels are shown in Figure 7. From this depiction it becomes obvious that the attribute "*probability of being identified*" yielded the highest range between part-worth utilities. When looking at absolute utility values, the level "probability of 10% of being identifiable" reached the highest utility score (76.7), followed by the 25% probability, which was still acceptable (reaching a positive value of 23.3). In contrast, the 50% probability was not acceptable for respondents (yielding a negative value of -36.2). The attribute level which received the lowest utility value was a 100% probability of being identifiable (-63.8).

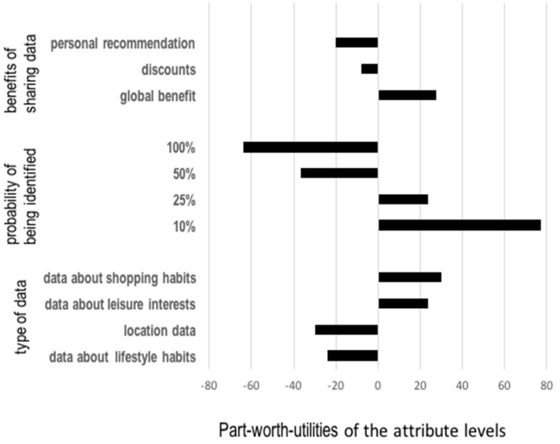


Figure 7: Part-worth utilities (zero-centered diffs) for all attributes and levels in the choice-based conjoint study.

When focusing on the type of *benefits*, also a mixed picture of acceptance was found. If data sharing leads to a global benefit for the public at large (e.g., research efforts, free software development, medical health or cultural issues), respondents would be willing to share their data (positive utility score of 27.4).

Negative utility scores were revealed for general discounts that are periodically offered by email (-7.5). The lowest utility scores (-19.9) were received by recommendations that specifically address personal habits (insurance, lifestyle tailored offers, leisure time activities in respondents' proximity).

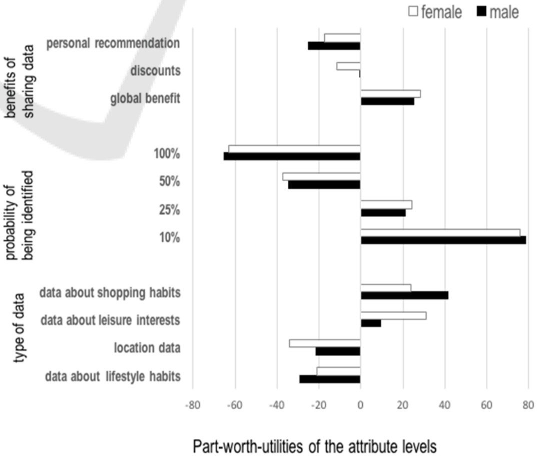


Figure 8: Part-worth utilities (zero-centered diffs) for all attributes and levels in the choice-based conjoint study from a gender perspective.

Looking at the type of data, "shopping habits" and "leisure time activities" might be shared from the respondents' point of view (shopping habit: 29.9;

leisure time data: 23.3). However, sharing of "location data" and lifestyle habits" receive negative scores (lifestyle habits: -23.8; location data: -29.5).

Finally, we analyzed gender differences with respect to part-worth utilities (Figure 8). Single values can be taken from Table 1.

Table 1: Part-worth utilities in both gender groups.

		Men	Women
benefit of sharing data	global benefit	25.4	28.5
	discounts	-0.3	-11.4
	personal recommendation	-25.3	-17.4
type of data	data about lifestyle habits	-29.1	-20.9
	location data	-21.3	-33.9
	data about leisure interests	9.2	30.9
	data about shopping habits	41.3	23.8
probability of being identified	10%	78.6	75.8
	25%	21.2	24.3
	50%	-34.4	-37.1
	100%	-65.3	-62.9

While the preferences with respect to the main attribute "probability of being identifiable" is the very same for women and men, there still were gender differences regarding the evaluation of the benefits of sharing data and the type of data users would be willing to share.

With respect to the benefits, personal recommenda-tions are more negatively evaluated by men (-25.3) than women (-17.4). Women evaluate discounts as more negative (-11.4) compared to men (-0.3), who are neutral in this regard. When looking at the type of data, men refuse to share data about life style habits (-29.1) more strongly than women (-20.9). Women, in contrast, are much less willing to share location data (-33.9) compared to men (-21.3). Sharing data about leisure time activities is more attractive to women (30.9), while men agree to share data with regard to leisure activities to lesser extent (9.2). Another difference between both gender groups regards shopping data, which male respondents would be much more willing to share (41.3) than women (23.8).

5 DISCUSSION AND FUTURE RESEARCH

This study revealed insights into users' preferences for sharing their data on the Internet. In order to understand behaviors and the tradeoffs between perceived benefits (that motivate users to share data) and perceived barriers (that motivate users not to share the data), we simulated usage scenarios in which different attribute combinations were experimentally varied. As attributes we explored three major factors, which were discussed as the most important factors in previously conducted focus group. The first factor was the data type (shopping preferences, leisure time interests, location data, and lifestyle habits), referring to the fact that users might have different privacy protection needs depending on the respective context. The second factor referred to the type of benefit which is to be expected from sharing data on the Internet. There were three different benefits, ranging from recommendations (information about interesting events, offers) over discounts (specifically tailored offers and bargains) to global benefits (data sharing helps the public at large, not only single company interests). Finally, the third factor dealt with the probability of being identified, which was also experimentally varied.

When looking at the single factors, the findings show that the anonymization extent is, in fact, the most relevant attribute for respondents, hinting at a high awareness of the risk of eavesdropping. The type of data was ranked as second most important criterion. Users' willingness to share data depends on the respective usage context, conforming Nissenbaum's theory of contextual integrity according to which users control their privacy needs depending on the respective context (Nissenbaum, 2014). Shopping habits and leisure time activities might be shared from the respondents' point of view, in contrast to lifestyle habits and location data which are perceived as too personal to be shared.

The lowest overall importance received the type of benefits. However, there were still differences between the respective types of benefits. Users' willingness to share data increases when the data sharing results in a global benefit for the public at large (e.g., research efforts, free software development, medical health or cultural issues). General discounts that are periodically offered by email were negatively evaluated. The same applies for specifically tailored recommendations (insurance, lifestyle tailored offers), because they rely on quite intimate data (personal habits). Thus, the more specific and personal the benefit is that is offered in return of sharing the data, the more negative is the respondents' attitude and the lower is the willingness to share data.

In conclusion, on this data basis, a "best case" and a "worst case" scenario of users' willingness to share data can be derived. As taken from the highest utility ratings for each attribute, the most accepted scenario for sharing their shopping or leisure activities data is a probability of 10% of being identifiable and a global benefit of data usage for the public at large (in contrast to the exclusive benefits for companies). The worst scenario is a 100% identification probability in combination with location data or even the identification of lifestyle habits that are used for an individually tailored personal recommendation.

Another research focus was directed to gender differences and the question if women and men have a similar perspective on the willingness to share their data on the Internet. The findings of the study showed both gender-sensitive and gender-insensitive findings. Men and women do have the same attribute importance ranking (anonymization is most important, followed by type of data, and the type of benefit as least important criterion). They also apply the same decision criterion for or against sharing the data. Thus, the above mentioned worst case and best case scenarios are valid for both gender groups.

However, gender differences showed with respect to the evaluation of the benefits of sharing data and the type of data users would be willing to share. In regard to the type of benefits, women evaluate personal recommendations as more positive and discounts as more negative than men. Sharing location data is a much stronger no-go for women than for men. Also, women are much more reluctant to share data regarding their shopping habits, in contrast to men who are more open in this regard. On the other hand, when it comes to the question of sharing leisure time activities on the Internet, women would share their data much easier than men would.

Even though our empirical research approach provided valuable insights into conditions of users' willingness to share their data, still, the outcomes here provide only a first glimpse into a highly complex phenomenon. Future studies will have to continue in this line of research, considering methodological limitations and broadening the research focus (Dartmann et al., under revision).

A first point in this context regards the sample size of this study. The findings should be replicated in larger and more representative samples.

In this context, integrating a larger portion of participants of higher and younger age could be insightful in order to learn if the perceived benefit of sharing data might be age-sensitive. Also, we only used three attributes for the conjoint study. Even though the attributes were empirically identified as most important for Internet users, of course there are further factors that should be integrated in future studies to receive a full picture.

Furthermore, the findings do reflect only empirical insights from one country - Germany. Naturally, social values, societal patterns, policy structures, and economic status of countries do impact Internet behaviors, and thus also risk behaviors which impact the willingness to share the data in the Internet. Therefore, respondents of other countries and cultures should be integrated in order to get a more international picture.

In addition, other usage contexts need to be examined. As such, data sharing in the medical context (with more sensible and person-related data) is an emerging field (Wilkowska & Ziefle, 2012) that needs a closer look as does the mobility and transport context in which car2x communication in automated driving (Schmidt et al., 2016) raises other privacy concerns that yet need to be explored. Overall it will be interesting to explore such critical situations in which data sharing on the one and data protection on the other hand make sense at the same time – however from the perspective of different interest groups.

ACKNOWLEDGEMENTS

This research has been funded by the Excellence Initiative of German state and federal governments (Project 'NEPTUN, no. OPSF316).

The authors express their special thanks to Chantal Lydinia and Juliana Brell for excellent research support as well as two anonymous reviewers for helpful comments on an earlier version of this manuscript.

REFERENCES

Akhter, S. H., 2014. Privacy concern and online transactions: the impact of internet self-efficacy and internet involvement. *Journal of Consumer Marketing,* 31, pp. 118–125.

Awad, N. F. and Krishnan, M. S., 2006. The personalization privacy paradox: an empirical evaluation of information transparency and the willingness to be profiled online for personalization. *MIS quarterly*, 30(1), pp. 13-28.

Barker, V. (2009). Older adolescents' motivations for social network site use: The influence of gender, group identity, and collective self-esteem. *CyberPsychology & Behavior*, 12(2), 209-213.

Boyd, D. and Hargittai, E., 2010. Facebook privacy settings: Who cares? *First Monday*, 15 (8).

Casaló, L. V., Flavián, C. and Guinalíu, M., 2007. The role of security, privacy, usability and reputation in the development of online banking. *Online Information Review*, 31, pp. 583–603.

Dartmann, G.; Demir, M.; Laux, H.; Lücken, V.; Bajcinca, N.; Kurt, G.; Ziefle, M. & Ascheid, G. (under revision). Privacy in Cyber-Physical Systems. In: S. Jeschke, C. Brecher, H. Song, & D. B. Rawat (Eds.). Cyber-Physical Systems - Foundations, Principles and Applications. Elsevier Series, *Intelligent Data-Centric Systems*".

Debatin, B., Lovejoy, J.P., Horn, A.-K., and Hughes, B.N., 2009. Facebook and Online Privacy: Attitudes, Behaviours, and Unintended Consequences. *Journal of Computer-Mediated Communication*, 15(1), pp. 83–108.

Dritsas, S., Gritzalis, D. and Lambrinoudakis, C., 2006. Protecting privacy and anonymity in pervasive computing: trends and perspectives. *Telematics and informatics*, 23(3), pp. 196-210.

Durndell, A. and Haag, Z., 2002. Computer self efficacy, computer anxiety, attitudes towards the Internet and reported experience with the Internet, by gender, in an East European sample. *Computers in human behavior*, 18(5), pp. 521-535.

Dwork, C., 2006. *Differential privacy.* Differential privacy. In Bugliesi, M. et al. (eds.), *Automata, Languages and Programming*, 4052, Lecture Notes in Computer Science, Springer Berlin Heidelberg, pp. 1-12.

Fogel, J. and Nehmad, E., 2009. Internet social network communities: Risk taking, trust, and privacy concerns. *Computers in Human Behavior,* 25, pp. 153–160.

Freestone, O. and Mitchell, V., 2004. Generation Y attitudes towards e-ethics and internet-related misbehaviours. *Journal of Business Ethics*, 54(2), pp. 121-128.

Gantz, J. and Reinsel, D., 2012. The digital universe in 2020: Big data, bigger digital shadows, and biggest growth in the far east. *IDC iView: IDC Analyze the Future*, pp. 1-16.

Halder, D., & Jaishankar, K., 2011. Cyber gender harassment and secondary victimization: A comparative analysis of the United States, the UK, and India. *Victims & Offenders*, 6(4), 386-398.

Hargittai, E., 2007. Whose space? Differences among users and non□users of social network sites. *Journal of Computer□Mediated Communication*, 13(1), 276-297.

Hornung, G. and Schnabel, C., 2009. Data protection in Germany I: The population census decision and the right to informational self-determination. *Computer Law & Security Review*, 25(1), pp. 84-88.

Kalwar, S. K., 2008. Human behavior on the internet. *Potentials, IEEE*, 27(5), pp. 31-33.

Karim, N. S. A., Zamzuri, N. H. A. and Nor, Y.M., 2009. Exploring the relationship between Internet ethics in university students and the big five model of personality. *Computers & Education*, 53(1), pp. 86-93.

Karabey, B., 2012. Big Data and Privacy Issues. In: Kurbanoğlu, S. et al. (eds.) *E-Science and Information Management*, Springer Berlin Heidelberg, p. 3.

Katal, A., Wazid, M. and Goudar, R. H., 2013. Big data: Issues, challenges, tools and Good practices. In: *Sixth International Conference on Contemporary Computing* (IC3), IEEE, Noida, India, pp. 404-409.

Kennedy, T., Wellman, B. and Klement, K., 2003. Gendering the digital divide. *It & Society,* 1(5), pp. 72-96.

Kim, D. Y., Lehto, X. Y. and Morrison, A. M., 2007. Gender differences in online travel information search: Implications for marketing communications on the internet. *Tourism management*, 28(2), pp. 423-433.

Kowalewski, S., Ziefle, M., Ziegeldorf, H., & Wehrle, K. (2015). Like us on Facebook!–Analyzing User Preferences Regarding Privacy Settings in Germany. *Procedia Manufacturing*, 3, 815-822.

Krasnova, H. and Veltri, N. F., 2010. Privacy Calculus on Social Networking Sites: Explorative Evidence from Germany and USA. In: *43rd Hawaii International Conference on System Sciences (HICSS),* Hawaii, USA, pp. 1–10.

Lahlou, S., 2008. Identity, social status, privacy and face-keeping in digital society. *Social science information,* 47(3), pp. 299-330.

Laufer, R. S. and Wolfe, M., 1977. Privacy as a concept and a social issue: A multidimensional developmental theory. *Journal of social Issues*, 33(3), pp. 22-42.

Luce, R. D. & Tukey, J.W., 1964. Simultaneous conjoint measurement: A new type of fundamental measurement," *Journal of Mathematical Psychology,* 1, 1-27.

Marx, G. T., 1998. Ethics for the new surveillance. *The Information Society*, 14(3), pp. 171–185.

Matsusaki, H., 2016. The Data Needs and the Scientific Methodologies of Marketing Studies: An Analysis From Ecological Perspectives. In: *Proceedings of the 1979 Academy of Marketing Science Annual Conference,* Springer International Publishing, pp. 374-374.

Mayeda, G., 2016. Privacy in the Age of the Internet: Lawful Access Provisions and Access to ISP and OSP Subscriber Information. *Alberta Law Review*, 53(3), pp. 2015-27.

Nissenbaum, H., 2011. A Contextual Approach to Privacy Online. *Daedalus*, 140, pp. 32-48.

Norberg, P. A., Horne, D. R. and Horne, D. A., 2007. The Privacy Paradox: Personal Information Disclosure Intentions versus Behaviors. *Journal of Consumer Affairs*, 41, pp. 100–126.

Phelps, J. E., D'Souza, G. & Nowak, G. J., 2001. Antecedents and consequences of consumer privacy concerns: An empirical investigation. *Journal of Interactive Marketing*., 15, pp. 2–17.

Ribble, M. S., Bailey, G. D. and Ross, T.W., 2004. Digital Citizenship:Addressing Appropriate Technology Behavior. *Learning & Leading with Technology*, 32(1), p. 6.

Rouvroy, A. & Poullet, Y., 2009. The right to informational self-determination and the value of self-development: Reassessing the importance of privacy for democracy. In Gutwirth, S. et al. (eds.) *Reinventing data protection*? Springer Netherlands, pp. 45-76.

Schmidt, T.; Philipsen, R. & Ziefle, M. 2016. Share to protect - Quantitative Study on Privacy Issues in V2X-Technology. Full paper at the 18th International Conference on Human-Computer Interaction, 2016, in press.

Spiekermann, S., 2005. Perceived control: Scales for privacy in ubiquitous computing. In 10th International conference on user modeling. Available at SSRN: http://ssrn.com/abstract=761109 or http://dx.doi.org/10. 2139/ssrn.761109

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

Sweeney, L., 2002. K-anonymity: A model for protecting privacy. *International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems*, 10(05), pp. 557-570.

Szongott, C., Henne, B. and von Voigt, G., 2012. Big data privacy issues in public social media. In: 6th IEEE International Conference on Digital Ecosystems Technologies (DEST), IEEE, Campione, Italy, 18-20 June 2012, pp. 1-6.

Takabi, H., Joshi, J. B. and Ahn, G. J., 2010. Security and privacy challenges in cloud computing environments. *IEEE Security & Privacy*, 6, pp. 24-31.

Tene, O. and Polonetsky, J., 2012. Big data for all: Privacy and user control in the age of analytics. N*w. J. Tech. & Intell. Prop.*, 11(5), pp. 239-273.

Thelwall, M., Wilkinson, D., & Uppal, S. (2010). Data mining emotion in social network communication: Gender differences in MySpace. *Journal of the American Society for Information Science and Technology*, 61(1), 190-199.

Thelwall, M. (2008). Social networks, gender, and friending: An analysis of MySpace member profiles. *Journal of the American Society for Information Science and Technology*, 59(8), 1321-1330.

Trstenjak, V., 2016. General report: The influence of human rights and basic rights in private law. In: Trstenjak and Weingerl (eds.) *The Influence of Human Rights and Basic Rights in Private Law*, Springer International Publishing, pp. 3-61.

Van Heek, J., Schaar, A. K., Trevisan, B., Bosowski, P. & Ziefle, M., 2014. User requirements for wearable smart textiles. Does the usage context matter (medical vs. sport)? In: S. Boll and F. H. Köhler (Eds.). User-Centered Design. 4th International Workshop on User-Centered Design of Pervasive Healthcare Applications (pp. 205-209). *Institute for Computer Science, Social-Informatics and Telecommunications Engineering (ICST).*

Van Heek, J., Arning, K., and Ziefle, M., 2015. Safety and privacy perceptions in public spaces: An empirical study on user requirements for city mobility. In Giaffreda, R., Caganova, D., Li, Y., Riggio, R., and Voisard, A. (Eds.). *Internet of Things 2014, LNICST* 151, Springer Berlin Heidelberg.

Van Heek, J.; Arning, K. & Ziefle, M. (2016). How Fear of Crime affects Needs for Privacy & Safety. Acceptance of Surveillance Technologies in Smart Cities. Full paper at the 5th International Conference on Smart Cities and Green ICT Systems (Smartgreens 2016), in press.

Wilkowska, W. and Ziefle, M., 2012. Privacy and data security in E-health: Requirements from the user's perspective. *Health Informatics J.,* 18, 191–201.

Wilkowska, W., Gaul, S., & Ziefle, M. (2010). A Small but Significant Difference – The Role of Gender on Acceptance of Medical Assistive Technologies (pp. 82-100). Springer Berlin Heidelberg.

Young, A. L. and Quan-Haase, A., 2013. Privacy Protection Strategies on Facebook. *Information, Communication & Society*, 16, pp. 479–500.

Ziefle, M., Himmel, S., & Wilkowska, W. 2011. When your living space knows what you do: Acceptance of medical home monitoring by different technologies (pp. 607-624). Springer Berlin Heidelberg.