Conference Paper

The Effect of a Blockchain-Supported, Privacy-Preserving System on Disclosure of Personal Data

Author(s):
Frey, Remo M.; Bühler, Pascal; Gerdes, Alexander; Hardjono, Thomas; Fuchs, Klaus L.; Ilic, Alexander

Publication Date:
2017

Permanent Link:
https://doi.org/10.3929/ethz-b-000204637 →

Originally published in:
http://doi.org/10.1109/NCA.2017.8171385 →

The Effect of a Blockchain-Supported, Privacy-Preserving System on Disclosure of Personal Data

Remo Manuel Frey
ETH Zurich
Zurich, Switzerland
rfrey@ethz.ch

Pascal Bühler
University of St. Gallen
St. Gallen, Switzerland
pascal.buehler@unisg.ch

Alexander Gerdes
Karlsruhe Institue of Technology
Karlsruhe, Germany
aljgerdes@autoidlabs.ch

Thomas Hardjono
Massachusetts Institute of Technology (MIT)
Cambridge, MA, USA
hardjono@mit.edu

Klaus Ludwig Fuchs
ETH Zurich
Zurich, Switzerland
fuchsk@ethz.ch

Alexander Ilic
ETH Zurich
Zurich, Switzerland
ailic@ethz.ch

***Abstract*—In light of digitalization, customers increasingly share private data through their online behaviors and actions. Yet, customers have become reluctant to share data due to privacy concerns. From a psychological perspective, a reduction of users' perceived risks should result in a higher willingness to share sensitive data. The development of blockchain-supported, multi-part computation thereby represents an interesting novel empirical context to study such willingness to disclose personal data, as such technologies involve a privacy-preserving approach that could not only technically solve privacy issues but also ought to address precisely the user's risk perception. Therefore, we conducted an online experiment with 420 participants to examine the willingness to disclose personal data dependent on different privacy protection mechanisms. A deception based experiment allowed to measure not only user intention, but also real user behavior. Surprisingly, our results demonstrate that participants shared similar amounts of personal data for blockchain-supported approaches and standard privacy policies. Even though an aversion to the blockchain system due to its novelty and potentially perceived complexity was not detected. Furthermore, we found that the willingness to share data increased significantly specifically for technically affine people when they were presented with the opportunity to monetize their data. We further discuss the effects of privacy awareness and whether prior knowledge of blockchain technology had a supporting effect for user acceptance.**

Keywords—personal data, data privacy, privacy-preserving system, blockchain, user behavior

I. Introduction

New privacy-preserving data sharing approaches are technically able to secure sensitive data with cryptographic techniques and provide some degree of technical-trust. This study focuses on a solution based on blockchain-supported multi-party computation. The blockchain acts as platform for a tamper-proof handling of the data access permissions while the data itself is stored encrypted off-chain. Computations are made by multi-party computation, which allows running algorithms on raw data, but without having access on it. Zyskind, Nathan, and Pentland [1] describe the technical details of such a system. The aim of the system is to guarantee data protection cryptographically and thus to minimize the real risks of misuse. The contribution of this article lies in the first attempts to investigate the system from a user perception point of view rather than from a purely technical perspective. The motivation behind this work aims to highlight and explain a widely seen phenomenon also present in similar areas. For example, despite existing email encryption solutions have been available on the market for many years now, their user acceptance is still very low despite the obvious benefits of increased privacy. Since blockchain-based systems are rather new, this allows us to study the phenomena in an adoption phase rather than an inertial habitual stage of a privacy-relevant application. However, blockchain-supported systems are significantly more complex and users might be overwhelmed by not being able to understand and verify all of the underlying fundamentals that ultimately result in the privacy protection. Therefore, it is important to get a better user understanding of the effect of such a system on user behavior in an early stage. Due to the novelty, it is not yet explored in research how a blockchain-supported multi-party computation system affects user behavior. Especially the perception of privacy risks plays a key role in user acceptance of security systems because the reduction of perceived privacy risks leads to a higher willingness to share personal data [2]. In the present article, we assess the willingness to share personal data with this type of system and compare it with other well-established risk reduction instruments in context of online activities.

II. Research Design

A. Hypotheses

Researcher propose new privacy-preserving data sharing systems, build proof-of-concepts and demonstrate cryptographic correctness. From a technology point of view, these approaches are ready for implementation in real business cases and are thus highly promising. Blockchain-supported

978-1-5386-1465-5/17/$31.00 ©2017 IEEE

multi-party computation represents one such application. Its novelty however imply that user experience or user study considering such a system do not yet exist. We know little about a users' perception of such systems in regards to his willingness to disclose private data, which is typically center stage of privacy discussions. We thus focus on four specific hypotheses for better understanding the expected effect of privacy-preserving technology on users' behavior.

As the blockchain-supported multi-party computation system allows privacy-preserving sharing of data, we expect to confirm the first hypothesis: *Blockchain-supported multi-party computation systems increase the willingness to share personal data (H1).* Due to the Privacy Calculus [3], we expect the user-based data disclosure to be better achieved by an attractive incentive system. Thanks to the blockchain technology, the system can be seamlessly expanded with micropayment without having to give up anonymity. Therefore, we examine the monetary extension in the second hypothesis: *Blockchain-supported multi-party computation systems with an option for data monetization increase the willingness to share personal data (H2).* Since the complexity of such a system is high, we hypothesize that user expertise has an influence on the user behavior: *Technical affinity (H3a) and previous knowledge about blockchain technology (H3b) increase the willingness to share personal data in a blockchain-supported multi-party computation system.* The fourth hypothesis considers users' individual privacy awareness. Since the system promises to be cryptographically secure and misuse of data is practically impossible, one can expect users with privacy concerns to trust the system for secure data sharing. Thus, the hypothesis is: *Due to the blockchain-supported multi-party computation system, users with privacy concerns share the same amount of data as users without privacy concerns do (H4).*

B. Online Experiment

The experiment take shape in form of an online survey with 420 participants. In order to better understand users' disclosure behavior we consider two widely acknowledged risk reduction instruments: presenting a privacy policy to the users [4], and customer empowerment [5]. To test the four hypotheses, we compare these two instruments with two versions of a blockchain-supported multi-party computation system. One version includes monetization incentives and a second version omits monetization of personal data. We compare the instruments against a control group, where no risk reduction efforts are carried out. The study participants are thereby randomly assigned to one of the described options. To recruit participants, a tablet as a monetary incentive is raffled.

Care must be taken to select a methodology in which data sharing is as realistic as possible [6]. According to Kehr et al. [7], the actual behavior is to be analyzed and not merely intentions, prevailing opinions and irrational behavior. Otherwise, a corruption of the true results due to the privacy-paradox and the privacy calculus ought to be expected. Therefore, we decided to use a deception mechanism, a classical instrument in psychology research. Since deceptions have been subject to ethical concerns, the procedure was examined and approved by the ethics committee of ETH Zurich. The mechanism is essential to overcome the well-known intention-behavior gap [8]. In this study, we hence ask participants to answer an online questionnaire about a topical subject ("future mobility"), which is unrelated to our hidden research hypotheses, where the majority of all questions in the questionnaire deals with said topic. The participants however are left in the belief that the research is about said topic until the end of the online experiment.

The workflow of the online survey is as follows: When participants start the questionnaire, they are automatically randomly assigned to one of the five treatment groups. This process happens in the background and participants are not aware of it. Then, the participants are primed: A short text explains how the responsible researchers deal with the collected data of the participants. The text differs according to participants' assigned group. The five text versions are presented in the next section. After the priming, participants answer the irrelevant questions about "future mobility" and the relevant questions about personal data like participants' address. In addition, some extra questions are posed concerning user characteristics like technical affinity. In the end, the participants are informed about the deception, and they have the possibility to delete all disclosed data and to refuse participation.

C. Treamtment Groups

Five groups are defined in the experiment in order to compare different privacy mechanisms and measure the respective user behavior. Each group is exposed to a different priming. The groups are treated as follows before the start of the questionnaire:

1) No Privacy Policy. No text is displayed to the user.

2) Standard Privacy Policy. The displayed text is: "All data collected will be treated anonymously and strictly confidential. No personal data will be passed on to third parties. The collected data are used exclusively for scientific purposes within this research project and are deleted after study completion."

3) Customer Empowerment. The displayed text is as follows and includes three checkboxes: "Please select one or more of the following privacy options: (a) Deletion of data within 42 hours: Your data will be deleted from our servers within 24 hours. (b) Data may only be used for internal optimization approaches. Your data will not be shared with others under any circumstances. (c) Personal 'data tracking': You will get an overview of all accesses to your data. This allows you to fully understand how and which data are shared with which legal entities and how these data are passed on."

4) Blockchain-Supported Multi-Party Computation (hereinafter referred as "blockchain"). The displayed text is: "To ensure your privacy, we use a decentralized, blockchain-supported data protection technology. This allows us to use your customer data for our purposes without owning this data. Your data is stored in an encrypted database. No one can view the information you have provided. We can only access aggregated indicators of the total data volume by means of computing operations. For instance, it is only possible for us to calculate the frequency of a selected color preference. We do

not know who has expressed his predilection and cannot assign the preferences given to them in a person-specific manner. We therefore have no information about the participant."

5) Blockchain-Supported Multi-Party Computation with Option for Monetization (hereinafter referred as "monetary blockchain"). The displayed text is the same as before but with an additional section at the end: "Each time we access a data point of your personal data, you will receive a reward of 0.01 Euro per record. Payment processing is carried out at the end of the questionnaire."

The answers of all groups are stored in the same way, in a relational database on our server. We do not use blockchain-supported multi-party computation for the latter two groups because it does not have an impact on the experiment results. In the end, the participants are informed about the deception.

D. Data Analysis

The hypotheses focus on the willingness to share personal data. Since the participants are not aware about the true purpose of the questions, the experiment is able to measure the true behavior of the participants as it may occur in typical online activities. Therefore, the willingness to share personal data can be measured by counting the number of entered data items. Participants are asked to enter the following personal data items: name, street, zip code, email address, phone number, driver license number, net salary. They have to provide an answer for each question. They can either answer the question or decide for the fallback option "I will not specify". Each question has to be answered in this sense.

We expect that the priming in the different treatment groups leads to differences in how often the option "I will not specify" is chosen. Therefore, we compare the amount of entered data items among the five treatment groups. The entered value itself is not considered, in order to preserve the participant's anonymity. We assume that not every data item is perceived as equally sensitive. For instance, the name as an explicit identifier is more sensitive than the zip code. Hence, instead of summarizing data items, we consider each data items as a single measurement for the willingness to share personal data. In other words, the analysis of each data item can be understood as a repeated measure of the same dependent variable. We hence apply a One-Way Repeated Measures ANOVA to test our hypotheses.

In order to increase test power and receive a test statistic comparable to values of the F-distribution, we test the assumption of sphericity with a Mauchly's test. Values greater than 0.05 means that the assumption is not violated. If the assumption is violated, the degrees of freedom are adjusted in order to make the F-ratio more conservative. The Huynh-Feldt correction is used if the Greenhouse-Geisser estimate of sphericity is greater than 0.75, otherwise the Greenhouse-Geisser correction. Mauchly's test for sphericity is only relevant for more than two independent variables (treatment groups). Therefore, the tests are only applied to the first hypothesis. The ANOVA does within-subjects effects comparisons. We additionally conduct pairwise comparisons (post hoc tests): Fisher's LSD and the Bonferroni procedure are performed. We present both in the result section.

III. Results

The study was conducted in Switzerland and Germany between August 19, 2016 and October 5, 2016. For recruiting, we used a survey service from the Institute of Psychology at the University of Zurich. 6.267 email invitations and were sent until September 2, 2016. In total, we received 295 questionnaires. To enlarge the sample, advertising on Facebook was used. In total, 1.111 views (clicks) on the questionnaire were recorded. 1.102 people started to answer the survey. 420 of them finished the whole survey and agreed to open the data for research, despite the deception. Not fully completed questionnaires or participants who rejected the data agreement were excluded from further consideration. Table I shows the answers which are relevant for the third and fourth hypothesis. The relatively new blockchain technology seems to be not yet well-known: 82.1% of the participants have never heard about blockchain technology in the context of data privacy.

A. Comparison Between Risk Reduction Instruments

As explained in a previous section, each participant was asked to enter seven personal data items (name, phone number, etc.). We had to remove the data item "email address" from data analysis because of an issue with the provided instructions on the welcome page. Fig. 1 provides an overview of the results. It shows the percentage of disclosed items per treatment group and for different user characteristics (highly technical affine, blockchain affine, and having privacy concerns). As expected, the participants want to reveal the least data if they do not receive a privacy policy or other risk reduction instrument. Compared to the standard privacy policy, it seems that the blockchain-supported system with option for monetization ("monetary blockchain") and customer empowerment are overall more successful, except for people with privacy concerns. It is interesting to note the difference between high technical affinity and blockchain affinity. While monetary blockchain is equally good for both, customer empowerment significantly reduced when blockchain affine participants were addressed. We assume that these people have a better understanding of risks that may arise from the handling of personal data than people who are "only" technical affine. They may recognize that customer empowerment does not provide proper protection against misuse and fraud, but view it as a strategy to merely gain the trust of customers. Further research could examine these differences, which could corroborate our assumptions.

Test results are shown in Table II. Mauchly's test indicates that the assumption of sphericity had not been violated for most cases, except for technical affine participants. Therefore, the degree of freedom of this group is adjusted according to the rules described in the analysis section. F is significant for all groups, which means that there are significant differences between the risk reduction instruments. In all result tables, the

TABLE I. Answers Concerning Technical Affinity, Blockchain Affinity, and Privacy Concerns (N=420)

Question	Yes	No	No Answer
High technology affinity?	137 (32.6%)	280 (66.7%)	3 (0.7%)
Blockchain technology known in context of data privacy?	70 (16.7%)	345 (82.1%)	5 (1.2%)
Concerns about personal data when disclosed to companies?	201 (47.9%)	212 (50.5%)	7 (1.7%)

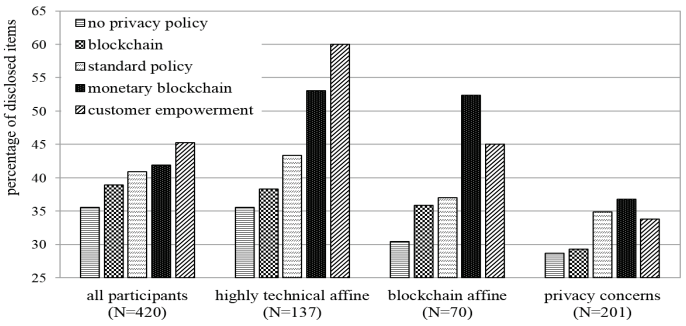

Fig. 1. Comparison between the five risk reduction instruments and three user characteristics.

TABLE II. TESTS OF WITHIN-SUBJECTS EFFECTS

	F	*df*	*df (error)*	*p*	*partial η2*
all participants	10.90	4	20	**.000**	.686
technical affine	20.71	1.83	9.14	**.000**	.806
blockchain affine	6.43	4	20	**.002**	.562
privacy concerns	3.78	4	20	**.019**	.430

significant values are marked in bold. In addition, we provide the values of partial Eta-squared. It means for all participants that the risk reduction instrument explains 68.6% of the variation in the willingness of sharing personal data.

Table III shows the pairwise comparisons (post hoc) between the different risk reduction instruments. If the Bonferroni correction is added, the risk reduction instruments no longer significantly differ from one another in the group of all participants. Therefore, the first hypothesis H1 cannot be confirmed despite the observed trends and significant differences indicated by the F-test. Thus, Blockchain-supported multi-party computation systems appear not to increase the willingness to share personal data in the scope of our study at least. Nonetheless, it is interesting to note that the system was not met with aversion and it did not perform worse compared to other the instruments.

However, significant differences in some subgroups exist. For technical affine users, the absence of a privacy policy performs significantly worse than all other options, and customer empowerment appears to perform significantly better than a standard privacy policy. This is in line with the literature. For blockchain affine people, blockchain-supported multi-party computation system with the option for monetization ("monetary blockchain") performs significantly better than without a privacy policy. For people with privacy concerns, the monetary blockchain significantly increased the willingness to share personal data compared to the system without the monetization mechanism. Thus, the second hypothesis H2 can be confirmed. There is a significant influence of privacy-preserving systems if an option for monetization of personal data is provided.

B. Comparison Between User Characteristics

1) Technical Affinity. Fig. 2 (left) displays the percentage of disclosed items of highly technical affine people and people who are not. While all risk reduction instruments for the latter group are more or less equal, customer empowerment and monetary blockchain significantly increase if people are highly technical affine (p=.003 and p=.014). Thus, technical affinity only has a significant effect on privacy-preserving systems if an option for monetization is provided. Therefore, the hypothesis H3(a) holds only partially.

2) Blockchain Affinity. The results for blockchain affinity are shown in Fig. 2 (middle). Blockchain affinity appears to impose a similar effect on sharing willingness as technical affinity: the monetary blockchain performs significantly better for people who know the blockchain technology in context of data privacy opposed to people who do not (p=.032). Surprisingly, customer empowerment remains at a high level, but cannot keep up with the monetary blockchain approach, as it was still the case with technical affinity. These findings suggest that blockchain educated people realize that customer empowerment offers advantages, but does not provide full control over personal data, necessary for adequate and proper protection. Thus, analogue to H3(a), hypothesis H3(b) holds only for blockchain-supported multi-party computation systems with the option for data monetization.

TABLE III. PAIRWISE COMPARISONS (POST HOC TESTS)

		all participants			technical affine			blockchain affine			privacy concerns		
group 1	**group 2**	***Δm***	p_{LSD}	$p_{Bonferroni}$	***Δm***	p_{LSD}	$p_{Bonferroni}$	***Δm***	p_{LSD}	$p_{Bonferroni}$	***Δm***	p_{LSD}	$p_{Bonferroni}$
no privacy policy	blockchain	-3.37	**.023**	.228	-2.778	.224	1.000	-5.441	.093	.927	-.611	.730	1.000
	standard policy	-5.36	**.007**	.071	-7.778	**.003**	**.026**	-6.645	.311	1.000	-6.202	**.014**	.136
	monetary blockchain	-6.35	**.018**	.181	-17.475	**.001**	**.015**	-21.989	**.001**	**.014**	-8.123	**.006**	.062
	customer empowerment	-9.72	**.007**	.067	-24.444	**.002**	**.017**	-14.608	**.011**	.112	-5.141	.298	1.000
blockchain	no privacy policy	3.37	**.023**	.228	2.778	.224	1.000	5.441	.093	.927	.611	.730	1.000
	standard policy	-1.98	**.011**	.108	-5.000	.076	.756	-1.204	.882	1.000	-5.591	**.007**	.073
	monetary blockchain	-2.98	.131	1.000	-14.697	**.017**	.166	-16.548	**.027**	.269	-7.513	**.000**	**.003**
	customer empowerment	-6.35	**.020**	.200	-21.667	**.005**	.054	-9.167	.069	.686	-4.531	.196	1.000
standard policy	no privacy policy	5.36	**.007**	.071	7.778	**.003**	**.026**	6.645	.311	1.000	6.202	**.014**	.136
	blockchain	1.98	**.011**	.108	5.000	.076	.756	1.204	.882	1.000	5.591	**.007**	.073
	monetary blockchain	-.99	.565	1.000	-9.697	**.028**	.283	-15.344	**.013**	.127	-1.922	.157	1.000
	customer empowerment	-4.37	.050	.504	-16.667	**.004**	**.042**	-7.963	.208	1.000	1.060	.801	1.000
monetary blockchain	no privacy policy	6.35	**.018**	.181	17.475	**.001**	**.015**	21.989	**.001**	**.014**	8.123	**.006**	.062
	blockchain	2.98	.131	1.000	14.697	**.017**	.166	16.548	**.027**	.269	7.513	**.000**	**.003**
	standard policy	.99	.565	1.000	9.697	**.028**	.283	15.344	**.013**	.127	1.922	.157	1.000
	customer empowerment	-3.37	**.023**	.228	-6.970	**.048**	.481	7.381	.098	.984	2.982	.387	1.000
customer empowerment	no privacy policy	9.72	**.007**	.067	24.444	**.002**	**.017**	14.608	**.011**	.112	5.141	.298	1.000
	blockchain	6.35	**.020**	.200	21.667	**.005**	.054	9.167	.069	.686	4.531	.196	1.000
	standard policy	4.37	.050	.504	16.667	**.004**	**.042**	7.963	.208	1.000	-1.060	.801	1.000
	monetary blockchain	3.37	**.023**	.228	6.970	**.048**	.481	-7.381	.098	.984	-2.982	.387	1.000

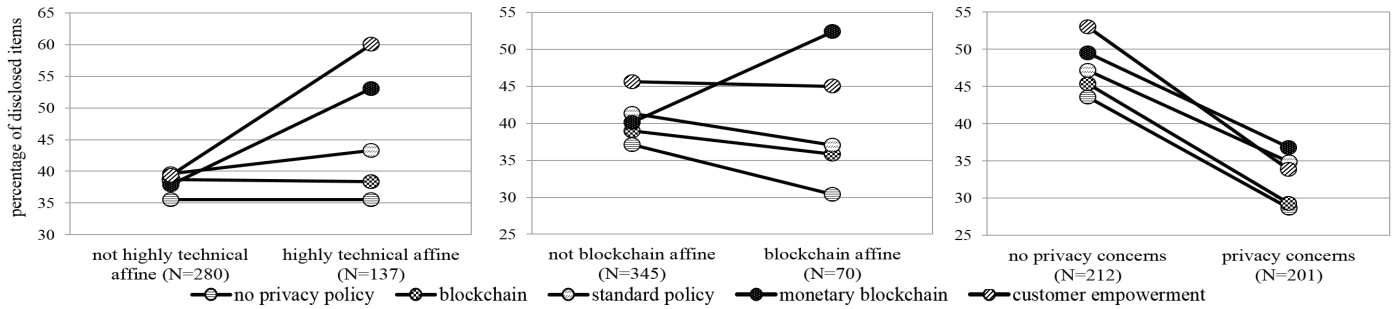

Fig. 2. Comparison with respect to technical affinity (left), blockchain affinity (middle), and privacy concerns (right).

3) Privacy Concerns. The findings suggest that privacy concerns have a significant effect on all risk reduction instruments (p<.05), as shown in Fig. 2 (right). The effect is strong for both investigated privacy-preserving systems, blockchain and monetary blockchain: partial Eta-squared is 75.8% and 81.4% respectively. The results clearly disconfirm the fourth hypothesis H4. As privacy concerns were not solved and participants did not share more personal data compared to other presented instruments, this suggests that participants with high security requirements were not convinced by the new technological solutions ("blockchain", "monetary blockchain") to share more personal data.

IV. Discussion

After Fisher's LSD post hoc test has been evaluated, the experiment indicates clear results, but must be relativized after the strict Bonferroni correction. The research design is based on many multiple comparisons for answering the first research question. This greatly increases the likelihood of making a Type I error, which is why the Bonferroni correction works very strongly. It is therefore quite possible that the chosen setting with five treatment groups had a negative effect on the meaningfulness of the results. However, when comparing the risk reduction instruments, it can be concluded at least that the presented privacy-preserving systems do not trigger aversion among the users and the willingness for disclosing personal data does not differ significantly from other instruments.

Several limitations of this work, which leads to opportunities for future research, are worth of further discussion. First, the presented results of the online experiment appeared strongly dependent on the priming. That is, the choice of displayed descriptions about the instruments per se. We discussed the descriptions intensively in advance and evaluated the feedback within a preliminary study. However, the development of an easily understandable description that feature the core principles of a sophisticated and complex system was challenging. It appeared easier to confuse the users rather than educating the user on privacy. The sample represent a second limitation. The participants of the study had a lower age and higher education level compared to the average population. Future studies may place more emphasis on demographic factors to increase the external validity of the study. Besides gender and age specific differences, cultural and geographic considerations could be particularly interesting. As the worldwide growing network and the accompanying increase in data exchange brings conflicts with data protection and privacy. Certain countries possess large differences in their legal systems, as a comparison between the US and Germany has shown by Kumar and Reinartz [9] for example. It opens a loophole to globally operating companies and, in particular, represents a privacy risk for modern customer groups [10]. Third, it seems that the monetarization option has a different nature than the others, so it would be interesting to consider it as a different variable and to analyze its interactions with the different data protection instruments. This would allow a deeper exploration of the psychology of "privacy calculus" by considering on the one hand the perceived level of data protection and, on the other hand, the potential economic advantages of taking such risks. Finally, the three user characteristics (technology affinity, blockchain affinity, privacy concerns) were considered as binary responses. Using a scaled survey questionnaire for the respective items could render more insights than asking participants to directly report these values.

References

[1] G. Zyskind, O. Nathan, and A. Pentland, "Enigma: Decentralized Computation Platform with Guaranteed Privacy," 2015.

[2] N. Olivero and P. Lunt, "Privacy versus willingness to disclose in e-commerce exchanges: The effect of risk awareness on the relative role of trust and control," *Journal of Economic Psychology*, vol. 25, pp. 243–262, 2004.

[3] M. J. Culnan and R. J. Bies, "Consumer privacy: Balancing economic and justice considerations," *Journal of Social Issues*, vol. 59, no. 2, pp. 323–342, 2003.

[4] N. F. Awad and M. S. Krishnan, "The Personalization Privacy Paradox: An Empirical Evaluation of Information Transparency and the Willingness to Be Profiled Online for Personalization," *MIS Quarterly*, vol. 30, no. 1, pp. 13–28, 2006.

[5] T. T. Van Dyke, V. Midha, and H. Nemati, "The effect of consumer privacy empowerment on trust and privacy concerns in e-commerce," *Electronic Markets*, vol. 17, no. 1, pp. 68–81, 2007.

[6] L. Miesler and A. Bearth, "'Willingness to share'" im Kontext Big Data: Wie entscheiden Kunden, ob sie ihre persönlichen Daten mit Unternehmen teilen?," in *Dialogmarketing Perspektiven 2015/2016*, Wiesbaden: Springer Fachmedien, 2016, pp. 49–66.

[7] F. Kehr, D. Wentzel, and T. Kowatsch, "Privacy Paradox Revised: Pre-Existing Attitudes, Psychological Ownership, and Actual Disclosure," in *Thirty Fifth International Conference on Information Systems*, 2014.

[8] S. Orbell and P. Sheeran, "'Inclined abstainers': a problem for predicting health-related behaviour.," *The British journal of social psychology / the British Psychological Society*, vol. 37, pp. 151–165, 1998.

[9] V. Kumar and W. Reinartz, "Customer Privacy Concerns and Privacy Protective Responses," in *Customer Relationship Management: Concept, Strategy, and Tools*, Berlin, Heidelberg: Springer Verlag, 2012, pp. 279–300.

[10] P. Schaar and J. Onstein, "Datenschutzrecht in der vernetzten Welt des 21. Jahrhunderts," *Bonner Rechtsjournal*, vol. 2, 2011.