Data Privacy in Ride-Sharing Services: From an Analysis of Common Practices to Improvement of User Awareness

Carsten Hesselmann[1], Delphine Reinhardt[2], Jan Gertheiss[3], and
Jörg P. Müller[1]

[1] Clausthal University of Technology, Adolph-Roemer-Straße 2A, 38678
Clausthal-Zellerfeld, Germany
[carsten.hesselmann,joerg.mueller]@tu-clausthal.de
[2] Georg August University Göttingen, Wilhelmsplatz 1, 37073 Göttingen, Germany
reinhardt@cs.uni-goettingen.de
[3] Helmut Schmidt University, Holstenhofweg 85, 22043 Hamburg, Germany
jan.gertheiss@hsu-hh.de

Abstract. Individuals are frequently confronted with privacy-related decisions under uncertainty especially in online contexts. The resulting privacy concerns are a decisive factor for individuals to (not) use online services. In order to support individuals to make more informed decisions, we assess the current state of practice of certain online services. This analysis is focused on ride-sharing services and includes popular services in Germany, Austria, and Switzerland and we investigate how they handle user data. The results show that services include a wide-ranging set of personal data and lack standardization. Furthermore, they offer limited privacy-related features. Based on this analysis, we developed a Transparency Enhancing Technology in the form of a browser extension that informs users about a service's data practices at the time of data disclosure. In addition to this, we conducted a scenario-based online experiment with a representative sample to evaluate our tool's usability and its effect on users' concerns and behavior. Results show significant improvements in awareness and decision reflection with limited decrease in disclosure rates of personal data.

Keywords: Data Privacy · Disclosure Behavior · Sharing Economy · Transparency Enhancing Technology

1 Introduction

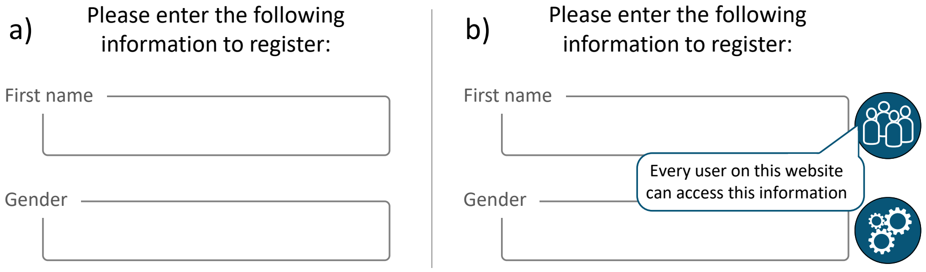
Privacy has been a topic of expanding interest for researchers, economists, and regulators alike [35]. This is a distinct indicator that the recent developments in data collection and processing are problematic as privacy regulations and privacy research are a reactive area [7]. More specifically, individuals — also referred to as users in this work — frequently make privacy-related decisions under uncertainty as a consequence of incomplete information and information asymmetry [3]. This, in turn, results in a lack of transparency and control over personal data while increasing individuals' anxiety and concern [6]. This is substantiated by examples such as the misuse of web browsers' device battery API by companies to increase prices [27] and the growing market for online personal data which stays obscure and out of reach for the individual [4]. Tools and technologies that try to aid individuals in their privacy choices exist in physical (shutter for privacy webcams [22]) and in digital form (tools that help with privacy choices [10]) but are scarce.

Digital data collection is substantially easier and more broadly applied compared to the physical world. In addition, concerns about digital privacy are similar to concerns in the physical world [46] and therefore have to be taken into consideration from the stakeholders of any data collection and processing. Moreover, the research on how the privacy practices of companies affect individuals is limited [6] and needs to be extended on.

This applies especially well to the sharing economy — a term summarizing Peer-to-Peer (P2P) markets for the temporary share or rental of goods — as sharing blends borders between the online and offline [31] as well as the private and the economic spheres [37]; in this work we focus on services for sharing rides. Markets in the sharing economy are partly based on reputation systems which enable consumers to evaluate other (unknown) participants in the market based on crowd-sourced information [34]. In addition to that, the quality of interactions in P2P markets can vary greatly depending on each individual and the market is regulated by reciprocity [29]. For these reasons, user retention and loyalty are of major importance and any barriers and impediments should be reduced. One of the dominant factors that prevent participation and transaction execution in e-commerce and the sharing economy is privacy concerns [1, 14, 42].

The topic of data collection and disclosure in sharing services is complex and equivocal. The disclosure of personal information required to gain access to (ride-sharing) services is applied in a ask-on-first-use principle — typically used in mobile applications — which is insufficient to match individuals' preferences [18, 47] and, furthermore, occurs at an early stage of digital interaction, making the individual feel hopeless about their data [19]. Supporting an individual's assessment of privacy choices is a needed area of research [10] as users are willing to rethink their decision when provided with a plausible reason [21].

In this paper, we present the results of our ongoing research on the disclosure behavior in the sharing economy. Our goal is to make data practices of ride-sharing services more transparent for the user. Therefore, we decided to design and implement a TET in the form of a browser extension that adds relevant information to the sign-up process – its functionality is depicted with a simplified example in Figure 1 and further detailed in Section 4. There are multiple reasons for our decision to create a TET. Firstly, personal preferences surrounding privacy are highly subjective [36]. Secondly, recent research has shown that individuals not only have differing preferences about partly giving up control but prefer to remain in control of their decisions [9, 19, 21, 24]. Thirdly, incomplete information hinders privacy decisions [2] especially due to the disconnect



Please enter the following information to register:

Please enter the following information to register:

Every user on this website can access this information

Fig. 1. Simplified registration form without a) and with b) the proposed TET

between data collection and its usage. This disconnect is felt on the users side as lack of awareness about the data usage [47]. Lastly, transparency positively affects individuals' reaction and a lack of it could result in future backlash [6] as lack of awareness is one of the key components of privacy concerns [32].

The contributions of this paper are threefold. After reviewing related research in Section 2, we analyze the state of practice of popular ride-sharing services in Section 3 and identify differences in data practices and privacy-related features, addressing the Research Questions (RQs):

RQ 1: *What personal data are commonly included in ride-sharing services?*
RQ 2: *How much of the personal information is exposed to other individuals?*
RQ 3: *What privacy-related features do services for ride-sharing offer?*

Our analysis shows that privacy-related features are rare and commonly not easy to access. In addition, most services' transparency about data practices could be improved, as opt-out options are commonly difficult to access [11]. A study has shown that data practices affect users' privacy concerns [15]. However, the data was attained in a self-reporting manner which usually correlates less with actual behavior.

Therefore, we implemented a tool which integrates privacy-related information into the websites of ride-sharing services via an icon-based approach, as described in Section 4. The tool is a browser extension that works with all common browsers. The goal of this is to empower individuals to make a more informed decision when deciding about disclosing personal information and is formulated as RQs:

RQ 4: *How much personalization and automation is applicable?*
RQ 5: *Which technical approach is applicable for the described goal?*
RQ 6: *Which design approach is suitable?*
RQ 7: *How can long-term usefulness be ensured and data set kept up-to-date?*

Furthermore, we conducted a scenario-based experiment with a representative sample ($n = 1093$) to get a primary evaluation of our tool and answer RQs:

RQ 8: *Do the icons change the participant's privacy concerns?*

RQ 9: *Do the icons change the participant's decision about data disclosure?*
RQ 10: *Do the participants perceive and understand the icons?*
RQ 11: *Are the participants aware of how much data they actively disclosed?*
RQ 12: *Do the participants use or perceive the available profile settings?*
RQ 13: *What is the most helpful information for the participants?*

We used the Internet Users' Information Privacy Concern (IUIPC) metric to assess participants' privacy concerns, as discussed in Section 5. The results indicate significant effects on the dimensions awareness and (partly) collection. Furthermore, the disclosure rates are affected for certain combinations of icon and personal data.

2 Related Work

Privacy research is multifaceted. We start by mentioning a set of notable works which include similar approaches based on visual cues and data disclosure. Similar to [20], we analyze the disclosure behavior during a sign-up process and aim to improve the transparency of underlying data practices for the individual based on visual cues. A key difference from their work is that we do not analyze affection in our experiment and therefore do not include framing in our visual cues. Instead, the visual cues and textual assistance in our experiment are designed and formulated neutrally. Furthermore, our experiment is adjusted to the sharing economy, i.e., including common practices based on a prior analysis of services, and therefore contributes specifically to this field of research.

The process of signing up for digital services is similar to the installation process of mobile applications at least as far as the disclosure of information is concerned. It is mandatory to disclose a set of personal information to complete a sign-up process. Equally, the installation of a mobile application demands the granting of permissions, which in turn leads to the disclosure of a variety of data. [16] study individuals' disclosure behavior during the installation process of mobile applications by adding visual cues hinting at potentially dangerous permissions. In consequence, the authors use the Mobile Users' Information Privacy Concern (MUIPC) metric instead of IUIPC which is used in our work. Additionally, our work does not include a three-color approach which typically conveys implicit information – for example, by using green, yellow, and red visual cues – as we firstly wanted to evaluate the effects of visual cues without framing. Furthermore, we do not include individuals' general privacy concerns as a separate set of questions in order to compare their decisions with their self-stated attitudes because we compare the results from the test and control group. In addition to that, the IUIPC metric includes broad privacy statements and the visual cues used in our work include the most imminent consequences of potential data disclosure by highlighting the data practices of the service and paying emphasis specifically to other individuals' access to the personal information disclosed.

Similarly to the installation of mobile applications, [18] investigate users' knowledge about browser extensions and their preferences for installation notifications. This directly relates to our proposed tool as it is a browser extension and we have incorporated minimal permission requests and limit the extension's activity to relevant websites of ride-sharing services. Furthermore, [18] found that users prefer more extensive dialogues with examples of what data is potentially accessed. We integrate this finding in our implementation and offer a summary function that includes all relevant data which are otherwise communicated via individual icons across the website and its sub-pages. [11] measure how risk-based icons (icons that convey the result of a risk analysis) can lead to a more informed consent of individuals. Our approach shows similarity to the work of [11] as both aim at providing means of comparability between different services and policies respectively. However, in contrast to [11], our work focuses on data disclosure instead of privacy policies. In summary, our work contributes to the privacy research specifically for the sharing economy by evaluating the disclosure behavior of individuals during a sign-up process in a scenario-based experiment without decision-nudging or framing.

3 Analysis of Ride-Sharing Services

To assess the current state of practice of ride-sharing services and how transparent they present their practices to the user for RQs 1-3, we analyzed the websites of the most popular ride-sharing services in Germany, Austria and Switzerland, according to various ratings [12, 39–41].

The results cover eleven ride-sharing services and 39 data attributes in total (after merging similar attributes, as summarized in Table 1). Our analysis shows which data attributes are included in a service and are exposed to other users, as discussed in Section 3.1. In addition, available privacy settings and the validation of information are reviewed in Section 3.2. This analysis was limited to shared rides. Therefore, if cargo transport or the like is offered, it is not included.

This analysis was limited to those areas and features of the websites which are accessible to users as these include, e.g. the privacy policies. Furthermore, the analysis only includes information that is directly linked to the individual, their preferences, or information about their vehicle. Consequently, information relating solely to a ride offer, e.g., locations and routes, is not included in this analysis as it is an extensive research area on its own.

Table 1. Merged data attributes

contact information	cellphone number, landline number, fax number
social media	Facebook, YouTube, personal website
personal description	personal characteristics, self description, things I like
interests	sport, hobby, movie
job	job description, job industry
address	country, city, zip code, street

The following steps were carried out to analyze a ride-sharing service: (i) register an account (if possible), (ii) complete the profile with data attributes, (iii) check for available profile settings, (iv) review profile pages, (v) review ride offers, (vi) create ride offers, and (vii) book ride offers.

3.1 Collection and Exposure of Personal Information

Our results show that there is a great variety of user data which is included in the services; the set of collected data attributes ranges between 5 and 29, and the set of exposed data attributes varies between 4 and 17. The statistics for all analyzed ride-sharing services are provided in Table 2. In addition, Figure 2 displays the full set of included data attributes across all analyzed services. The data points are widely spread. This indicates the diversity in included personal information and a lack of standardization in this regard. The fact that each data attribute is on average included in 36% of ride-sharing services confirms this lack of standardization and concludes RQ 1.

To assess the exposure of personal information for RQ 2, we investigated whether disclosed personal information is accessible for other users. Information about smoking behavior, the vehicle and profile picture are shared most often. On average, services show 76% of the disclosed personal information to other users. This means one quarter of the disclosed personal information is not part of any user-to-user interaction and remains only with the service provider. Some services are close to the 50% margin, which emphasizes how different the data practices depend on the choice of service. This raises the question as to why the user should disclose their personal information to the service provider if not even half of it is accessible to the other users, especially in the context of sharing rides, where – by design – the interaction with other users is arguably the main reason for an individual to use such a service. This question is further aggravated by the fact that explanations to the user on why this information should be disclosed

Table 2. Details of collection and exposure of personal data

	bessermitfahren	blablacar	clickapoint	e-carpooling	fahrgemeinschaft	foahstmit	greendrive	mifaz	mitfahrportal	pendlerportal	twogo
collected data	9	23	12	16	14	5	6	16	29	21	14
exposed data	9	12	10	11	12	5	4	15	17	11	10
mandatory	3	14	3	9	5	4	2	2	4	12	10
optional	6	9	9	7	9	1	4	14	25	9	4
profile page	0	2	0	1	3	0	0	0	5	0	2
ride offer	9	4	5	5	3	5	4	7	10	11	8
both	0	6	5	5	5	0	0	8	2	0	0

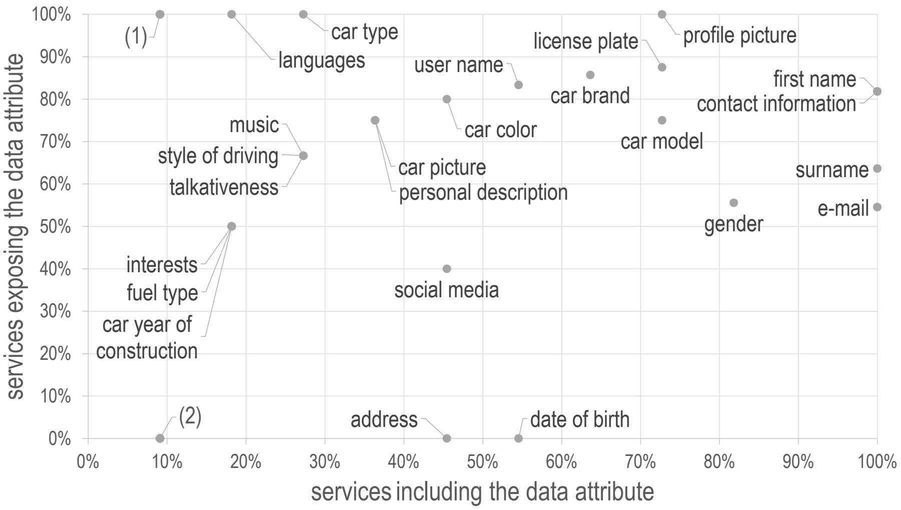


Fig. 2. Services that collect (x-axis) and expose (y-axis) data attributes. Abbreviations: (1) COVID test results and/or vaccination status, job, membership automobile club, phone owner, phone provider (2) marital status, bank account, PayPal account, air-condition, car mileage, country of car registration, fuel consumption

are lacking in almost all instances as the privacy policies provide basic legal terminology. Only a limited number of services mention privacy settings (e.g., change exposure of information towards other users) in the privacy policies.

The analysis also includes the type of disclosure of personal information which can be either *mandatory* or *optional*. On average, 29% of the considered 39 data attributes is mandatory. The rest is optional, which advocates a tendency towards a user-friendly type of collection at a first glance. However, only in few cases is the optional disclosure made transparent, enticing users to disclose more information due to the over-disclose phenomenon [28].

Furthermore, the exposure of each attribute towards other users is either not exposed or exposed on the *profile page* and/or together with the *ride offer*. Certain information is exposed in a reduced fashion; for instance, if the date of birth is disclosed, only the age (in years) is made accessible to other users. Ten data attributes are never displayed for other users, which makes the collection of this information questionable from a user perspective.

3.2 Privacy-Related Features

An analysis of the privacy-related features for RQ 3 shows that only a limited number of shared mobility services offer these features. Moreover, the range and type of these features vary greatly. Some services offer or require the user to complete a process to validate the authenticity of personal information while other services offer privacy settings to change the exposure of certain data. Those

Table 3. Validation of authenticity (V) and privacy settings (S) offered by ride-sharing services

	email		phone number		first name		last name		license plate		vehicle model		driver's license		auto mobile club	
	S	V	S	V	S	V	S	V	S	V	S	V	S	V	S	V
bessermitfahren	x		x													
blablacar		x		x												
clickapoint	x	x	x													
e-carpooling														x		
fahrgemeinschaft	x	x	x				x		x		x					x
foahstmit																
greendrive	x															
mifaz			x		x		x									
mitfahrportal	x		x													
pendlerportal									x							
twogo				x												

can affect the communication with other users, matchmaking among groups, and whether specific information (e.g., email address or phone number) is accessible for others (as shown in Table 3). In almost all services, the availability of profile settings is not communicated to the user at the time of disclosure.

4 Proposed Transparency Enhancing Technology

In order to address the differences in data practices and the improvable degree of transparency, we implemented a tool and tested its functionality with the services included in the prior analysis. We first decided on the degree of personalization and automation (as stated in RQ 4) to be included in the implementation, as referred to in Figure 3. The fact that we choose *low* for both dimensions has a number of reasons. On the one hand, a higher degree of personalization is considerate of the subjective and contextual nature of privacy. However, the loss of privacy accumulates with every disclosure of information [26] and the additional collection and processing of personal preferences pose a risk of (future) privacy infringements. On top of that, the protection of privacy based on collection of personal data is contradictory. On the other hand, a higher degree of automation promises a reduced cognitive burden for the individual but is difficult to achieve especially due to the mentioned subjective nature of privacy, the conception of privacy, and the subsequent decisions. Furthermore, recent research has made apparent that individuals do prefer to remain in control over their privacy-related decisions [9,19,21,24]. In this context, the transparency gained by our approach can be sufficient for individuals to make informed decisions. We achieve this by displaying the imminent consequences of the underlying data practices, since individuals tend to devalue and underestimate decisions and consequences due

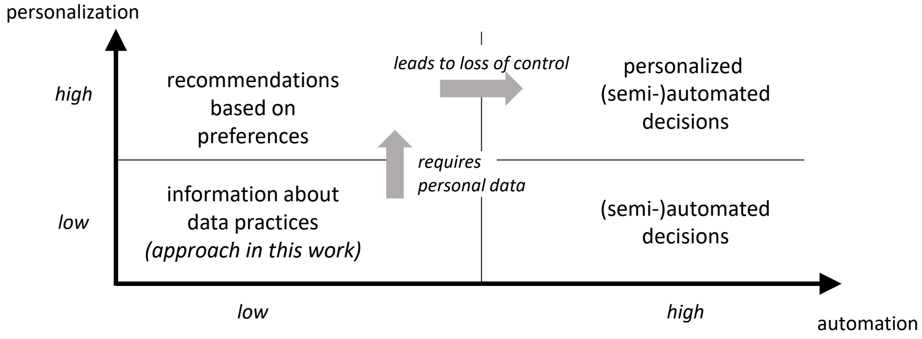


Fig. 3. Degree of personalization and automation of privacy tools

to psychological distance [5]. With that in mind, we chose to design a TET focused on the underlying data practices of the respective ride-sharing service and aimed at increasing the awareness on the side of the individuals. This enables individuals to take more informed decisions and react accordingly if needed, e.g., by not disclosing or purposefully falsifying information.

In order to make the data practices of service providers transparent to individuals, we then decided on the technical approach (formulated as RQ 5). It is important to reduce barriers on the side of the user as far as possible as adoption is difficult to achieve [36]. For that reason, we committed to the premise to focus on a technology that could integrate additional information directly into the website of the ride-sharing service. The corresponding mobile applications of ride-sharing services are closed systems and a third-party app based on the service's API (if available) would have contradicted our premise. Therefore, we chose to implement our idea as a browser extension. This also incorporates the well-established notion that privacy is contextual [2, 25].

Subsequently, we chose an icon-based approach for the design of our tool and the integration of information (for RQ 6), as depicted in Figure 1. Icons have multiple advantages over text- or color-based designs. The memorization of icons works effortlessly as a results of picture superiority [8]. This is crucial, as individuals can only dedicate a limited time for privacy protection [36] and are at the same time confronted with considerable number of privacy-related decisions [47]. Icons are free from linguistic barriers and if used as a standardized set across multiple instances can create comparability [11], in our case between different ride-sharing services. However, the set of icons needs to be small enough to not risk an information overload [11] similar to a notification fatigue which leads to inattentive permission granting since receiving too many notifications has the same effect as receiving no notifications at all [47].

Our tool covers most of the ride-sharing services included in the prior analysis. For each website, additional icons are displayed next to the input fields; a simplified example is depicted in Figure 1. These icons indicate privacy-related information, e.g., whether the corresponding personal information is exposed to

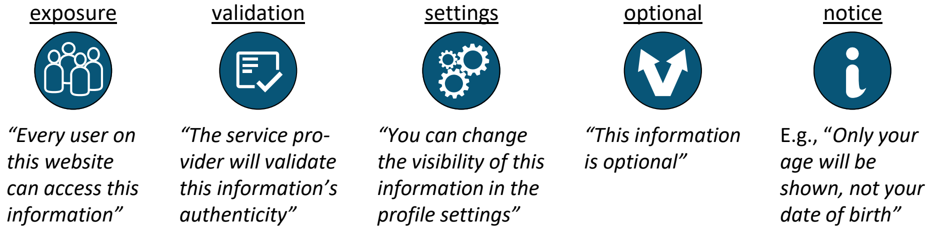

exposure
"Every user on this website can access this information"
validation
"The service provider will validate this information's authenticity"
settings
"You can change the visibility of this information in the profile settings"
optional
"This information is optional"
notice
E.g., "Only your age will be shown, not your date of birth"

Fig. 4. Current set of icons included in the proposed TET and their respective tool-tip

other registered users. In total, the current implementation features five different icons; *exposure*, *validation*, *settings*, *optional*, and *notice* (covering information that does not fit in the four prior categories), as shown in Figure 4. Each icon has a tool-tip explaining its meaning, accessible via mouse-over/touch. The icons were chosen based on prior interviews in which multiple icon options were displayed and the participants interpreted their meaning.

The browser extension communicates with a server to receive the relevant data about websites and icons. In order to keep the data set up-to-date (for RQ 7), we have further implemented feedback functionalities that report input fields to the server. This function only reports strings in the name or id attribute of certain elements in the website, e.g., surname, and does not include any personal information.

5 Scenario-based Online Experiment

After implementing our tool, we conducted a scenario-based online experiment in order to receive a first evaluation of its impact on user behavior and gain feedback on the tool's usability. The experiment was reviewed and approved by the Data Protection Officer of the Clausthal University of Technology. The experiment was conducted between the 10th and 22nd of March 2022.

5.1 Sample

A total of 1093 participants contributed to our study. They were a representation of the German online population and were recruited by a panel provider (certified ISO 20252:2019). The average age of our sample is 44 years and gender distributions are 51% female and 49% male. The majority of participants with education levels 2 and 3 are 40 years or older while education levels 4 and 5 are predominantly young adults (30 years or younger). The full demographic information is presented in Table 7 in the Appendix.

5.2 Setup

In the experiment, we showed the participants a fictional ride-sharing service and asked them to create a personal profile and adjust it to their liking. This includes disclosure of personal information and adjustment of privacy-related profile settings. To capture the disclosure behavior of participants, meta data (i.e., *dirty fields*) was stored during the experiment. Consequently the resulting data set does not include any personal information. After finishing the profile creation we investigated the participants' privacy concerns and asked further questions. Since privacy requires a proxy to be measured [35], we used the IUIPC metric as it is specifically designed for online contexts [23]. We adapted the IUIPC metric to the context of our experiment and re-formulated the questions to match our fictional ride-sharing service. The statements for each question are listed in Table 8 in the Appendix.

To evaluate the effects of our privacy tool, we used a control ($n = 551$) and a test ($n = 542$) group. Both groups' demographics are representative of the German online population. The latter had access to additional visual cues at the time of data disclosure based on our proposed tool. The experiment included the icons indicating the *exposure* of personal data, the *validation* of authenticity, and the availability of user-specific *profile settings*, as shown in Figure 1. An icon marking *optional* data disclosure is not included since existing research demonstrated its effectiveness [21] and we focused on the remaining icons. In addition to that, it is difficult to differentiate between a user's decision on to not disclose information for privacy reasons and the decision on to not disclose information because it is optional. Since the disclosure of *first name* and *email* was mandatory during the experiment, they are not listed in the results.

5.3 Methods

To evaluate the participants' answers to the IUIPC questions and the disclosure behavior, we use logistic additive regression models (ordinal/binary), fitted through R add-on package `mgcv` [30,43–45] with *group* and *gender* being considered as binary factors. For *age* and *education level* we allow for smooth, potentially nonlinear effects, with age-effects being modeled as (penalized) thin plate regression splines (`mgcv` default). For the ordinal education factor, a discrete, second-order smoothing penalty is used as proposed in [13,38].

6 Results

We used the IUIPC metric to measure participant's privacy concerns to answer RQ 8 and collected the disclosure rates of personal information via meta data for RQ 9. On top of that, we asked the participants further questions, to answer RQs 10-13.

6.1 Privacy Concerns

The results show significant differences between the control and test group in the awareness dimension and significance in part of the collection dimension as displayed in Table 4. The difference in awareness indicates a higher degree of users'

Table 4. Logistic regression coefficients for IUIPC metric responses for *group* (test) and *gender* (female) with usual significance codes *** (0.001), ** (0.01), * (0.05).

	cont.1	cont.2	cont.3	coll.1	coll.2	coll.3	awar.1	awar.2
group	0.021	-0.024	0.109	0.085	0.480	0.153	0.628	0.550
p-value	0.846	0.825	0.310	0.430	0.000***	0.153	0.000***	0.000***
gender	-0.062	0.354	0.017	0.318	0.480	0.311	0.245	0.036
p-value	0.573	0.002**	0.875	0.004**	0.000***	0.005**	0.026*	0.741

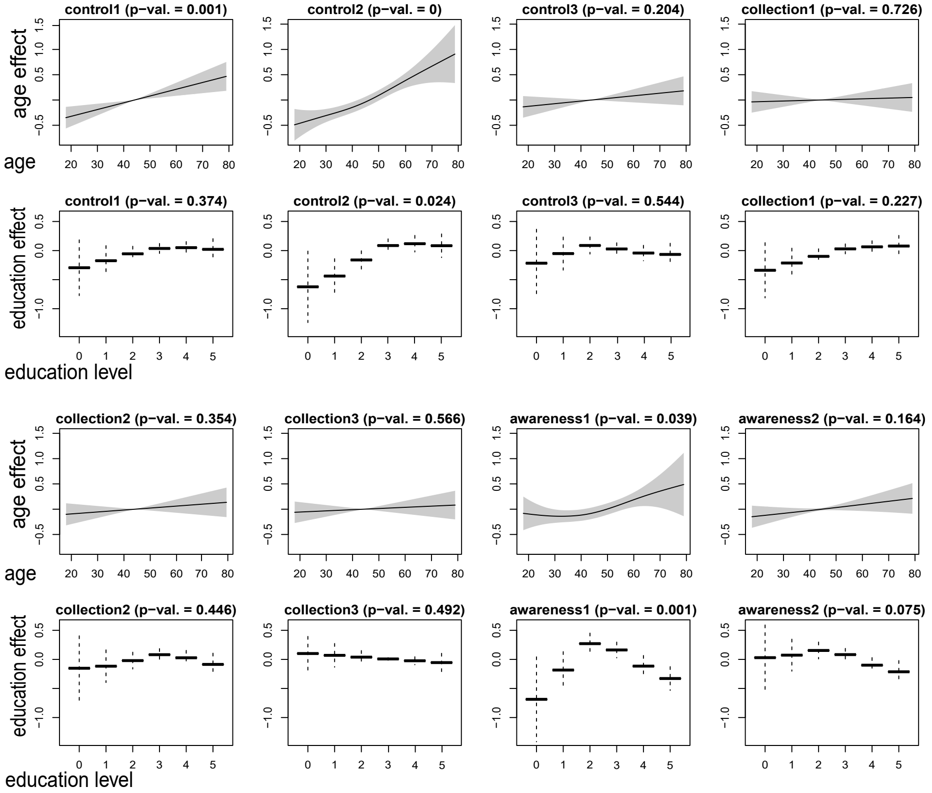


Fig. 5. Effect of *age* and *education level* on the IUIPC metric responses (note, for model identifiability, effects are centered around zero across the data observed [45]). Shaded regions and dashed lines indicate approximate, pointwise 95% confidence intervals.

certainty about the data practices of the service provider, since the respective statements refer directly to the service providers' transparency about their data practices. Additionally, coll.2 shows an increase in decision reflection. The gender variable displays additional, though smaller, increases in awareness (awar.1) and decision reflection (coll.2) for women, but also an increase in expressed discomfort (coll.1) and concern (coll.3), which aligns with prior literature [3, 33].

Higher age is typically associated with stronger agreement to IUIPC statements, but effects vary in terms of size, shape, and significance — see Figure 5. Agreement to general privacy statements (cont.1 & cont.2) shows a (rather) linear increase with age while awar.1 shows increases starting around age 40. Positive association between education level and IUIPC is observed for *cont.1*, *cont.2* and *coll.1*, but with some statistical uncertainty as indicated by p-values and (pointwise) confidence intervals. For *awar.1* and *awar.2*, the association seems to be negative, at least for higher education levels (where uncertainty is lower). For the remaining statements, no clear effects are observed.

6.2 Disclosure Rate

Applying the procedure from subsection 6.1 to the disclosure rates leads to the results summarized in Table 5 and Figure 6. Looking at the differences between control and test group in Table 5 we can see some behavioral changes depending on of data sensitivity and type of icon; particularly *date of birth* and *last name* show a decrease in disclosure rate when combined with the *exposure* icon. In contrast to that, the disclosure of *license plate* increases when it is presented in combination with the *profile settings* icon (which, however, does not apply for the disclosure of *sex*). Furthermore, gender differences show that men tend to disclose less information about themselves but more about their vehicle and vice versa for women. Except for *sex*, higher age is associated with increased disclosure rates, but with varying effect sizes — see Figure 6. With respect to education, higher levels show lower disclosure rates of information about the individual. In contrast to this, participants with higher levels of education tend to disclose more information about their vehicles.

6.3 Icon Recognizability & Understandability

Next, for RQ 10, we evaluate the recognizability of the used icons. Therefore, we ask all participants in the test group to identify a given icon. In addition, we ask participants to recall who was able to access one piece of personal information they had disclosed during the profile creation. Our results show that more participants are able to correctly recall privacy-related information linked to one of

Table 5. Logistic regression coefficients for disclosure rate for *group* (test) and *gender* (female) with usual significance codes *** (0.001), ** (0.01), * (0.05). Attached icons for test group: 1 = *exposure*, 2 = *validation*, 3 = *settings*

	address	date of birth[1]	driver's license[2]	last name[1]	sex[3]	vehicle color[1]	license plate[3]	vehicle model[2]
group	-0.102	-0.352	-0.065	-0.507	0.065	-0.149	0.328	-0.073
p-value	0.500	0.016*	0.671	0.004**	0.668	0.379	0.010*	0.681
gender	0.581	0.202	0.450	0.595	0.588	-0.341	-0.273	-0.465
p-value	0.000***	0.178	0.004**	0.001**	0.000***	0.053	0.037*	0.013*

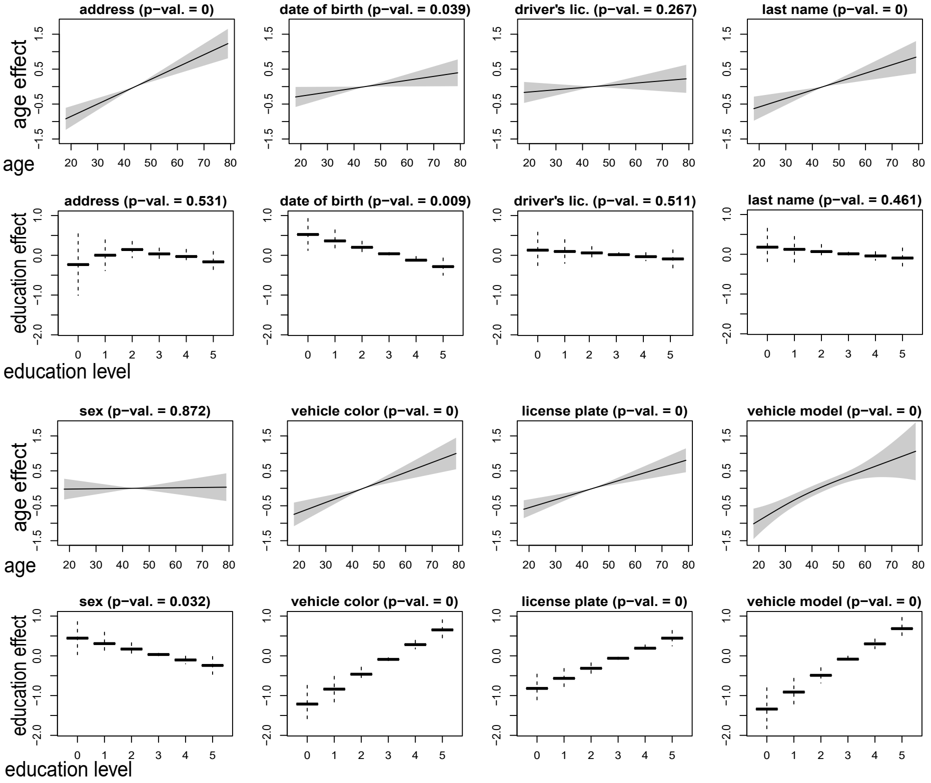


Fig. 6. Effect of *age* and *education level* on the disclosure rate (note, for model identifiability, effects are centered around zero across the data observed [45]). Shaded regions and dashed lines indicate approximate, pointwise 95% confidence intervals.

their personal information; 28% of participants identified the given icon correctly and 37% correctly recalled who was able to access their personal information. Moreover, we ask participants to state how often the additional privacy-related information did influence their disclosure decision. In total, 55% stated it had an influence on their decision (combining the answers *sometimes*, *most of the time*, and *always*), with 14% being unsure and 31% stating *never.*

6.4 Data Disclosure

Then, additionally, the participants had to reflect on the amount of personal information they had disclosed during the experiment (for RQ 11). This question was raised to both groups and the answers indicate differences; in the control group 19% are able to correctly state how many attributes of personal information they had disclosed compared to 28% in the test group. This indicates improved – yet small – reminiscence and awareness. While these numbers remain relatively low they are higher than usually recorded in privacy research with failure rates of above 90% [17].

6.5 Use of Profile Settings

We note a similar difference in the data for RQ 12. As the icon set includes one icon specifically dedicated to the availability of profile settings, we evaluated the usage of these settings or (in case they were not adjusted by the participant) asked participants if they had perceived them. In the control group, 26% used or at least perceived the profile settings compared to 35% in the test group.

6.6 Information Usefulness

Finally, for RQ 13, we asked the participants to state the additional information which they regard most useful (or in case of the control group; which they would most liked to have), presenting them multiple options, as stated in Table 6. The results show interesting differences. While both groups state a piece of information's exposure as most useful, the test group shows noteworthy changes in the distribution of answers. The option for information exposure scores about 20% fewer answers while, at the same time, the number of uncertain answers more than doubles. This supports the theory that individuals are often unable to correctly evaluate their own privacy preferences as, once explicitly confronted with a given scenario, the individual's evaluation of helpful information shifts.

7 Limitations

Some shortcomings and limitations are worth mentioning. Firstly, the analysis of ride-sharing services is limited to the user perspective. That means that internal data processing and further data practices on side of the service provider are not included. To complement our analysis with the provider's side remains a task of our ongoing research. Secondly, the participants were notified twice before participating in the experiment that any disclosed personal data would not be stored. Though the term privacy was not mentioned explicitly to the participants, this possibly affected some participants in their disclosure decisions but was inevitable as this notification was required by the panel provider. Thirdly,

Table 6. Answers about most useful information

	control group	test group
If my data was exposed to others	44%	35%
If the service provider offered to validate my data	15%	11%
If profile settings for this information were available	10%	10%
If I should not provide a piece of information	17%	13%
I am not sure	15%	32%

our data indicate that participants with a higher level of education are more likely to disclose information about their vehicle. However, we did not ask participants about the possession of a vehicle, which could be a mediating factor. Lastly, the combinations of data attributes and icon type could be changed to examine if different combinations of information and icon type yield insights about individuals' disclosure behavior.

8 Conclusion and Outlook

With this work, we contribute to the current state of privacy research in the sharing economy. Firstly, we analyzed the current practices of ride-sharing services and uncover part of their data practices. Secondly, we proposed and implemented a TET in form of a browser extension that is capable of integrating additional information seamlessly into the services' websites and support individuals to make more informed decisions. Thirdly, we conducted a scenario-based online experiment with two representative samples for the test and control groups to evaluate our tool. For this experiment, we used a fictional ride-sharing service and asked participants to create a profile with their personal information. Based on our findings, we can confirm that a higher degree of transparency of data practices does not necessarily lead to less disclosed information. Consequently, we recommend service providers to offer profile settings for sensitive personal information and illustrate data practices which apply to their service and website more clearly. The direction of our future research includes the provider's perspective on the data practices and the overall affect data practices have on the individual's choice of service when they are able to compare services with the help of our tool. In addition to that, our prior analysis of ride-sharing services only covered the countries Germany, Austria, and Switzerland. Therefore, broader cultural differences and their effects on privacy concerns and disclosure behavior are not accounted for in this work.

Appendix

Table 7. Demographic data of respondents

%	gender	%	age	%	education
51	female	3	18-20	1	not finished school (yet)
49	male	20	21-30	6	primary school certificate
		20	31-40	21	primary school certificate & vocational training
		19	41-50	32	secondary school certificate or equivalent
		23	51-60	19	higher education entrance qualification
		15	>60	21	higher education

Table 8. Context-specific formulations of IUIPC metric

question	statement
cont.1	My privacy is really a matter of my right to exercise control and autonomy over how MyCarPool collects, uses, and shares my information
cont.2	The control of my personal information lies at the heart of my privacy
cont.3	I believe that MyCarPool has taken or reduced my control over my data as a result of a marketing transaction
coll.1	It bothered me when MyCarPool asked me for personal information
coll.2	When MyCarPool asked me for personal information, I sometimes thought twice before providing it
coll.3	I am concerned that MyCarPool collected too much personal information about me
awar.1	MyCarPool did disclose the way my data are collected, processed, and used
awar.2	I was aware and knowledgeable about how MyCarPool uses my personal information

References

1. Acquisti, A., Grossklags, J.: Privacy Attitudes and Privacy Behavior. In: Camp, L.J., Lewis, S. (eds.) Economics of Information Security, Advances in Information Security, vol. 12, pp. 165–178. Kluwer Academic Publishers, Boston (2004). https://doi.org/10.1007/1-4020-8090-5_13
2. Acquisti, A., Grossklags, J.: Privacy and rationality in individual decision making. IEEE Security & Privacy **3**(1), 26–33 (2005). https://doi.org/10.1109/MSP.2005.22
3. Acquisti, A., John, L.K., Loewenstein, G.: The Impact of Relative Standards on the Propensity to Disclose. Journal of Marketing Research **49**(2), 160–174 (2012). https://doi.org/10.1509/jmr.09.0215

4. Agogo, D.: Invisible market for online personal data: An examination. Electronic Markets **31**(4), 989–1010 (2021). https://doi.org/10.1007/s12525-020-00437-0
5. Bandara, R., Fernando, M., Akter, S.: Is the Privacy Paradox a Matter of Psychological Distance? An Exploratory Study of the Privacy Paradox from a Construal Level Theory Perspective. In: Proceedings of the 51st Hawaii International Conference on System Sciences (2018). https://doi.org/10.24251/HICSS.2018.465
6. Beke, F.T., Eggers, F., Verhoef, P.C.: Consumer Informational Privacy: Current Knowledge and Research Directions. FNT in Marketing (Foundations and Trends in Marketing) **11**(1), 1–71 (2018). https://doi.org/10.1561/1700000057
7. Bélanger, F., Crossler, R.E.: Privacy in the Digital Age: A Review of Information Privacy Research in Information Systems. MIS Quarterly **35**(4), 1017 (2011). https://doi.org/10.2307/41409971
8. Childers, T.L., Houston, M.J.: Conditions for a Picture-Superiority Effect on Consumer Memory. Journal of Consumer Research **11**(2), 643 (1984). https://doi.org/10.1086/209001
9. Colnago, J., Feng, Y., Palanivel, T., Pearman, S., Ung, M., Acquisti, A., Cranor, L.F., Sadeh, N.: Informing the Design of a Personalized Privacy Assistant for the Internet of Things. In: Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems, pp. 1–13. Association for Computing Machinery, New York, NY, USA (2020). https://doi.org/10.1145/3313831.3376389
10. De, S.J., Le Metayer, D.: Privacy Risk Analysis to Enable Informed Privacy Settings. In: 2018 IEEE European Symposium on Security and Privacy Workshops (EuroS&PW). pp. 95–102. IEEE (2018). https://doi.org/10.1109/EuroSPW.2018.00019
11. Efroni, Z., Metzger, J., Mischau, L., Schirmbeck, M.: Privacy Icons: A Risk-Based Approach to Visualisation of Data Processing. European Data Protection Law Review **5**(3), 352–366 (2019). https://doi.org/10.21552/edpl/2019/3/9
12. Entega Plus GmbH: Mitfahrgelegenheit und Co.: Fahrgemeinschaft 2.0, https://www.entega.de/blog/fahrgemeinschaft-die-wichtigsten-onlinemitfahrportale/
13. Gertheiss, J., Scheipl, F., Lauer, T., Ehrhardt, H.: Statistical inference for ordinal predictors in generalized additive models with application to bronchopulmonary dysplasia. BMC research notes **15**(1), 112 (2022). https://doi.org/10.1186/s13104-022-05995-4
14. Hann, I.H., Hui, K.L., Lee, S.Y.T., Png, I.P.: Overcoming Online Information Privacy Concerns: An Information-Processing Theory Approach. Journal of Management Information Systems **24**(2), 13–42 (2007). https://doi.org/10.2753/MIS0742-1222240202
15. Hesselmann, C., Gertheiss, J., Müller, J.P.: Ride Sharing & Data Privacy: How Data Handling Affects the Willingness to Disclose Personal Information. Findings (2021). https://doi.org/10.32866/001c.29863
16. Jackson, C.B., Wang, Y.: Addressing The Privacy Paradox through Personalized Privacy Notifications. Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies **2**(2), 1–25 (2018). https://doi.org/10.1145/3214271
17. Kamleitner, B., Sotoudeh, M.: Information sharing and privacy as a socio-technical phenomenon. TATuP - Zeitschrift für Technikfolgenabschätzung in Theorie und Praxis **29**(3), 68–71 (2019). https://doi.org/10.14512/tatup.28.3.68

18. Kariryaa, A., Savino, G.L., Stellmacher, C., Schöning, J.: Understanding Users' Knowledge about the Privacy and Security of Browser Extensions. In: Seventeenth Symposium on Usable Privacy and Security (SOUPS 2021). pp. 99–118 (2021)
19. Kitkowska, A., Shulman, Y., Martucci, L.A., Wästlund, E.: Facilitating Privacy Attitudes and Behaviors with Affective Visual Design. In: ICT Systems Security and Privacy Protection, vol. 580, pp. 109–123. (2020). https://doi.org/10.1007/978-3-030-58201-2_8
20. Kitkowska, A., Warner, M., Shulman, Y., Wästlund, E., Martucci, L.A.: Enhancing Privacy through the Visual Design of Privacy Notices: Exploring the Interplay of Curiosity, Control and Affect. In: Sixteenth Symposium on Usable Privacy and Security (SOUPS 2020). pp. 437–456. USENIX Association (2020)
21. Krol, K., Preibusch, S.: Control versus Effort in Privacy Warnings for Webforms. In: Proceedings of the 2016 ACM on Workshop on Privacy in the Electronic Society - WPES'16. pp. 13–23. ACM Press, New York, New York, USA (2016). https://doi.org/10.1145/2994620.2994640
22. Machuletz, D., Laube, S., Böhme, R.: Webcam Covering as Planned Behavior. In: Mandryk, R., Hancock, M., Perry, M., Cox, A. (eds.) Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems. pp. 1–13. ACM, New York, NY, USA (2018). https://doi.org/10.1145/3173574.3173754
23. Malhotra, N.K., Kim, S.S., Agarwal, J.: Internet Users' Information Privacy Concerns (IUIPC): The Construct, the Scale, and a Causal Model. Information Systems Research **15**(4), 336–355 (2004). https://doi.org/10.1287/isre.1040.0032
24. Marsch, M., Grossklags, J., Patil, S.: Won't You Think of Others? Interdependent Privacy in Smartphone App Permissions. Proceedings of the ACM on Human-Computer Interaction **5**(CSCW2), 1–35 (2021). https://doi.org/10.1145/3479581
25. Nissenbaum, H.: Privacy in Context: Technology, Policy, and the Integrity of Social Life. Stanford University Press, Stanford, CA (2009). https://doi.org/10.1515/9780804772891
26. Nissim, K., Wood, A.: Is privacy privacy? Philosophical transactions. Series A, Mathematical, physical, and engineering sciences **376**(2128) (2018). https://doi.org/10.1098/rsta.2017.0358
27. Olejnik, L., Englehardt, S., Narayanan, A.: Battery status not included: Assessing privacy in web standards. CEUR Workshop Proceedings **1873**, 17–24 (2017)
28. Preibusch, S., Krol, K., Beresford, A.R.: The Privacy Economics of Voluntary Over-disclosure in Web Forms. In: Böhme, R. (ed.) The Economics of Information Security and Privacy, pp. 183–209. Springer Berlin Heidelberg, Berlin, Heidelberg (2013). https://doi.org/10.1007/978-3-642-39498-0_9
29. Proserpio, D., Xu, W., Zervas, G.: You get what you give: theory and evidence of reciprocity in the sharing economy. Quantitative Marketing and Economics **16**(4), 371–407 (2018). https://doi.org/10.1007/s11129-018-9201-9
30. R-Core-Team: R: a Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria (2022), https://www.R-project.org/
31. Ranzini, G., Etter, M., Lutz, C., Vermeulen, I.E.: Privacy in the Sharing Economy. SSRN Electronic Journal (2017). https://doi.org/10.2139/ssrn.2960942
32. Rath, D.K., Kumar, A.: Information privacy concern at individual, group, organization and societal level - a literature review. Vilakshan - XIMB Journal of Management **18**(2), 171–186 (2021). https://doi.org/10.1108/XJM-08-2020-0096

33. Reinhardt, D., Khurana, M., Hernández Acosta, L.: I still need my privacy: Exploring the level of comfort and privacy preferences of German-speaking older adults in the case of mobile assistant robots. Pervasive and Mobile Computing **74**, 101397 (2021). https://doi.org/10.1016/j.pmcj.2021.101397
34. Schor, J., et al.: Debating the sharing economy. Journal of Self-Governance and Management Economics **4**(3), 7–22 (2014)
35. Smith, J.H., Dinev, T., Xu, H.: Information Privacy Research: An Interdisciplinary Review. MIS Quarterly **35**(4), 989 (2011). https://doi.org/10.2307/41409970
36. Story, P., Smullen, D., Yao, Y., Acquisti, A., Cranor, L.F., Sadeh, N., Schaub, F.: Awareness, Adoption, and Misconceptions of Web Privacy Tools. Proceedings on Privacy Enhancing Technologies **2021**(3), 308–333 (2021). https://doi.org/10.2478/popets-2021-0049
37. Teubner, T., Flath, C.M.: Privacy in the Sharing Economy. Journal of the Association for Information Systems pp. 213–242 (2019). https://doi.org/10.17705/1jais.00534
38. Tutz, G., Gertheiss, J.: Regularized regression for categorical data. Statistical Modelling **16**(3), 161–200 (2016). https://doi.org/10.1177/1471082X16642560
39. Utopia GmbH: Die besten Mitfahrgelegenheiten, https://utopia.de/ratgeber/mitfahrgelegenheiten/
40. VCS Verkehrs-Club der Schweiz: Carpooling, https://www.verkehrsclub.ch/ratgeber/auto/autoteilen/carpooling/
41. VGL Verlagsgesellschaft: Mitfahrzentralen im Vergleich, https://www.vergleich.org/mitfahrzentrale/
42. Wang, Y., Wang, S., Wang, J., Wei, J., Wang, C.: An empirical study of consumers' intention to use ride-sharing services. Transportation **47**(1), 397–415 (2020). https://doi.org/10.1007/s11116-018-9893-4
43. Wood, S.N.: On p-values for smooth components of an extended generalized additive model. Biometrika **100**(1), 221–228 (2013). https://doi.org/10.1093/biomet/ass048
44. Wood, S.N.: Fast stable restricted maximum likelihood and marginal likelihood estimation of semiparametric generalized linear models. Journal of the Royal Statistical Society: Series B (Statistical Methodology) **73**(1), 3–36 (2011). https://doi.org/10.1111/j.1467-9868.2010.00749.x
45. Wood, S.N.: Generalized Additive Models. Chapman and Hall/CRC (2017). https://doi.org/10.1201/9781315370279
46. Yao, M.Z., Rice, R.E., Wallis, K.: Predicting user concerns about online privacy. Journal of the American Society for Information Science and Technology **58**(5), 710–722 (2007). https://doi.org/10.1002/asi.20530
47. Zhang, S., Feng, Y., Bauer, L., Cranor, L.F., Das, A., Sadeh, N.: Did you know this camera tracks your mood? Proceedings on Privacy Enhancing Technologies **2021**(2), 282–304 (2021). https://doi.org/10.2478/popets-2021-0028