University of Warsaw
Faculty of Economic Sciences

Working Papers
No. 14/2021 (362)

PRIVACY TRADE-OFFS IN THE RIDE-HAILING SERVICES

Michał Paliński

Warsaw 2021

WORKING PAPERS 14/2021 (362)

Privacy trade-offs in the ride-hailing services

Michał Paliński

University of Warsaw, Faculty of Economic Sciences; Digital Economy Lab (DELab UW)
m.palinski@delab.uw.edu.pl

Abstract: We test for users readiness for co-financing ride-hailing service with their personal data applying a Discrete Choice Experiment. We design an experiment in which respondents are asked to choose between hypothetical app-based taxi rides which offered discounts as a compensation for intruding their privacy and a regular service. Our analysis compare how awareness of rights stemming from GDPR affects respondent's privacy preferences. Cross-group analysis indicate that reminding users about their rights stemming from the GDPR significantly increases their valuation of personal data. The results of WTA analysis suggest that there is a market for "pay with your data" business models.

Keywords: economics of privacy, mobile apps, DCE, mixed logit, WTA

JEL codes: C25, D12, L51

Acknowledgements: This research was supported by National Science Centre, Poland (grant number: 2017/25/N/HS4/01214.

1 Introduction

Privacy management became a fundamental aspect of our daily online activity. By selecting what data to disclose and to whom, people build and manage their online identities (Sayaf 2016). In light of this, it is utmost worrisome that users often do not feel in control of their data and express concern about it (European Commission 2019; Auxier et al. 2020).

On the other hand, it is well established that declared privacy attitudes often are not consistent with the real behavior of users, even privacy concerned ones, who widely disclose personal data online in trade-off for convenience or small remuneration (Acquisti 2004; Beresford et al. 2012). This ambivalent attitude, known as privacy paradox, is a contextual phenomenon which has several causes. One of them is the lack of proper instruments to control the sharing of online personal data leading to pervasive digital resignation (Kokolakis 2017; Draper and Turow 2019).

On May 2018 the new EU General Data Protection Regulation (GDPR) came into force aiming at enhancing users' control over online privacy (van Ooijen and Vrabec 2019). Although GDPR has a high recognition among Europeans (67% heard of the regulation), it is still poorly understood what kind of users' rights it exactly entails as half of the respondents who heard of it does not know what it is (European Commission 2019).

In this study we test for users readiness for co-financing online services with their personal data through conducting a Discrete Choice Experiment (DCE). We design an experiment in which respondents are asked to choose between hypothetical app-based taxi rides which offer discounts as a compensation for intruding their privacy and a regular service. Our survey compares two treatment groups: one which is explicitly informed about main mechanisms and legal provisions of GDPR in relation to data shared in the hypothetical scenarios, and the second one which is not reminded about them. Following behavioral privacy literature, we briefly consider how our results resonate with psychological factors such as information asymmetry, present bias or illusion of control affecting privacy decisions (Acquisti et al. 2020). We hypothesise that inclusion of the GDPR notice in the experiment should decrease respondents valuation of privacy. We assume that privacy management mechanisms stemming from the regulation would boost respondents' confidence about effective data control.

A growing body of literature aims to quantitatively analyse privacy preferences using DCE (e.g. Daly et al. 2012; Potoglou et al. 2015; Glasgow et al. 2020; Goad et al. 2020). We

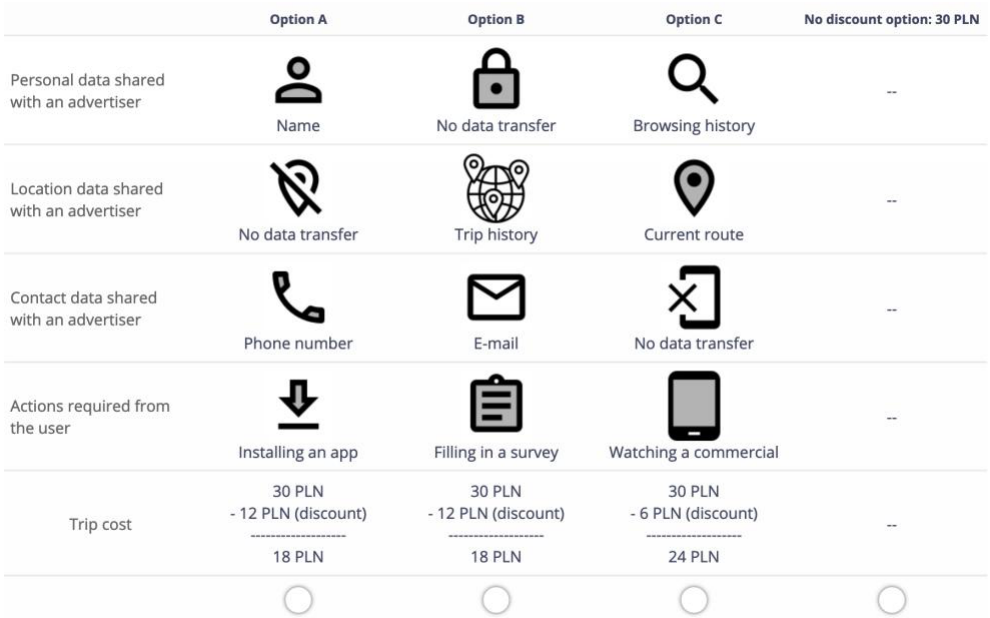
add to this literature by assessing how awareness of rights stemming from the GDPR affects respondent's willingness to accept discounts for sharing personal data.

2 The Discrete Choice Experiment

Survey was conducted online by a professional polling agency in Q3 2020 among 400 participants from major Polish agglomerations who stated that they used an app-based ride-hailing service at least once. We have confirmed representativeness of the sample regarding the basic privacy attitudes by a cross-check with the most recent Eurobarometer study (European Commission 2019). Respondents were presented with eight choice tasks, each having four options: three hypothetical scenarios and the status quo (no discount option). In all tasks respondents were asked to choose the most suitable option.

Hypothetical scenarios described in the questionnaire announce an appearance of a new type of service which would offer a discount for a ride in exchange for data sharing with a third party or performing a privacy sensitive action. Hypothetical options (A-C) varied with respect to privacy intrusiveness and the discount value. Respondents were informed that choosing options A-C would entail sharing data with an advertising company which will use it for marketing profiling. In selecting the most preferred ride, respondents considered three categories of data shared with an advertiser regarding personal data (e.g. name or social media likes), location data (e.g. current trip's route), contact data (e.g. e-mail) and actions required during the trip (e.g. installing an advertised app). An example of a choice card is presented in Figure 1 and attributes levels are shown in Table 3.

Figure 1: Example of a choice card (translation)



Furthermore, survey respondents were randomly assigned to one of the two treatment groups. Members of the first group were reminded about their rights stemming from the GDPR while the other group was directly presented with the choice cards. In particular, the first group was informed that i) they will be able to restrict, modify or ask for deleting their data acquired by a third party company; ii) they will be able to uninstall the required app at any time; iii) and that their personal data will be processed in compliance with the GDPR.

In this study, we applied an efficient experimental design optimized with respect to D-Error. We obtained priors from declared reservation prices collected in the pilot phase of the survey. Experimental design was recalibrated three times throughout the main study in order to obtain more accurate information about respondents' preferences. Price was coded as a continuous variable, while all the other attributes were coded as dummy variables relative to a baseline category (i.e. no data sharing).

3 Results

3.1 MXL estimation

Our final dataset consisted of 2749 observations related to 345 respondents (55 respondents were excluded from the final study due to a short time of filling in a survey). We used these data to estimate mixed-logit models (MXL) for a pooled sample and two treatments: with and without the GDPR notice (Train and Weeks 2005). We assumed that all the 24 preference parameters were random, following normal distributions and lognormal distribution for the discount coefficient. We used 1000 Sobol draws with Owen scrambling for estimating parameters and their standard errors (Owen 1995). We assumed the following form of the utility function of respondent $i \in I$ from choosing alternative $j \in J$ in choice situation $t \in T$ (time subscript is suppressed):

$$U_{ij} = \beta_{1i} PD_{ij} + \beta_{2i} LD_{ij} + \beta_{3i} CD_{ij} + \beta_{4i} UA_{ij} + \beta_{5i} Discount_{ij} + \boldsymbol{\varepsilon}_{ij} \quad (1)$$

where β is the vector of parameters associated with their respective variables and ε_{ij} is a random component of utility associated with alternative j. We have tested a model specification which included alternative specific constant (ASC) for the SQ scenario. ASC parameter was large, negative and significant, which means that on average respondents derive utility from choosing rides offering discounts for data sharing. Other estimates followed similar path as the ones reported in Table 1. We chose the specification without ASC as it allowed us to deal with the problem of confounding SQ and zero price effects (Tjiong et al. 2019). The estimation results for all three MXL models are reported in Table 1 and variables' names are explained in Table 3.

Most of the MXL parameters for the pooled sample are highly significant with the exception of sharing birth date, browsing and trips history with an advertiser. Standard deviations of all parameters are significant which suggests a considerable individual heterogeneity of preferences in the sample. Parameters related to data sharing variables (PD, LD, CD) have expected negative signs, indicating that, on average, these attributes were viewed as undesirable. Ranking the significant parameters by their values reveals that respondents are much more concerned about sharing their social media likes, names or email addresses with a third party company than about sharing the route of their taxi ride. It shows us that respondents tend to underestimate the sensitivity of location data. It is in line with the previous studies which reported that users are willing to share location of their smartphone (Danezis et al. 2005; Glasgow et al. 2020) or IoT device (Goad et al. 2020) for a small remuneration or discount.

Interestingly, all three parameters of variables related to privacy intrusive actions (UA) are positive and significant. This result suggests that respondents are willing to install advertised application, participate in a marketing survey or watch a commercial on their smartphones during a taxi ride in exchange for a discount. From the fact that these actions do not decrease respondents' utility we infer that they treat performing them as a fair trade for a discounted ride.

Overview of choices shows us that SQ option entailing no discount and no privacy intrusion was chosen only in around 12% of cases and its popularity did not vary significantly across our treatment groups. It suggests that there is a market for "pay with your data" business models in Poland.

Results of MXL models estimation for two treatment samples follow a similar pattern in terms of parameters' signs and their relative importance. In order to compare choice models generated from two subgroups of respondents we need to consider differences in unobserved variability between the data sources which is often referred as a scale parameter (Hess and Train 2017). We used the Swait-Louviere test to reject a hypothesis that both our subgroups share coefficients that are statistically indifferent (Swait and Louviere 1993). The test result implies that preference and attribute valuation (WTA) are heterogeneous and vary across two respondent groups considered in our study. The most important insights from our study come from the comparison of WTA values obtained from these two scenarios, which we describe in the next subsection.

Table 1: Estimation results for different choice models

	Pooled sample				GDPR condition				No condition			
	Mean		SD		Mean		SD		Mean		SD	
Variable	Coef.	SE	Coef.	SE	Coef.	SE	Coef.	SE	Coef.	SE	Coef.	SE
pd.name	-0.594***	(0.113)	0.893***	(0.15)	-0.88***	(0.2)	1.261***	(0.249)	-0.43***	(0.139)	0.641***	(0.21)
pd.birth	-0.147	(0.104)	0.782***	(0.154)	-0.12	(0.16)	0.946***	(0.213)	-0.208	(0.138)	0.606**	(0.249)
pd.sm	-0.735***	(0.123)	1.05***	(0.163)	-0.821***	(0.186)	1.112***	(0.235)	-0.688***	(0.162)	0.94***	(0.227)
pd.search	-0.035	(0.107)	0.973***	(0.133)	-0.17	(0.156)	0.878***	(0.207)	0.086	(0.147)	0.975***	(0.187)
ld.route	-0.218***	(0.081)	0.499***	(0.135)	-0.229*	(0.121)	0.5**	(0.207)	-0.268**	(0.116)	0.666***	(0.162)
ld.history	-0.04	(0.082)	0.67***	(0.11)	-0.08	(0.13)	0.829***	(0.159)	-0.057	(0.108)	0.618***	(0.145)
cd.email	-0.429***	(0.086)	0.77***	(0.114)	-0.381***	(0.134)	0.829***	(0.184)	-0.465***	(0.118)	0.785***	(0.158)
cd.phone	-0.159*	(0.086)	0.796***	(0.115)	-0.01	(0.13)	0.744***	(0.194)	-0.262**	(0.117)	0.812***	(0.144)
ua.app	0.36***	(0.097)	0.786***	(0.136)	0.419***	(0.152)	0.94***	(0.206)	0.311**	(0.126)	0.611***	(0.221)
ua.ad	0.19**	(0.093)	0.532***	(0.158)	0.21	(0.148)	0.736***	(0.21)	0.15	(0.123)	0.438**	(0.207)
ua.survey	0.49***	(0.09)	0.613***	(0.14)	0.588***	(0.143)	0.844***	(0.181)	0.393***	(0.12)	0.489**	(0.215)
discount[a]	-0.44***	(0.132)	1.558***	(0.136)	-0.563***	(0.209)	1.668***	(0.206)	-0.295*	(0.169)	1.513***	(0.186)
LL	-3118.242				-1453.061				-1652.571			
ρ^2	0.1818				0.1805				0.1891			
AIC	6284.48				2954.12				3353.14			
n (obs.)	2749				1279				1470			

3.2 WTA estimation

With linear utility function we can calculate a respondent's willingness-to-accept (WTA) remuneration for a change in a given attribute. It can be defined as the ratio between the parameter of interest and the discount attribute multiplied by -1 (Bliemer and Rose 2013). Standard errors for the derived WTA values were obtained using the Delta method (Daly et al. 2012)[1]. Comparing WTA values between the different attributes and their levels helps us identify which aspects of online privacy respondents find the most sensitive. Positive WTA estimates imply that respondents would accept sharing a particular information or performing privacy intrusive action only when being offered a discount. The exact WTA values should not be overinterpreted given that stated preference experiments suffer from hypothetical bias (Colombo et al. 2020). We tried to mitigate this issue by recruiting respondents using ride-hailing services.

Intuitively, we could expect that the GDPR notice might give respondents a sense of control over their privacy and encourage them to share more in exchange for smaller discounts. Interestingly, the reverse has proven to be the case. Results presented in Table 2 indicate that respondents who were reminded about their rights stemming from the GDPR require significantly larger discounts for accepting all types of data sharing. We disregard WTA values for phone number and route history which are statistically insignificant.

This difference is most noticeable in the case of sharing such personal data as name and social media likes (WTA in the GDPR notice group was respectively 60% and 36% higher in comparison to the group not reminded about the GDPR). On the other hand, this group of respondents was more likely to choose alternatives which entailed some kind of privacy intrusive action.

Privacy literature provides rich evidence of psychological factors impairing users' privacy decisions (Acquisti et al. 2020). Our results suggest that inclusion of plain-language explanation of user's data processing[2] might effectively diminish information asymmetry between data

[1] All models were estimated in the preference space with a use of the Apollo package in R (Hess and Palma,2019). Estimation in preference space as opposed to WTP/WTA space results in better fit of MXL models (Hole and Kolstad 2012). On the other hand, estimation in the preference space and subsequent calculation of the moments and confidence intervals of WTP/WTA using the ratio of parameters has been proven problematic. Delta method employed in this study for WTA calculation tends to produce misleading standard error estimates unless the cost/discount parameter is highly significant (Carson and Czajkowski 2019).

[2] Envisioned by GDPR's right to information: Before any processing of personal data takes place, a data subject has to be informed, among others, about the purposes for which data will be processed, about the data controller's identity, about the recipients of his personal data and about the period of data storage (GDPR articles 14-15).

subjects (consumers), and data controllers (companies) and as a result decrease the latter's ability to collect and utilize users' personal data at low or near zero cost.

Moreover, results of our experiment show no indication of an illusion of control effect i.e. an increased risk-taking situation in which users divulge more data when given a more granular control over their privacy settings (Brandimarte et al. 2013). On contrary, it seems that the GDPR reminder might to some extent mitigate the present bias under which consumers accept small benefits in exchange for their data because of overemphasizing immediate, and underweighting delayed, costs and benefits (Acquisti et al. 2020). However, we need to take into account that this result might suffer from the presence of social desirability bias under which respondents downplay their willingness to trade privacy (Braunstein et al. 2011).

Table 2: WTA estimation (results in PLN/trip)

	Pooled sample		GDPR condition		No condition	
Variable	Coef.	Rob. SE	Coef.	Rob. SE	Coef.	Rob. SE
pd.name	0.923***	(-0.229)	1.545***	(-0.57)	0.577***	(-0.224)
pd.birth	0.228	(-0.169)	0.211	(-0.286)	0.279	(-0.198)
pd.sm	1.141***	(-0.278)	1.441***	(-0.536)	0.924***	(-0.285)
pd.search	0.055	(-0.176)	0.3	(-0.304)	-0.116	(-0.219)
ld.route	0.339**	(-0.144)	0.402	(-0.258)	0.359**	(-0.177)
ld.history	0.062	(-0.129)	0.145	(-0.228)	0.076	(-0.147)
cd.email	0.666***	(-0.17)	0.668**	(-0.325)	0.624***	(-0.185)
cd.phone	0.247*	(-0.14)	0.017	(-0.234)	0.352**	(-0.165)
ua.app	-0.559***	(-0.198)	-0.736*	(-0.404)	-0.417**	(-0.196)
ua.ad	-0.295*	(-0.162)	-0.374	(-0.309)	-0.202	(-0.169)
ua.survey	-0.761***	(-0.212)	-1.032**	(-0.476)	-0.528***	(-0.2)

At 1 PLN 0.22 EUR
*** p value < 1%, ** p value in [1, 5%), * p value in [5, 10%)

4 Conclusions

Our experiment indicates that there might be a positive demand for a new type of online services which explicitly offer discounts in exchange for privacy intrusion. We found that our respondents are much more concerned about sharing their social media likes, names or email addresses with a third party company than about sharing their location data. On the other hand,

respondents were likely to choose alternatives which entailed installing advertised application, participating in a marketing survey or watching a commercial on smartphones during a taxi ride even for relatively smaller discounts. A considerable heterogeneity of preferences discovered in our sample requires further examination.

Contrary, to our intuition our experiment revealed that informing respondents about their privacy rights guaranteed by the GDPR does not lead to the illusion of control effect. Our results suggest that broader awareness about this regulation might contribute to development of more user-centric online services.

Appendix

Table 3: List of attributes and their levels

Attribute	Attribute's levels	Variable name
Personal data (PD)	Name	pd.name
	Birth date	pd.birth
	Social media likes	pd.sm
	Search history	pd.search
	No data	(baseline)
Location data (LD)	Current route	ld.route
	Routes' history	ld.history
	No data	(baseline)
Contact data (CD)	E-mail address	cd.email
	Phone number	cd.phone
	No data	(baseline)
User's action (UA)	Watching an ad	ua.ad
	Installing an app	ua.app
	Filling in a survey	ua.survey
	No activity	(baseline)
Discount	[0,3,6,9,12]	discount

References

Acquisti, A. (2004). Privacy in electronic commerce and the economics of immediate gratification. In *Proceedings of the 5th ACM conference on Electronic commerce*, pp. 21–29.

Acquisti, A., L. Brandimarte, and G. Loewenstein (2020). Secrets and likes: The drive for privacy and the difficulty of achieving it in the digital age. *Journal of Consumer Psychology*.

Auxier, B., L. Rainie, M. Anderson, A. Perrin, M. Kumar, and E. Turner (2020). Americans and privacy: Concerned, confused and feeling lack of control over their personal information. pew research centre (2019).

Beresford, A. R., D. Kübler, and S. Preibusch (2012). Unwillingness to pay for privacy: A field experiment. *Economics Letters 117*(1), 25 – 27.

Brandimarte, L., A. Acquisti, and G. Loewenstein (2013). Misplaced confidences: Privacy and the control paradox. *Social Psychological and Personality Science 4*(3), 340–347.

Braunstein, A., L. Granka, and J. Staddon (2011). Indirect content privacy surveys: mea-suring privacy without asking about it. In *Proceedings of the Seventh Symposium on Usable Privacy and Security*, pp. 1–14.

Carson, R. T. and M. Czajkowski (2019). A new baseline model for estimating willingness to pay from discrete choice models. *Journal of Environmental Economics and Management 95*, 57–61.

Colombo, S., W. Budzinski, M. Czajkowski, and K. Glenk (2020). Ex- ante and ex-post measures to mitigate hypothetical bias. are they alternative or complementary tools to increase the reliability and validity of dce estimates? Working Papers 2020-20, Faculty of Economic Sciences, University of Warsaw.

Daly, A., S. Hess, and G. de Jong (2012). Calculating errors for measures derived from choice modelling estimates. *Transportation Research Part B: Methodological 46*(2), 333 – 341. Emerging and Innovative Directions in Choice Modeling.

Daly, A., S. Hess, B. Patruni, D. Potoglou, and C. Rohr (2012). Using ordered attitudinal indicators in a latent variable choice model: a study of the impact of security on rail travel behaviour. *Transportation 39*(2), 267–297.

Danezis, G., S. Lewis, and R. J. Anderson (2005). How much is location privacy worth? In *WEIS*, Volume 5. Citeseer.

Draper, N. A. and J. Turow (2019). The corporate cultivation of digital resignation. *New media & society 21*(8), 1824–1839.

European Commission (2019). Special Eurobarometer 487a. The General Data Protection Regulation.

Glasgow, G., S. Butler, and S. Iyengar (2020). Survey response bias and the 'privacy para-dox': evidence from a discrete choice experiment. *Applied Economics Letters*, 1–5.

Goad, D., A. T. Collins, and U. Gal (2020). Privacy and the internet of thingsan experiment in discrete choice. *Information Management*, 103292.

Hess, S. and K. Train (2017). Correlation and scale in mixed logit models. *Journal of choice modelling 23*, 1–8.

Hole, A. R. and J. R. Kolstad (2012). Mixed logit estimation of willingness to pay distri-butions: a comparison of models in preference and wtp space using data from a health-related choice experiment. *Empirical Economics 42*(2), 445–469.

Kokolakis, S. (2017). Privacy attitudes and privacy behaviour: A review of current research on the privacy paradox phenomenon. *Computers Security 64*, 122 – 134.

Owen, A. B. (1995). Randomly permuted (t, m, s)-nets and (t, s)-sequences. In *Monte Carlo and quasi-Monte Carlo methods in scientific computing*, pp. 299–317. Springer.

Potoglou, D., J. F. Palacios, and C. Feijóo (2015). An integrated latent variable and choice model to explore the role of privacy concern on stated behavioural intentions in e-commerce. *Journal of Choice Modelling 17*, 10–27.

Sayaf, R. (2016). Contextual privacy: The interplay of sensitivity and context. *KU Leuven*.

Swait, J. and J. Louviere (1993). The role of the scale parameter in the estimation and comparison of multinomial logit models. *Journal of marketing research 30*(3), 305–314.

Tjiong, J., T. Dekker, S. Hess, M. Giergiczny, and M. Ojeda-Cabral (2019). Capturing zero-price effects in stated choice surveys and implications for willingness-to-pay computation. In *International Choice Modelling Conference 2019*.

Train, K. and M. Weeks (2005). Discrete choice models in preference space and willingness-to-pay space. In *Applications of simulation methods in environmental and resource economics*, pp. 1–16. Springer.

van Ooijen, I. and H. U. Vrabec (2019). Does the GDPR enhance consumers' control over personal data? An analysis from a behavioural perspective. *Journal of Consumer Policy 42*(1), 91-107.

University of Warsaw
Faculty of Economic Sciences
44/50 Długa St.
00-241 Warsaw
www.wne.uw.edu.pl