European Commission

JRC TECHNICAL REPORT

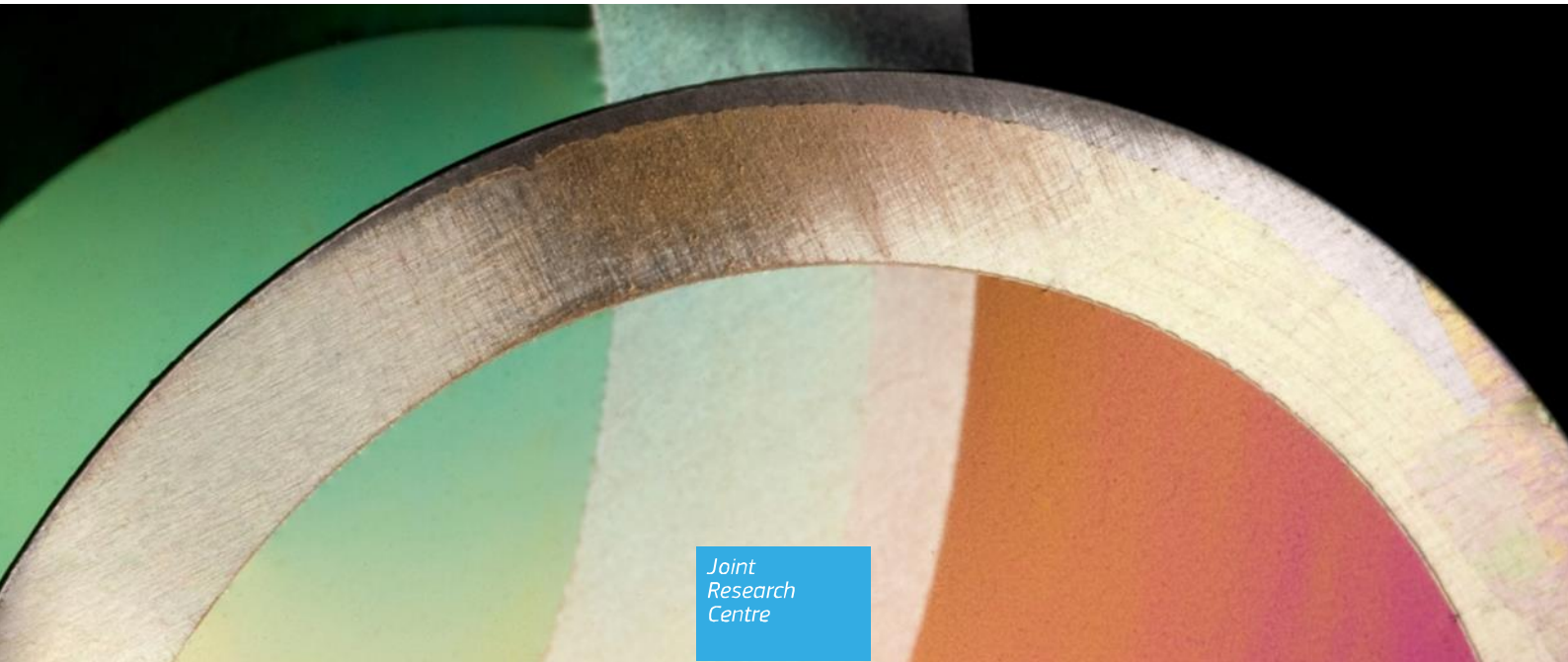
JRC Digital Economy Working Paper 2021-04

Measuring consumer well-being from using zero price digital services

The case of navigation apps and location-based services

Sobolewski, Maciej

2021



Joint Research Centre

Contact information

Name: European Commission, Joint Research Centre, Digital Economy Unit (JRC/B6)
Address: Edificio Expo, c/ Inca Garcilaso 3, 41092 Seville (Spain)
Email: maciej.sobolewski@ec.europa.eu

EU Science Hub
https://ec.europa.eu/jrc

JRC123535

Seville: European Commission, 2021

How to cite this report: Sobolewski, M., *Measuring consumer well-being from using zero price digital services. The case of navigation apps and location-based services,* European Commission, Seville, 2021, JRC123535.

Contents

Foreword

This study has been conducted within the work programme of Digital Economy Unit of the Joint Research Centre. The data used in this research project comes from a discrete choice experiment executed by the JRC and funded from "European Location Interoperability Solutions for e-Government" (ELISE) Action of the ISA^2 Programme.

For more information on the scope of ELISE visit:

https://ec.europa.eu/isa2/actions/elise_en

For other studies within the JRC Working Papers on Digital Economy visit:

https://ideas.repec.org/s/ipt/decwpa.html

Acknowledgements

The author would like to express his gratitude to Alessandro Annoni, Massimo Craglia, Nestor Duch-Brown, Bertin Martens, Daniel Nepelski and Francesco Pignatelli from the Joint Research Centre of the European Commission for providing valuable comments and support for this study on various stages.

The scientific output and views expressed in this study do not represent the official position of the European Commission, but are solely those of the author.

Author

Maciej Sobolewski (PhD) is an industrial economists working at the Joint Research Centre of the European Commission. At the JRC, his research focuses on the regulatory and competition aspects of online platforms and the economic impact of emerging technologies, such as blockchain and artificial intelligence.

Abstract

Digital maps and navigation applications are considered an essential tool by 70% of smartphone users. As these apps come predominantly free of charge, their contribution to consumer well-being cannot be captured by the common economic measures, like the GDP. This study demonstrates how the discrete choice experiment approach can be applied to measure, in an economically consistent way, consumer surplus from a navigation service. We elicit preferences for a satellite navigation with the two optional location-based functionalities: real-time traffic information and location-sensitive commercial information. In the experiment, the respondents are confronted with a range of location-sharing conditions set by a navigation provider. Finally, we estimate a demand model and derive welfare measures from the collected choices. Median consumer surplus from using basic satellite navigation without location-based functionalities is estimated at 8.06 EUR per month. Adding location-based services can increase this gain by 36% to 10.98 EUR, provided that users maintain control over location disclosure. Location-sharing terms set by a provider and privacy concerns of users both affect the size of the surplus from a navigation service.

1 Introduction

Billions of users worldwide use digital information services like a search, an email or navigation without any payment. It is perhaps the most noticeable and intriguing outcome of the digital revolution. Nevertheless, the fact that these services are offered free of charge on a permanent basis, without posing any threat to economic sustainability, has a convincing explanation. First, transformation to digital has reduced drastically marginal costs of information goods, making the cost of an extra subscription negligible (Varian et al., 2004). Second, with the emergence of multi-sided markets, a new way of recovering fixed development costs using the cross-side subsidization has become available (Tirole & Rochet, 2006). The so-called freemium business model is the extreme form of this general pricing rule applied to many multi-sided markets. The users who generate large positive cross-side externalities are not charged with money but instead give away their attention and data, like location or search history. Platforms use these insights to improve matching between the sides and monetize them through targeted advertising and in-app sales. Advertisers and app developers finally pay the fixed cost of information service.[1] On many platform markets, free-services serve essentially for attention and data extraction. However, the zero-pricing strategy is sustainable also in one-sided settings, like cloud storage (Dropbox) or text processing (Overleaf). On these markets, with strong direct network effects, the basic package is offered for free to expand the user base, with an intention to upgrade more intensive users to paid extended service later.

Zero price services have an unambiguously positive impact on consumer welfare, but in most cases affect traditional markets. This is clearly observed in communications and media industry, where text and voice communicators (Messenger, WhatsApp), social media (Facebook) or video and music streaming (YouTube) crowd out the traditional business of telecom operators, media publishers and music record labels. Business cannibalisation also affects proprietary software markets where open-source alternatives compete away proprietary solutions, for example in statistical computing (R and Python) and operating systems (Linux). Economic quantification of these adverse effects with output measures is not possible (Brynjolfsson et al., 2019). Standard measures, like GDP, capture the value of economic transactions at market prices. Hence, unpaid consumption is completely disregarded. To quantify well-being from zero-price goods one has to use welfare-based measures from the economic toolbox. Such measures are more demanding as they link directly to the reservation prices of consumers. Welfare-based methods require detailed micro level data, often proprietary, which restricts the use of this approach severely. Limited evidence from welfare-based measures indicates that consumers receive multibillion gains from access to specific zero price digital services.[2] Such studies are important for understanding the size of digital economy and for guiding investments in private and public infrastructures necessary for provision of these free services.

The present study focuses on the economic benefits from digital navigation service for individual use. Navigation technology has changed radically over time. Early explorers identified their position against destination based on observation of characteristic points on the terrain or in the sky. Terrestrial and astronavigation significantly improved in the 12th century with the invention of a compass and more precise mapping. Next major, technological changes came in the early 20th century with advancements in electromagnetism that led to the development of radio and radar navigation. The use of satellites marks another major breakthrough in the reliability, coverage and precision of navigation. It dates back to 1978 when the first satellite of GPS system was placed on Earth's orbit by the US. The GPS became operational in 1995 and was opened later for civil applications in industry, aviation, maritime transport, agriculture, construction and logistics. The segment of consumer portable navigation devices, which is a primary focus of this paper, is relatively young. Due to the miniaturisation of GPS receivers, first mobile handsets and devices dedicated for car navigation appeared in 2003. Since 2009, every new smartphone has had a free map and a navigation application preloaded with the operating system. The total number of GPS-enabled smartphones is estimated at 3.5 billion worldwide.[3] Out of that number, personal navigation assistant service is declared an essential tool by 70% of smartphone users. Despite potentially large gains for consumers, navigation service received little attention in economic research. The present paper aims to narrow this gap. Among prior studies

1 From the platform's perspective, a free service is an instrument to reveal and accumulate more data. A user is required to allow on his device tracking technologies, such as cookies. These technologies collect and pass back to the platform information on all clicks and actions made during an interaction with a service. Data from a single user have little value in isolation, yet if aggregated can be converted to valuable insights about the preferences and interests of users. In the freemium business model, personal information and attention of users of digital services constitute core intangible asset for digital platforms. Over 95% of multibillion revenues of Google or Facebook come from targeted advertising based on the intelligence from user data. These companies perform sophisticated profiling and classification of demand in order to display ads that maximize click-through rate.

2 For example, Brynjolfsson, Eggers, & Gannamaneni (2018) estimate the annual median willingness-to-accept of losing access to Wikipedia at 150$ per person, which amounts to $50 billion of consumer surplus created annually only for the US users.

3 The estimate taken from Statista (2020).

that estimate the value of free digital goods, only one concerns digital maps. Brynjolfsson, Collis, & Eggers (2019) estimate the compensating surplus for losing access to digital maps on the phone in the range of 59 EUR to 304 USD monthly. The current paper extends the analysis in several direction. It analyses how location-based services contribute to the value of navigation and what are the effects of user privacy concerns and location-sharing conditions on the valuation.

From a techno-economic perspective, personal navigation assistant is a system good composed of three complementary items: (i) a signal provided from a satellite navigation system (SNS), (ii) a receiver of that signal inside user's device and (iii) a digital map of the terrain with signal processing software. The software component calculates the position of the user's device on a map, based on signals received from at least three satellites. The hardware part may come in different forms, for example as an integrated car system, dedicated device or a smartphone. Unlike other digital services, satellite navigation relies on expensive public infrastructure, which has been developed, in the first place, based on security and military considerations. Major political powers have chosen to develop independent systems duplicating multibillion investments from public funds.[4] Notwithstanding geopolitical and industrial considerations, these public infrastructures increasingly contribute to the well-being of individual private users. The nature of this economic contribution is interesting from a public policy perspective but has not been a subject of a prior investigation.

Modern navigation apps benefit from rapid innovation in location-based services. Thanks to data connectivity, digital maps are enriched with additional layers of dynamic information, such as real-time traffic and transportation information (Nagaraj & Stern, 2020). Ride-sharing apps integrate digital maps with location tracking of car drivers paired with passengers. Online advertising and display of contextual commercial information are other arenas of location-driven progress. Big providers integrate platform model with their navigation apps, exposing users to location-sensitive search and display advertisements. Location is an essential ingredient for revenue generation from these in-app ads. Thus in modern navigation apps, location sharing is often enforced by design. The incentives to track users position every time and everywhere conflict however with the perception of location as sensitive personal data. Collection of user location by free navigation apps rises privacy concerns, especially that notice and consent can be purposely obscure and confusing about how often location data is collected from a device and if it is anonymised or not before sending.

The present study reports the results from an extended choice experiment, which elicited preferences for different variants of navigation apps, including those that display traffic information and location-sensitive commercial information. The stated preference survey was held in 2019 on a representative random sample of navigation users in Poland. The study makes two specific contributions. First, consumer surplus[5] from a range of navigation services is estimated. In particular, we show that a state-of-the-art navigation service with location-based traffic and commercial information generates nearly 40% higher gain than an 'old' satellite-only navigation. Second, we demonstrate how valuation of navigation with location-based services is affected by the terms for location sharing set by a navigation app provider. In the experiment the respondents were confronted with conditions of varying frequency and intrusiveness with respect to the mode of gathering location (permanent vs. optional) and a form of collected data (anonymised vs. personalised). On one extreme, a navigation provider can demand a permanent sharing of location with personal identifiers.[6] On the other extreme, sharing can be optional and location anonymised prior to disclosure. The preferences of users for the mode of sharing clearly differ and affect the valuation of the service. Finally, the study demonstrates an additional effect of privacy concerns held by the users on the size of consumer surplus from navigation services with location-based functionalities and different terms of location sharing.

The rest of the paper is organized in the following way. In section 2 we review relevant literature. Section 3 explains the experiment designed to elicit users' preferences for navigation and the econometric framework utilized for modelling preference data. In section 4 we provide the main estimates of consumer surplus from navigation and demonstrates the effects of sharing terms and privacy concerns. Section 5 concludes.

---

4 There are four satellite navigation systems in operation on Earth's orbit: GPS (developed by the US), GALILEO (developed by the EU), GLONASS (developed by Russia) and BAIDU (developed by China). Navigations systems are controlled by the governments and serve for civil, industrial and military purposes. GPS and GALILEO are global systems, while the others are regional. Except of GPS, remaining systems are incomplete and still need to shoot out a number of satellites to cover the entire globe. All the systems provide compatible signals, and user devices can switch from one SNS to another. Delays in deployment and rising costs of satellites triggered the discussion in Europe about the economic rationale for duplicating GPS system. However, for geopolitical reasons, the EU have chosen to deploy its own Earth observation system.

5 Formally, we measure compensating variation measured with willingness-to-pay.

6 Under the provisions of General Data Protection Regulation, a provider needs to obtain a prior consent from the user in this case. However, anonymised location can be collected and processed without a formal consent. Often users are unaware that their anonymised location is tracked permanently, even when they do not use navigation app.

2 Relevant literature

Prior studies attempting to estimate consumer surplus from using free digital goods are relatively scarce. The data requirements of direct estimation of welfare are quite demanding, which explains this scarcity. From the theoretical viewpoint, changes in consumer welfare can be approximated in monetary terms with Hicksian measures of compensating or equivalent income. In empirical research, these concepts are operationalized with the two corresponding measures: willingness-to-pay and willingness-to-accept. Both WTA and WTP assume a form of indirect utility and derive demand functions that can be estimated from revealed or stated preference data. The former source refers to a situation where real consumer choices on the market have been observed and collected. On the other hand, stated preference data are collected in an experiment, in which subjects are asked to declare their choices in hypothetical choice situations.

Early studies used revealed preference data to measure the value of the entire Internet and then tried to break it down to particular groups of services. The dominant paradigm in this line of research based on the value of time was developed by Goolsbee & Klenow (2006). They argue that using the Internet involves little monetary expenditure but is time-intensive. Hence, a proper measurement has to account for the opportunity cost of leisure time. Their estimation of a demand curve for the Internet consumption and corresponding consumer surplus is carried on household data containing information on time distribution among different activities. The estimated monthly per capita surplus from using the entire Internet is in a range of 208-316 USD for the US consumers. Pantea & Martens (2014) adopt the same approach to consumer micro data from five European countries. They obtain monthly estimates of consumer surplus from the use of the Internet in a similar range of 173 - 343 EUR. Brynjolfsson & Oh (2012) expand the value of time approach by introducing multiple 'leisure' products of different degree of substitutability. Between 2007 and 2011, the average value of consumer surplus from using the Internet was 324 USD per capita monthly. The surplus from the Internet increased over time due to quality improvements and more time spent online.

Among studies that use stated preference data, Bughin (2011) derives the value of unpaid web services from a conjoint survey with internet users in the US. He estimates the total value of free services at 24 USD per capita monthly in the US. The top four service categories include an e-mail (3 EUR), search engines (3.1 EUR) social networking (2.2 EUR) and instant messaging (2.1 EUR). Brynjolfsson et al. (2019) use a binary online choice experiment to estimate the value of Facebook and a range of categories of free services, including digital maps. Consequentiality of choices has a huge effect on valuations of Facebook obtained in their discrete choice experiment.[7] Monthly WTA for maps on phones is 304 USD in inconsequential treatment, while in a more controlled lab experiment conducted in Europe, the same authors obtained a much lower value of 59 EUR. In another study on the value of social media, Allcott et al. (2020) show that willingness-to-accept is affected by the length of deactivation period and credibility of enforcement. Evidence from stated preference studies for digital services is in line with the further observations on the direction of hypothetical bias in choice experiments: WTA (WTP) obtained from inconsequential choices is lower (higher) than for incentive compatible treatments. Because a full alignment of incentives is often impossible, weaker forms of enforcement and consequentiality checks adopted in choice experiments.

[7] The experiment on Facebook had two treatments to demonstrate the size of the hypothetical bias, which decreased WTA by the factor of 3.5.

3 Development of empirical study

3.1 Discrete choice experiment

In order to measure consumer surplus from using navigation services with optional location-based functionalities, a discrete choice experiment has been designed. Choice experiments are a prominent method for elicitation of stated preferences in applied economics. Typical experiment of this type has a form of a discrete-choice survey in which respondents are asked to make hypothetical choices between products or policy alternatives (Louviere et al., 2000). Choice experiments should not be confused with conjoint analysis, which is another survey-based technique of preference elicitation (Louviere et al., 2010). Formally, conjoint surveys are based on set orderings and as such do not comply with the economic theory of consumer choice.

Choice experiment method has a long self-standing history starting from Louviere & Woodworth (1983) building on Lancaster's (1966) theory of value and McFadden's (1973) random utility theory. Lancaster argued that goods are not direct objects of utility but rather the utility is derived from a vector of their properties or characteristics. McFadden formulated conditional logit model, which uses observed choices of individuals and compares utility levels associated with the choice alternatives in a way that is consistent with the classical utility maximization paradigm. He assumed that utility associated with any choice alternative is a sum of contributions that can be observed by a researcher and a component that cannot be observed, hence is treated as random. The role of the experiment is to introduce systematic variation into the characteristics of alternatives shown to respondents. Next, the effects of the attributes on the choices can be derived by linking in a functional form the changes in respondents' stated choices with variation in the choice situations. In the final step, after the estimation of parameters of the utility function, relevant welfare analyses can be carried out.

Choice experiments offer certain advantages for a researcher. First, such an experiment enables to elicit preferences for goods for which there is no revealed preference data available. The two notable examples are non-market goods and products with particular characteristics that are not yet offered on the market.[8] The experiment designed in the current study represents a similar case. We introduce two attributes, which cover a range of navigation variants, including older and state-of-the-art systems with hypothetical feature of an enhanced user control over sharing location data. Secondly, experiment introduces orthogonal differentiation in the choice attributes, which guaranties sufficient variability of stated preference data. By contrast, real market data are less variable and often suffer from collinearity problem, which may impair parameter identification. There are also disadvantages, however. The major criticism of choice experiments is their exposure to a hypothetical bias. It arises because declared choices are not budget-constrained and respondents do not bear any cost of stating non-optimal option. Depending on the topic of experiment, respondents may even have incentives to cheat in order to influence the results of the survey and avoid some unwanted policy options. Lack of incentive compatibility might result in biased estimates and consequently undermine validity of post-estimation outcomes. The long-lasting research on the hypothetical bias has led to a set of postulates and conditions for ensuring incentive compatibility (Carson & Groves, 2007; Vossler et al., 2012). As long as experimental protocols satisfy these conditions, parameter estimates from stated and revealed preference data are found to be consistent (Carlsson & Martinsson, 2001; Carson et al., 1996). In the current experiment, we could apply only light measures supporting consequentiality of choices, without being able to request respondents to pay declared amounts. To increase integrity of the final data we have deleted from the sample respondents who had strange response patterns or showed disregard for the topic of the survey. One particular type of anomaly that is often encountered in a discrete choice experiment are the so-called protest respondents (Meyerhoff & Liebe, 2008). Protest respondents are those who always chose status quo alternative because they do not agree on the rules defined in the experiment.

The data, analysed in the present study, were collected in March and April 2019 from an online survey conducted on a random sample of N=762 users of portable or in-car navigation systems in Poland. The sample was drawn from a representative panel of the internet users. The experiment introduced a hypothetical 'blackout' situation, in which satellite signals from the GPS system will cease to be available in Europe in six months. The potential reasons for the blackout given to the respondents was a trade war or geopolitical conflict. The respondents were informed that in such a case, they would be able to use the European Galileo satellite system and that, thanks to compatibility with the GPS, no hardware or software upgrades would be needed. In order to introduce the price attribute to choice alternatives, respondents were instructed that Galileo is under construction and could be completed in a shorter period only at an additional,

8 A nice illustration of such application is the experiment on electric cars in Daziano, Sarrias, & Leard (2017).

huge cost. This argument served as a motivation for introducing a subscription fee for the use of navigation service. In the next step, attributes of choice alternatives were introduced and explained and the section with choices followed. The final part of the questionnaire contained socio-demographic and attitudinal questions. The latter served to measure privacy concerns and filter out from the sample protest respondent. In order to elicit preferences for navigation with optional location-based functionalities, two non-price attributes were defined, as shown in Table 1. They describe conditions on to which respondents had to agree in order to get access to traffic information and commercial information.[9] We assumed that in exchange for real-time traffic information (on traffic jams, travel time, road accidents) a provider of navigation requests momentary and anonymised location from a user. In case of commercial information (local search and display ads, sponsored location pins on the map) a provider collects personalised location from a user and may store it.

For each of the two location-based functionalities, an attribute determines if this functionality is offered in the app. If it is available, the attribute indicates whether location sharing is mandatory or optional. Mandatory sharing entails that a user position is collected constantly whenever the app is open, even if he does not use location-based functionality in a given moment. With optional sharing, a user can deactivate location sharing and activate it again at any time. During deactivation period, related location-based functionality cannot be used. However, the user keeps access to the main satellite-based service inside the app. In essence, optional sharing leaves more control over a location to the user, while mandatory sharing grants it to the app provider.[10] The two attributes were identified during qualitative interviews with navigation users carried out in the preparatory phase of the study. Participants attached great importance to the way and the purpose for which their location data is collected. Interviews showed that users perceive location as sensitive data and prefer more controlled, optional sharing. There is also common awareness that providers are interested in gaining permanent access to location data because it has a commercial value. The levels of both choice attributes were set according to this perception of misaligned interests in order obtain choice alternatives with clear discriminatory power.

Table 1. Attributes of navigation service and their levels used in the choice experiment.

Choice attribute	Attribute levels
Access to location-based commercial information[(a)] in exchange for personalised location data (a) local search and display ads, sponsored location pins on the map	☐ Functionality not available ☐ Optional location sharing ☐ Mandatory location sharing
Access to location-based traffic information[(b)] in exchange for anonymised location data (b) real time information on road accidents, traffic jams and travel time	☐ Functionality not available ☐ Optional location sharing ☐ Mandatory location sharing
Monthly price of subscription to navigation service in PLN[11]	Subscription fee assigned from the following values: 5, 15, 20, 25, 35, 45, 55.

9 In spirit of data protection regulation (GDPR) we have adopted the minimal design principle, which states that only data necessary for the operation of the service should be collected from a user.

10 In the experiment, it was assumed that by closing an app a user stops sharing his data. In practice, this is not necessarily true as various applications keep collecting and sending out location data in the background.

11 1 PLN ≈ 0.25 EUR.

Based on the levels of both choice attributes we differentiate four main configurations of navigation service for welfare analysis:

- **Satellite-only 'old' navigation** utilizes just a satellite signal and the pre-installed digital map. Location-based functionalities are not available and there is no location sharing. Both choice attributes have values set to 'no sharing'.
- **User-centric 'state-of-the-art' navigation** offers location-based functionalities. A user preserves control over location sharing in both modes (personalised and anonymised). He can freely activate and deactivate sharing. If a user disables location sharing in one or both modes, he loses access to the corresponding functionality, but may still use satellite-based navigation. Both choice attributes are set to 'optional location sharing'.
- **Provider-centric 'state-of-the-art' navigation** offers location-based functionalities. The navigation provider can collect location data constantly. If a user disables location sharing, he cannot access the satellite-based navigation on his device (application). Both choice attributes have value 'mandatory location sharing'.
- **Balanced 'state-of-the-art' navigation** offers location-based functionalities. The control over location sharing is divided between the navigation provider and a user. Anonymous location is collected constantly. If a user disables anonymous location sharing, he cannot access the satellite-based navigation on his device (application). In this variant of a service, anonymous location attribute is set to 'mandatory sharing' while personal location attribute is set to 'optional sharing'.

The first three listed navigation variants are self-explained. The last variant assumes a permanent collection of anonymised location data for wider use. Public authorities could, for example use it to facilitate traffic management or in an emergency. We want to check if such a service variant, which potentially generates positive externalities without posing a threat to sensitive user data is acknowledged and how it compares to the user-centric navigation in terms of economic surplus.

3.1.1 Experimental design

The design of choice alternatives, i.e., setting the combinations of attribute levels presented to a respondent in a choice situation, determines the amount of information that may be extracted from respondents' choices. In the so-called efficient designs, the choice situations are generated in a way that maximizes Fisher information, given prior expectations on what the parameters of a representative respondent's utility function are (Scarpa & Rose, 2008).[12] Effective designs reveal more information, compared to standard orthogonal plans, thus allowing for smaller samples. Finally, the state-of-the-art choice experiments utilise Bayesian efficient designs (Bliemer & Rose, 2010). The benefit of this approach is that it accounts for uncertainty with respect to parameters' priors, by allowing these priors to be random variables following a probability distribution over a range of plausible values (Sandor & Wedel, 2001). In the reported experiment, two experimental designs have been developed, each with 18 choice situations. The first design assumed a binary choice between one generic variant of paid navigation: 'GN1' and the 'opt out' option (no access to navigation service). In the second design, choice situations contained two generic alternatives: 'GN1', 'GN2' and the 'opt-out' option. Both designs were optimized with respect to the D-error, using Bayesian priors taken from the pilot survey on N=200 respondents. Participants were assigned to the designs randomly.[13] Each respondent had individualised choice sets of 12 choice situations drawn randomly from the pool of 18. The example of a choice card with three alternatives is shown in Table 2.

[12] In technical terms, this is equivalent to the minimisation of the determinant of the asymptotic variance-covariance matrix of the model parameters (D-error).

[13] In this paper, observations from the two designs are pooled and analysed jointly.

Table 2. Example of a choice card.

Which of the following solutions would you consider the best for yourself?			
	GN1: generic navigation option 1	GN2: generic navigation option 2	Opt-out: no access to navigation
Commercial information functionality in exchange for your personalised location data with:	Optional location sharing	Functionality not available	
Traffic information functionality in exchange for your anonymised location data with:	Mandatory location sharing	Mandatory location sharing	
Monthly fee	25	10	0
Your choice	☐	☐	☐

3.2 Characteristics of the sample

Official census statistics do not provide data on the use of digital applications. Hence, no suitable information can be found to set out representativeness criteria for sampling navigation users from general population. Therefore, the survey was executed on the internet panel maintained by the external provider. The panel is representative for the adult (18+) internet users population in Poland across age, size of the place of residence and country province. The qualification to the survey was based on the three filtering questions, which ensured jointly that each respondent owns at least one GPS-enabled device and they had used navigation service on their device during a trip, a journey, a walk or any similar situation at least once, over the last 12 months. Based on the above condition around 33% of the subjects who initially signed up for the survey, were disqualified. The final sample consisted of 762 navigation users and the average duration of an interview was 20 minutes. The questionnaire was accessible from desktops and portable devices.

Table 3. Descriptive statistics of the sample.

Demographics and user profile				
age [years]	15-29	30-44	45-59	60+
N=762	27%	23%	35%	15%
gender	women	men		
N=762	50%	50%		
education	primary/voc.	secondary	higher	
N=762	35%	35%	30%	
Frequency of use	at most once per month	once per week	few times per week	every day
N=762	50%	21%	19%	10%

Main device used	smartphone	in-car navigation	portable car device	portable touristic
N=762	80%	5%	13%	2%
Gave explicit consent for location sharing?	yes	no, didn't want to share	not appl./old device	do not know/ note sure
N=762	50%	23%	7%	20%
Which smartphone app do you use?	Automapa	Google Maps	Yanosik	Apple Maps
N=694	18%	91%	27%	9%
Portable device used	Garmin	Goclever	Manta	TomTom
N=281	19%	13%	22%	26%

Descriptive statistics of the final sample are presented in Table 3. Navigation users represent all age and income groups. The sample is balanced concerning gender and education. Looking at the intensity of use, half of the respondents use navigation infrequently, at most once per month. Nearly 20% declare to use navigation few times per week, while 10% are everyday users. As expected, navigation service is mostly used on a smartphone (80%), followed by dedicated portable device (13%) from specialised suppliers like TomTom or Garmin. In-car integrated navigation was indicated by 5% of respondents, as their principal form of the service. In the smartphone segment, there is an absolute dominance of Google Maps, followed by the two automotive-specific apps: Automapa and Yanosik. These two apps are popular among drivers for their specific functionalities like radar information. Descriptive statistics bring essentially three insights. First, the market standard of navigation system consists of a smartphone with free-of-charge application from Google. Second, among navigation apps users, nearly half (45%) engages in multihoming. In particular, general-purpose apps such as Google Maps or Apple Maps are complemented with apps dedicated to car driving. Lastly, location sharing is indeed perceived as sensitive data. As much as 50% of respondents remember giving their explicit consent to the navigation provider for collecting location, followed by 23% who explicitly disallowed it. For 20% of navigation users, location sharing is not a big deal. They admitted to being unaware about how this setting is configured in their navigation service.

Sharing of location data rises some privacy concerns among users. To capture this aspect, we have included a block of questions measuring attitudes towards privacy and actual protection measures undertaken in various occasions. Privacy concerns are measured in various ways in the literature. In this study we have adopted psychometric scales developed by Buchanan, Paine, Joinson, & Reips (2007). As shown in Table 4, respondents have quite strong subjective concerns about the use of their personal data in the online context. After applying exploratory principal component analysis to these scales, we have isolated two orthogonal indices of privacy concerns. Similar to other studies (Potoglou et al., 2015) these variables entered as covariates into the choice model in Section 4.1.

Table 4. General privacy concerns related to the online activity of users.

Questions on a Likert scale	**weak**	**moderate**	**strong**
Your personal data is shared with third party providers	20%	17%	62%
Your online behaviour and history are collected and shared with third party providers	23%	20%	56%
Your personal data is kept by the provider after you from the service	20%	18%	62%

Your private messages are monitored by algorithms for the purpose of personalised ads	21%	19%	60%
Your personal data is sold after anonymization to the third party providers	19%	21%	60%

Notes: Respondents were asked the following question: To what extent would you be concerned if the following situation happened? The answers were given on the 5-point Likert scale, starting from 'not at all concerned' and ending on 'very much concerned'. Results presented in Table 4 have been recoded for clarity to three levels.

3.3 Econometric framework

Formally, the data from discrete choice experiment is modelled with the random utility framework (McFadden, 1973). The utility function of consumer i from alternative $j \in J$ in choice situation $t \in T$ can be expressed as:

$$U_{ijt} = \boldsymbol{\beta}'\mathbf{x}_{ijt} + \varepsilon_{ijt} \qquad (1)$$

In equation (1), $\boldsymbol{\beta}$ is the vector of utility parameters, $\mathbf{x}$ is the vector of observed attributes specific to the consumer, alternative j and choice situation t. A random component ε represents the joint impact of all unobserved factors that influence decision-making. By assuming that the random component is identically and independently Gumbel distributed, the multinomial logit (MNL) model is obtained. The MNL model is convenient, for its closed-form logit expression for the choice probabilities of each alternative, but is overly simplistic. Most importantly, it implies independence from irrelevant alternatives and does not account for unobserved heterogeneity and correlation (Train, 2009). In this study, we apply a mixed logit (MXL) to take the respondents' preference heterogeneity into account (Hensher & Greene, 2003). MXL model assumes that consumer i has specified, albeit non-observable, parameters of the utility function. Individual parameters follow probability distributions specified by the researcher: $\boldsymbol{\beta}_i \sim f\ (\mathbf{b} + \boldsymbol{\Pi}\boldsymbol{z}_i, \boldsymbol{\Sigma})$, hence are treated as random variables. Means of these distributions, $\mathbf{b}$ can be affected by the observed characteristics of respondents given in a vector $\boldsymbol{z}_i$. Variance-covariance matrix of parameters: $\boldsymbol{\Sigma}$ can be non-diagonal to account for correlations across alternatives or choice situations. In this case, it is useful to express individual parameters of the utility in the following way:

$$\boldsymbol{\beta}_i = \mathbf{b} + \boldsymbol{\Pi}\boldsymbol{z}_i + \boldsymbol{\Gamma}\boldsymbol{\eta}_i \qquad (2)$$

Where: $\boldsymbol{\Gamma}$ is a lower-triangular Cholesky matrix, resulting from factorization of covariance matrix of random parameters ($\boldsymbol{\Gamma}\boldsymbol{\Gamma}' = \boldsymbol{\Sigma}$); $\boldsymbol{z}_i$ is a vector of characteristics of an individual that influence means of selected preference parameters; $\boldsymbol{\Pi}$ is a vector of additional shifting parameters to be estimated and $\boldsymbol{\eta}_i$ is a random disturbance with mean 0 and standard deviation 1. If $\boldsymbol{\Gamma}$ is diagonal, than random parameters are independently distributed, otherwise it introduces correlation among them. By assuming a structured variation of individual tastes in the sample, in the form of random parameters, the MXL model is more realistic and typically yields a much better fit to the data. This benefit comes at the cost of a more complicated estimation procedure because mixed logit probabilities involve integrals, which do not have a closed form. Unconditional probabilities must be simulated by taking multiple random draws from the respective joint distribution and averaging (Train, 2009). In the final step, the sequence of T choices made by each person during the experiment is represented in the likelihood form. Estimators of random parameters $\mathbf{b}, \boldsymbol{\Sigma}$ can be obtained numerically from maximization of the log-likelihood function:

$$LL = \sum_{i=1}^{I} \log \frac{1}{D} \sum_{d=1}^{D} \prod_{t=1}^{T} \sum_{j=1}^{J} y_{ijt} \frac{\exp(\mathbf{x}_{ijt}\boldsymbol{\beta}_i)}{\sum_{j=1}^{J} \exp(\mathbf{x}_{ijt}\boldsymbol{\beta}_i)} \qquad (3)$$

In the above formula, y_{ijt} is a dummy variable equal to 1 if respondent i selected alternative j in choice situation t and 0 otherwise and D represents the number of draws taken from a joint normal distribution. In the empirical analysis, we first assume homogenous means and variances of all K utility parameters across individuals. Later, we introduce the effects observed respondent characteristics $\mathbf{z}_i$ (privacy attitudes) to means of the distributions and allow random parameters to be correlated.

3.4 Willingness-to-pay and consumer surplus measures

In the post estimation stage, we simulate willingness-to-pay (WTP) for particular levels of attributes and consumer surplus (CS) of particular alternatives, given by the combinations of choice attributes (Hensher & Greene, 2003). For a linear utility function, consumer's willingness-to-pay for a change in an attribute is defined as the ratio between the utility coefficient of interest and a negative price parameter. This ratio is equivalent to a marginal rate substitution between attribute k and money. In MNL model, both coefficients are fixed, but uncertain due to sampling variance. Since these estimates are derived from maximum likelihood, they have asymptotically normal distributions. Hence, WTP given in Eq. (4) is a random variable with undefined moments. Often this variable is assumed asymptotically normal itself and confidence intervals for WTP can be obtained from the Delta method. In mixed logit, individual utility parameters are random variables following specific distributions assumed by the modeler. In such a case alternative, non-parametric methods, such as Krinsky and Robb simulation are often adopted to estimate WTP and calculate corresponding confidence intervals (Bliemer & Rose, 2013). We follow this approach in the current study. First multiple vectors of the model parameters are generated. Then for each vector, multiple samples of consumers with individual level utility parameters are drawn to obtain welfare measures. Finally, percentile intervals around selected measures are obtained. More specifically, let $\mathbf{Y}$ be the vector of estimated MXL parameters and let $\mathbf{\Omega}$ be the full covariance matrix of (all) model parameters. In the first step we draw a sample of $\mathbf{Y}$ according to the process given by:

$$\widetilde{\mathbf{Y}} = \mathbf{Y} + \mathbf{c}'\mathbf{u} \qquad (4)$$

In equation (4), ***c'*** is a lower triangular *Cholesky* matrix of the full variance-covariance matrix and **u** is the vector of standard normal variates. We account for the full $\mathbf{\Omega}$, including off-diagonal elements in order to use information from the entire multivariate distribution. For a single vector $\widetilde{\mathbf{Y}}$, we calculate the implied covariance matrix $\mathbf{\Sigma}$ of random parameters from the elements of the Cholesky matrix $\mathbf{\Gamma}$ included in $\widetilde{\mathbf{Y}}$. In the mixed logit with correlated random parameters $\mathbf{\Gamma}$ is non-diagonal, implying non-zero covariance in $\mathbf{\Sigma}$. Let '*k*' and '*price*' be the two attributes of interest for calculating WTP with random parameters distributed according to the choice of a modeler. Let $(\tilde{b}_k, \tilde{b}_{price}, \tilde{\sigma}_k, \tilde{\sigma}_{price})$ be the means and standard deviations of these distributions given by $\widetilde{\mathbf{Y}}$. In the second step of the procedure, we take many draws of individual parameters from these respective distributions and calculate individual WTP values:

$$\widetilde{WTP}_{i,k} = \frac{\tilde{b}_k + \tilde{\varrho}_k \cdot \xi_i}{|\tilde{b}_{price} + \tilde{\sigma}_{price} \cdot \nu_i|} \qquad (5)$$

Where ν_i, ξ_i are draws from standard normal distribution. For each sample, we take median as a central tendency of WTP distribution to avoid exploding means when a price draw approaches zero.

Next, for each sample we compute changes in individual consumer surpluses from shifting between any two choice alternatives, given by the combinations of attributes. For example, in the case of a binary choice between alternative '*j*' and '*opt-out*', the expected change of consumer surplus from shifting to *j* is given as (Small & Rosen, 1981):

$$\Delta CS = E(CS_j - CS_{opt-out}) = \int \left(\frac{1}{|\beta_{price}|}\left(V_\beta^j - V_\beta^{opt-out}\right) f(\beta|\boldsymbol{b}, \mathbf{\Sigma}) d\beta \qquad (6)$$

where β_{cost} is the parameter measuring marginal utility of income and V_β^k is the observed part of the utility function from alternative *k*. This integral can be approximated by summing individual net utilities and averaging.

4 Main results

For mixed logit estimation, we assume 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):

$$U_{ij} = \beta_{NA_i} NA_{ij} + \beta_{PL1_i} PL1_{ij} + \beta_{PL2_i} PL2_{ij} + \beta_{AL1_i} AL1_{ij} + \beta_{AL2} AL2_{ij} + \beta_{PRICE} PRICE_{ij} + \epsilon_{ij} \quad (7)$$

In the above specification $\boldsymbol{\beta}$ is the vector of parameters associated with respective choice attributes and ε_{ij} is a random component of utility associated with alternative j. The interpretations of the coefficients are given in the first column of Table 5. Except of the price, all variables entering utility model are dummies. They denote characteristics of a particular navigation option. *NA* is alternative-specific constant for the opt-out. The baseline alternative for which utility is normalized to zero is the satellite-only 'old' navigation.

In Table 5 we present results of three different mixed logit models. The estimated utility coefficients are not meaningful in absolute terms. They should be interpreted as changes relative to the baseline. Model 1 is a preliminary model without correlation, estimated on the entire sample of respondents (N=762). Model 2 uses the restricted sample (N=575), after filtering out protest respondents but leaving the so-called 'genuine zeros' (Hanley et al., 2007).[14] Model 3 is estimated on the restricted sample and assumes correlation between random parameters. All three models display expected signs of the means of utility parameters, except of the variable PL2, which has a positive sign in models without correlation. Price coefficient is always negative as expected. We note that lower coefficient for the price in model 1 arises due to inclusion of protest respondents in the sample. These subjects chose the 'opt out' alternative across all choice situations, to manifest their disagreement with the idea of introducing payment instrument. Therefore, the valuation of navigation service based on model 1 would have been significantly lower. Because of the above considerations, model 3 is our reference specification.

Looking at the estimation results from model 3, we note that a lack of access to navigation is associated with a large disutility, compared to the baseline option: a service that uses only satellite signal. Turning to the utility coefficients for location-based functionalities, respondents always prefer traffic information functionality than commercial information. Commercial information generates positive utility only if location sharing is optional, otherwise a contribution of this functionality to utility is negative. Consequently, any navigation variant with enabled commercial information functionality, which permanently collects personalised location, will be considered inferior to an otherwise identical service without such functionality. Lastly, optional location sharing is always preferred to mandatory sharing, regardless of the functionality. This is not surprising, as users perceive the former terms as less privacy intrusive and preserving control over location in their hands. Looking at the coefficients for standard deviations of random parameters, we note that they have large relatively to the means and are highly statistically significant. This observation points to the large unobserved individual preference heterogeneity concerning particular choice attributes. The results of the model indicate that users have varying sensitivity to the price instrument. Hence, introducing subscription fee for navigation would restrict demand for this service on the part of the most price-elastic users. Utility coefficients obtained from model 3, allow for computing several post-estimation economic indicators of great interest. Apart from welfare measures, which are the focus of this study, it is possible to compute implicit demand elasticity with respect to price and other choice attributes. Such elasticities can be particularly useful for conducting market delineation and running hypothetical monopolist tests.[15]

In essence, our results provide a clear evidence that users derive higher utility from traffic information functionality than commercial information in either mode of sharing their location. Moreover, optional sharing is preferred to mandatory sharing regardless of the provided functionality. Lastly, the users enjoy using commercial information if it can be deactivated when not needed, without losing access to the main satellite-only service. On the other hand, if location-sharing requirements change to permanency, ads become undesired.

[14] 'Genuine zeros' are the respondents who always choose 'opt out' alternative due to a very low valuation of navigation service. The need to be distinguished from protest respondents who select 'opt out' because they disagree for example with the idea of paid navigation service. In the larger sample we have classified only 20 respondents as genuine zeros and 187 subjects as protest respondents.

[15] Additional to classical test for small, significant, non-transitory increase in price (SSNIP) one can run quality version of this test (SSNIQ) by changing for example the terms of location-sharing from mandatory to optional.

Table 5. Mixed logit model of preferences for navigation and location data attributes.

Utility function parameters	distribution	model (1)	model (2)	model (3)
		means (s.e.)		
NA: no access to navigation (alternative specific constant)	normal	-4.656***	-4.933***	-6.129***
		(0.253)	(0.282)	(0.338)
PL1: comm. info & pers. location: optional sharing vs. not available	normal	0.404***	0.457***	0.451***
		(0.069)	(0.069)	(0.075)
PL2: comm. info & pers. location: mandatory sharing vs. not available	normal	0.686***	0.600***	-1.012***
		(0.126)	(0.126)	(0.147)
AL1: traffic info & anon. location: optional sharing vs. not available	normal	1.207***	1.207***	1.420***
		(0.101)	(0.101)	(0.133)
AL2: traffic info & anon. location: mandatory sharing vs. not available	normal	0.747***	0.715***	1.024***
		(0.109)	(0.108)	(0.137)
PRICE: price per month (PLN)	- lognorm.	-1.584***	-1.980***	-1.909***
		(0.043)	(0.047)	(0.054)
		standard deviations (s.e.)		
NA: no access to navigation (alternative specific constant)		3.976***	3.515***	5.25***
		(0.206)	(0.252)	(0.293)
PL1: comm. info & pers. location: optional sharing vs. not available		0.589***	0.590***	0.533***
		(0.130)	(0.100)	(0.099)
PL2: comm. info & pers. location: mandatory sharing vs. not available		1.629***	1.592***	1.875***
		(0.160)	(0.169)	(0.190)
AL1: traffic info & anon. location: optional sharing vs. not available		0.144	0.271	1.438***
		(0.186)	(0.176)	(0.143)
AL2: traffic info & anon. location: mandatory sharing vs. not available		0.898***	0.851***	1.754***
		(0.093)	(0.104)	(0.154)
PRICE: price per month (PLN)		1.109***	0.849***	1.189***
		(0.036)	(0.038)	(0.030)
model diagnostics				
Log-likelihood		-4278.859	-4046.826	-3918.11
BIC		8667.169	8199.722	8074.683
AIC		8581.719	8117.651	7890.22
n (no. of observations)		9144	6900	6900
k (no. of parameters)		12	12	27

Notes: ***, ** and * indicate 0.1%, 1% and 5% significance levels, respectively. All models are mixed logits estimated in R using gmnl package with 100 Halton draws. Parameter estimates represent moments (mean, standard deviation) of the distributions of utility coefficients. Standard errors are provided in parentheses. All parameters were assumed to be normally distributed except of price, which is distributed log-normally. For log-normal distribution the estimated coefficients of the underlying normal distribution are provided.

4.1 Introducing covariates to explain heterogeneity

A strong preference for optional sharing, following from the estimation results, suggests that privacy considerations might play an important role in the valuation of the navigation service. To test this hypothesis we introduce covariates to the mixed logit model to explain respondents' preference heterogeneity. The covariates are constructed from the five psychometric scales presented in Table 4, using factor analysis. In total, we have obtained two factors that correspond to different dimensions of privacy concerns. Their loadings are given in Table 6. The first factor (PC1) is related to the disclosure of personal data to online providers and potential uncontrolled use of these data by the third parties. The fear of unauthorised used is quite pervasive among the users. This is because once the data is shared on the Internet, it stays there

forever. The second component is related to monitoring of private messages by algorithms for personalization of ads (PC2). People widely perceive such a practice as braking the secrecy of correspondence, even when only machines read it. The in-sample values of covariates are computed as predicted values from Bartlett regression. Descriptive statistic of these predicted scores are given in the lower section of Table 6. Greater values of covariates correspond to the larger privacy concerns.

Table 6. Principal component analysis of privacy concerns.

A. Loadings from questions	**PC1: data disclosure**	**PC2: algorithmic processing**
Personal data shared with third party providers	0.73	
Online behaviour and history collected and shared with third party providers	0.76	
Personal data kept by the provider after user quits from the service	0.49	0.51
Private messages monitored by algorithms for personalised ads		0.83
Personal data sold after anonymization to third party providers	0.51	
B. Descriptive statistics	**PC1**	**PC2**
Mean	0	0
St. dev	1.22	1.22
Min	-4.22	-3.96
Max	2.92	2.85
q.10	-1.80	-1.74
Median (q.50)	0.25	0.09
q.90	1.24	1.42

Notes: Cut-off level = 0.4, varimax rotation used.

Variables PC1 and PC2 are explanatory variables for the means of the four random parameters related to location-based functionalities and location sharing: PL1, PL2, AL1, AL2. For example, parameter β_{PL1_i} has now the following distribution:

$$\beta_{PL1} \sim N\left(b_{PL} + \Pi_{PC1,PL1}\, PC_1 + \Pi_{PC2,PL1} PC_2,\ \sigma^2_{PL1}\right) \qquad (8)$$

The estimated mixed logit model (model 4) contains additional vector $\mathbf{\Pi}$ with eight coefficients that shift means of selected random parameters, as can be seen in Table 7. Because covariates are standardised, the estimated means $\mathbf{b}$ of random parameter characterise consumers with average levels concerns in the sample: $\boldsymbol{z_i} = \mathbf{0}$. Means of random parameters for users with different levels of privacy concerns take the form of sums: $\mathbf{b} + \mathbf{\Pi}\boldsymbol{z_i}$. All covariate coefficients $\mathbf{\Pi}$ have expected negative signs.

Table 7. Mixed logit model with correlated random parameters and privacy concerns as covariates.

Utility function parameters	distribution	model (4)	
		means (s.e.)	standard deviations (s.e.)
NA: no access to navigation (alternative specific constant)	normal	-5.292***	4.917***
		(0.279)	(0.307)
PL1: comm. info & pers. location: optional sharing vs. not available	normal	0.448***	0.540***
		(0.068)	(0.097)
PL2: comm. info & pers. location: mandatory sharing vs. not available	normal	-0.904***	1.642***
		(0.137)	(0.193)
AL1: traffic info & anon. location: optional sharing vs. not available	normal	1.432***	1.372***
		(0.133)	(0.184)
AL2: traffic info & anon. location: mandatory sharing vs. not available	normal	1.006***	1.447***
		(0.134)	(0.187)
PRICE: price per month (PLN)	- lognorm.	-1.950***	0.963***
		(0.048)	(0.025)
		covariates for means (s.e.)	
disclosure * comm. info & pers. location: optional		-0.092	(0.057)
algorithmic * comm. info & pers. location: optional		-0.043	(0.056)
disclosure * comm. info & pers. location: mandatory		-0.283**	(0.108)
algorithmic * comm. info & pers. location: mandatory		-0.105	(0.112)
disclosure * traffic info & anon. location: optional		-0.059	(0.095)
algorithmic * traffic info & anon. location: optional		-0.017	(0.094)
disclosure * traffic info & anon. location: mandatory		-0.129	(0.099)
algorithmic * traffic info & anon. location: mandatory		-0.162	(0.093)
model diagnostics			
Log-likelihood		-3910.979	
BIC		8131.332	
AIC		7891.957	
n (no. of observations)		6900	
k (no. of parameters)		35	

Notes: ***, ** and * indicate 0.1%, 1% and 5% significance levels, respectively. Model 4 is estimated in R using gmnl package with 100 Halton draws. Parameter estimates represent moments (mean, standard deviation) of the distributions of utility coefficients and covariates for means. Standard errors are provided in parentheses. All random parameters were assumed to be normally distributed except of price, which is distributed log-normally. For lognormal distribution the estimated coefficients of the underlying normal distribution are provided.

This implies that individuals with higher than average levels of privacy concerns ($\boldsymbol{z_i} > \mathbf{0}$) receive lower utility from location-based services. We note also that the estimated coefficients $\boldsymbol{\Pi}$ for mandatory sharing have much larger absolute values than the remaining parameters for optional sharing. Apparently, mandatory sharing of location is indeed perceived as problematic on privacy grounds. Several covariate effects lack statistical significance and hence should be treated with caution. Disclosure concern related to mandatory sharing of personalised location has the largest and statistically significant negative effect on the utility from commercial information functionality. The above results provide clear evidence for the role of privacy concerns on utility and valuations of navigation services with location-based functionalities that require location sharing. In particular, the largest effect can be observed for the case of permanent sharing of personalised location data and is related with unauthorized use of these data by third parties.

4.2 WTP and consumer surplus from navigation

We now turn to the quantification of the overall well-being from using navigation. In what follows, we present the results of welfare estimates, using the results from model 4 with all elements of vector $\boldsymbol{\Pi}$. We begin with

willingness-to-pay measures for selected location-based options set in the choice experiment. In Below we present WTP values for free levels of privacy concerns: sample average, 90th and 10th quantile.[16] See horizontal sections (A-C) in Table 8.

Table 8. Willingness-to-pay estimates.

	WTP for opt-out (losing access to navigation)	WTP for commercial information with optional location sharing	WTP for commercial information with mandatory location sharing	WTP for traffic information with optional location sharing	WTP for traffic information with mandatory location sharing
	[A] sample average (covariates at sample means: PC1=0; PC2=0)				
50%	-8.06	0.67	-1.33	2.21	1.69
95% c.i.	(-8.75; -7.3)	(0.46; 0.87)	(-1.76; -0.92)	(1.8; 2.69)	(1.2; 2.17)
(se)	(0.35)	(0.1)	(0.22)	(0.23)	(0.25)
	[B] high concerns about data disclosure and algorithmic processing (covariates values at q.90)				
50%	-8.06	0.41	-2.13	2.05	0.98
95% c.i.	(-8.75; -7.3)	(0.01; 0.81)	(-3.09; -1.25)	(1.32; 2.91)	(0.24; 1.81)
(se)	(0.35)	(0.21)	(0.47)	(0.4)	(0.41)
	[C] low concerns about data disclosure and algorithmic processing (covariates values at q.10)				
50%	-8.06	1.03	-0.26	2.45	2.59
95% c.i.	(-8.75; -7.3)	(0.52; 1.56)	(-1.27; 0.7)	(1.46; 3.5)	(1.49; 3.68)
(se)	(0.35)	(0.26)	(0.50)	(0.52)	(0.56)

Notes: Standard errors (se) and 95% confidence intervals of are obtained with Krinsky and Robb simulations using 10^4 draws of consumers and 10^4 draws of random parameters from the underlying mixed logit model 4.

The median willingness-to-pay for traffic information functionality ranges from 1.69 to 2.21 EUR per month, depending on the mode of sharing user location. These values correspond to the average level of privacy concerns in the sample, shown in section A. Higher WTP corresponds to a more preferred optional sharing. Willingness-to-pay for commercial information smaller and positive only if sharing is optional (0.67 EUR). Otherwise, a user is willing to pay 1.33 EUR for lack of obligation to use this functionality. Willingness-to-pay for an opt-out amounts to -8.06 EUR monthly, which can be interpreted as the level of a median user payment for lack of necessity to give up satellite-only navigation service. Finally, WTP levels for location-based functionalities are affected by the intensity of privacy concerns, as exhibited in sections B and C of Table 8. Most notably, WTP for choice attributes with mandatory location sharing are the most affected. The difference in WTP between users with high and low privacy concerns amounts to 1.87 EUR in case of commercial information and 1.62 EUR in case of traffic information. Privacy covariates affect significantly less WTP estimates for functionalities involving optional sharing. Willingness-to-pay provides information about the monetary valuation of a particular level of a single choice attribute. We now turn to the estimation of consumer surplus, which provides a comprehensive value of entire choice alternatives in money metric terms.

Based on expression (7) and estimates from model 4 we simulate changes in gross consumer surplus from shifting from 'no access' alternative to the four variants of navigation services, described in Section 3.1. Estimates of ΔCS are presented in Table 9, in four horizontal sections (1-4) corresponding to each considered navigation variant. We simulate the entire distributions of gross consumer surplus, but to ease exposition, we provide the estimates for the three points from these distributions: median (50th), 90th and 10th quantile. See the corresponding vertical sections [A-C] in Table 9. The estimates of gross consumer surplus are additionally differentiated based on the levels of privacy covariates in order to account for the impact of privacy concerns inside each section A-C.

We start from highlighting median values of consumer surplus estimates for an average level of privacy concerns, presented in the last column in section A. The state-of-the-art and user-centric navigation generates the highest monthly consumer surplus of 10.98 EUR. Compared to the value of satellite-only 'old' type service (8.06 EUR), adding location-based functionalities increases surplus from the service by 36%. However, these services very much depend on how the location sharing conditions are shaped. Most notably, if the provider requires mandatory location sharing of both types of location, the value of such provider-centric, state-of-the-

[16] The exact values for PC1 and PC2 taken for computations are provided Table 6, section B.

art navigation would not stay on the same level as for the satellite-only variant (8.2 EUR). The users are, of course, very suspicious about sharing personalised location and this condition keeps the value of the navigation service at a significantly lower level. For example, consumer surplus from the balanced variant equals 10.26 EUR and is close to the most preferred user-centric navigation. This variant assumes mandatory sharing of only anonymised location in exchange for traffic information and optional sharing of more sensitive personal location. The second set of results refers to the impact of privacy concerns on median valuations. Obviously, for modelling, we have assumed that privacy covariates may affect only the means of random parameters of choice attributes involving location sharing. Hence, satellite-only navigation is not influenced by, as it can be seen in the last horizontal section of Table 9. The most affected variant by privacy concerns is the provider-centric state-of-the-art navigation. In this case consumer surplus ranges from 6.78 for highly concerned to 10.13 EUR for little afraid users. These numbers represent a quite noticeable change of -18% and +23% relative to the median surplus level in the entire sample. The remaining two variants of state-of-the-art navigation are much less affected by privacy concerns because they allow for a safer, user-controlled sharing of personalised location. Finally, looking at sections B-C of Table 9, we can get an idea of how consumer surplus is distributed in the tails. The value of user-centric navigation can range from 0 to 30.68 EUR monthly.[17] The high-end users (from the 90th quantile) have roughly 3 times higher valuation than the median user.

[17] The negative values of surplus for the 10th quantile should be treated with caution. They arise from the assumption of normal distribution for several choice parameters in the model.

5 Discussion and conclusions

At present, navigation services are available to billions of individual users, primarily in the form of free-of-charge mobile applications, like Google Maps. While free digital services undoubtedly contribute to consumer welfare, the value of this contribution is difficult for rigorous quantification. Typical economic measures based on expenditure, such as GDP, do not capture the value of navigation and other zero price services. One solution to the above problem is a direct estimation of welfare from stated-preference data extracted in a choice experiment. The present study followed this path and demonstrated how discrete choice methodology can be applied to a zero price service. In essence, we have estimated utility function on micro-level data and computed changes in welfare from a range of alternatives in money metric terms. We have designed a discrete choice experiment to elicit stated preferences regarding the use of satellite-enabled navigation and optional location-based functionalities: traffic information and commercial information. The latter functionality includes local search and display ads and sponsored location pins on the map. Traffic information functionality offers real-time insights to road accidents, traffic intensity and travel time.

The experiment was conducted on a representative panel of internet users in Poland. The respondents were choosing multiple times between different generic variants of paid navigation and an opt-out option. It was assumed that in exchange for location-based functionality, an app provider requires location data from the users either in anonymised or personalised form. Additionally, the experiment has tested two conditions of location sharing: optional and mandatory. After controlling for the sources of hypothetical bias, the stated-preference data was modelled with the mixed logit framework. Finally, in the post-estimation phase, willingness-to-pay and consumer surplus measures were estimated. The median value from navigation ranges from 8.06 EUR monthly for satellite-only service to 10.98 EUR for navigation with both location-enabled functionalities: real-time traffic information and commercial information. By means of a quick approximation, assuming 15 million of navigation users in Poland and 11 EUR as the median monthly estimate of consumer surplus, the value of the entire navigation ecosystem amounts to 2 billion EUR annually just in the non-professional, civil segment. By nature of zero pricing, these benefits are not included in national accounts.

Free digital services have negligible marginal costs and substantial fixed development that has to be covered from a revenue stream. Zero price navigation services rely mainly on monetization of location data through local search and display ads or sponsored location pins on the map. Location with personal identifiers has far greater potential for monetisation than anonymised data. Therefore, navigation providers have strong incentives to collect the former type of location on a permanent basis. Our study shows that incentives of service providers are, largely, misaligned with the preferences of the users. The users attach importance to the way and the purpose their location data is collected for. They perceive location as sensitive data and prefer more controlled, optional sharing for a particular purpose. This preference for an optional sharing location is quite pronounced in our study and heavily affects the value of state-of-the-art navigation variants. On this end, location-based functionalities can increase the value of satellite-only navigation by 36%. On the other extreme, provider-centric navigation variant with mandatory sharing of personal and anonymous location generates just the same value as the 'old' navigation. Additionally, as this study has shown, individual preferences for sharing location data, in particular on a permanent basis, are affected by privacy concerns that the users hold. These concerns relate in the first place to the perceived risk of unauthorised use of personal data in the future.

The present study offers two main takeaways for shaping policies around the use of free digital services. First, if user data serves as an alternative payment for a service, then the terms of sharing will heavily affect the valuation of the service. The example of commercial information functionality from the present study illustrates that the contribution to the well-being can be positive or negative depending on the sharing conditions. From a user perspective, providers should collect personal data in the minimal possible extent necessary for the operation of a service. The user-centric approach, already adopted in the GDPR, conflicts with the appetite of the service providers for data. Therefore, it is an obvious area for the regulatory oversight and enforcement. Regulators should enforce certain best practices and minimalistic design principles, if they want to maximize the surplus of users of digital services. Second, for the valuation of free digital services, not only conditions but also the purpose sharing matters. There is no such a thing like general propensity to share data. The users are willing to disclose their data in exchange for access to a particular service, such as traffic information or commercial information in the case of navigation. Some services are more valuable than others are and this differentiation will affect users' propensity to share data. In real-life examples, service terms and conditions are often not precise in stating the purpose of disclosing user data. This introduces uncertainty and limits the comparability of services from different providers. Greater transparency of service

agreements would improve the conditions of making informed and optimal choices by users of free digital services. Lastly, even with transparent terms of service, the well-being will still be affected by the risk of unauthorised reuse of the data later. Because of the economic nature of data, regulators should mitigate such risk, for example by promoting data portability and the right to be forgotten. The regulators should also consider prohibitions of certain practices related to pooling user data or enforcing a sign-in to a new service solely with the intention to combine new sources of data.

Table 9. Monthly change in consumer surplus from shifting between different navigation service configurations, in EUR.

	[A] ΔCS sample median			**[B] ΔCS sample q.90**			**[C] ΔCS sample q.10**		
	high privacy concerns (q.90)	low privacy concerns (q.10)	sample avr. privacy concerns	high privacy concerns (q.90)	low privacy concerns (q.10)	sample avr. privacy concerns	high privacy concerns (q.90)	low privacy concerns (q.10)	sample avr. privacy concerns
	[1] change from 'no access'[(a)] to 'user-centric, state-of-the-art' navigation [(b)]								
50%	10.52	11.62	10.98	29.86	31.83	30.68	0.2	1,27	0.66
95% c.i.	(9.56; 11.52)	(10.43; 12.89)	(10.27; 11.72)	(26.86; 33.15)	(28.54; 35.44)	(27.87; 33.75)	(-1.34; 1.48)	(-0.29; 2.55)	(-0.59; 1.7)
(se)	(0.51)	(0.63)	(0.37)	(1.6)	(1.77)	(1.49)	(0.73)	(0.71)	(0.59)
	[2] change from 'no access'[(a)] to 'balanced, state-of-the-art' navigation [(c)]								
50%	9.34	11.51	10.26	29.1	33.15	• 30.8	-1.38	0.83	-0.4
95% c.i.	(8.41; 10.3)	(10.28; 12.78)	(9.59; 10.97)	(16.17; 32.26)	(29.67; 37.06)	(27.99; 33.86)	(-3.17; 0.12)	(-0.76; 2.19)	(-1.78; 0.81)
(se)	(0.48)	(0.64)	(0.36)	(1.54)	(1.89)	(1.49)	(0.84)	(0.75)	(0.67)
	[3] change from 'no access'[(a)] to 'provider-centric, state-of-the-art' navigation [(d)]								
50%	6.78	10.13	8.2	22.69	28.64	25.13	-4.02	-0.01	-2.23
95% c.i.	(5.68; 7.82)	(8.75; 11.6)	(7.53; 8.89)	(20.18; 25.43)	(25.24; 32.35)	(22.8; 27.69)	(-5.95; -2.31)	(-1.69; 1.46)	(-3.52; 1.07)
(se)	(0.55)	(0.73)	(0.34)	(1.33)	(1.81)	(1.24)	(0.93)	(0.81)	(0.63)
	[4] change from 'no access'[(a)] to 'satellite-only, old' navigation[(e)]								
50%	8.06	8.06	8.06	25.4	25.4	25.4	-1.65	-1.65	-1.65
95% c.i.	(7.37; 8.75)	(7.37; 8.75)	(7.37; 8.75)	(22.88; 28.08)	(22.88; 28.08)	(22.88; 28.08)	(-2.88; -0.62)	(-2.88; -0.62)	(-2.88; -0.62)
(se)	(0.35)	(0.35)	(0.35)	(1.33)	(1.33)	(1.33)	(0.58)	(0.58)	(0.58)

Notes: (a) 'No access' option denotes lack of access to navigation; (b) 'user-centric, state-of-the-art' navigation stands for navigation with both location-based functionalities (traffic and commercial information) and optional sharing of both personal and anonymized location; (c) 'balanced, state-of-the-art' navigation requires optional sharing of personal location and optional of anonymized location; (d) 'provider-centric, state-of-the-art' navigation has user location sharing mandatory in personalized and anonymized modes; (e) 'Plain, off-line' navigation has location sharing completely blocked. ΔCS stands for change in consumer surplus. Values of ΔCS correspond to the three statistics of a simulated surplus distribution: [A] sample median and [B] tenth and [C] ninetieth quantiles. For each statistic [A-C], three estimates of the ΔCS are provided, which correspond to different levels of privacy concerns (PC1, PC2): high concerns; low concerns and sample average. For exact values of privacy covariates consult Table 8. Standard errors (se) and 95% confidence intervals of are obtained with Krinsky and Robb simulations using 10^4 draws of consumers and 10^4 draws of random parameters from the underlying mixed logit model 4.

References

Allcott, H., Braghieri, L., Eichmeyer, S., & Gentzkow, M. (2020). The welfare effects of social media. *American Economic Review*, *110*(3), 629–676.

Bliemer, M. C. J., & Rose, J. M. (2010). Construction of experimental designs for mixed logit models allowing for correlation across choice observations. *Transportation Research Part B: Methodological*, *44*(6), 720–734.

Bliemer, M. C. J., & Rose, J. M. (2013). Confidence intervals of willingness-to-pay for random coefficient logit models. *Transportation Research Part B: Methodological*, *58*, 199–214. https://doi.org/10.1016/j.trb.2013.09.010

Brynjolfsson, E., Collis, A., & Eggers, F. (2019). Using massive online choice experiments to measure changes in well-being. *Proceedings of the National Academy of Sciences*, *116*(15), 7250–7255.

Brynjolfsson, E., Eggers, F., & Gannamaneni, A. (2018). Measuring Welfare with Massive Online Choice Experiments: A Brief Introduction. *AEA Papers and Proceedings*, *108*, 473–476. https://doi.org/10.1257/pandp.20181035

Brynjolfsson, E., & Oh, J. (2012). *The attention economy: measuring the value of free digital services on the Internet*.

Buchanan, T., Paine, C., Joinson, A. N., & Reips, U. (2007). Development of measures of online privacy concern and protection for use on the Internet. *Journal of the American Society for Information Science and Technology*, *58*(2), 157–165.

Bughin, J. (2011). The Web's€ 100 billion surplus. *McKinsey Quarterly*, *2*(4), 1–4.

Carlsson, F., & Martinsson, P. (2001). Do hypothetical and actual marginal willingness to pay differ in choice experiments?: Application to the valuation of the environment. *Journal of Environmental Economics and Management*, *41*(2), 179–192.

Carson, R. T., Flores, N. E., Martin, K. M., & Wright, J. L. (1996). Contingent valuation and revealed preference methodologies: comparing the estimates for quasi-public goods. *Land Economics*, 80–99.

Carson, R. T., & Groves, T. (2007). Incentive and informational properties of preference questions. *Environmental and Resource Economics*, *37*(1), 181–210.

Daziano, R. A., Sarrias, M., & Leard, B. (2017). Are consumers willing to pay to let cars drive for them? Analyzing response to autonomous vehicles. *Transportation Research Part C: Emerging Technologies*, *78*, 150–164.

Goolsbee, A., & Klenow, P. J. (2006). Valuing consumer products by the time spent using them: An application to the Internet. *American Economic Review*, *96*(2), 108–113.

Hanley, N., Wright, R. E., & Alvarez-Farizo, B. (2007). Estimating the economic value of improvements in river ecology using choice experiments: an application to the water framework directive. In *Environmental value transfer: Issues and methods* (pp. 111–130). Springer.

Hensher, D. A., & Greene, W. H. (2003). The mixed logit model: the state of practice. *Transportation*, *30*(2), 133–176.

Lancaster, K. J. (1966). A new approach to consumer theory. *Journal of Political Economy*, *74*(2), 132–157.

Louviere, J. J., Flynn, T. N., & Carson, R. T. (2010). Discrete choice experiments are not conjoint analysis. *Journal of Choice Modelling*, *3*(3), 57–72.

Louviere, J. J., Hensher, D. A., & Swait, J. D. (2000). *Stated choice methods: analysis and applications*. Cambridge university press.

Louviere, J. J., & Woodworth, G. (1983). Design and analysis of simulated consumer choice or allocation experiments: an approach based on aggregate data. *Journal of Marketing Research*, *20*(4), 350–367.

McFadden, D. (1973). *Conditional logit analysis of qualitative choice behavior*.

Meyerhoff, J., & Liebe, U. (2008). Do protest responses to a contingent valuation question and a choice experiment differ? *Environmental and Resource Economics*, *39*(4), 433–446.

Nagaraj, A., & Stern, S. (2020). The Economics of Maps. *Journal of Economic Perspectives*, *34*(1), 196–221.

Pantea, S., & Martens, B. (2014). The value of the Internet for consumers. *Available at SSRN 2446962*.

Potoglou, D., Palacios, J.-F., & Feijóo, C. (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.

Sandor, Z., & Wedel, M. (2001). Designing conjoint choice experiments using managers' prior beliefs. *Journal of Marketing Research*, *38*(4), 430–444.

Scarpa, R., & Rose, J. M. (2008). Design efficiency for non-market valuation with choice modelling: how to measure it, what to report and why. *Australian Journal of Agricultural and Resource Economics*, *52*(3), 253–282.

Small, K. A., & Rosen, H. S. (1981). Applied welfare economics with discrete choice models. *Econometrica (Pre-1986)*, *49*(1), 105.

Tirole, J., & Rochet, J. C. (2006). Two Sided Markets: A Progress Report. *RAND Journal of Economics*, *37*(3), 645–667. https://doi.org/10.1080/0092623X.2014.966397

Train, K. E. (2009). *Discrete choice methods with simulation*. Cambridge university press.

Varian, H. R., Farrell, J., & Shapiro, C. (2004). *The economics of information technology: An introduction*. Cambridge University Press.

Vossler, C. A., Doyon, M., & Rondeau, D. (2012). Truth in consequentiality: theory and field evidence on discrete choice experiments. *American Economic Journal: Microeconomics*, *4*(4), 145–171.

List of abbreviations and definitions

CS	Consumer surplus
GDP	Gross domestic product
GDPR	General Data Protection Regulation
GPS	Global Positioning System
MNL	Multinomial logit model
MXL	Mixed logit model
SSNIP	Small, significant, non-transitory increase in price
SSNIQ	Small, significant, non-transitory increase in quality
WTA	Willingness-to-accept
WTP	Willingness-to-pay

List of tables

The European Commission's science and knowledge service

Joint Research Centre

JRC Mission

As the science and knowledge service of the European Commission, the Joint Research Centre's mission is to support EU policies with independent evidence throughout the whole policy cycle.

EU Science Hub
ec.europa.eu/jrc

@EU_ScienceHub

EU Science Hub - Joint Research Centre

EU Science, Research and Innovation

EU Science Hub