Willingness to Pay for the Protection of Different Data Types

Vera Schmitt
Quality and Usability Lab Technical University Berlin
Berlin, Germany
vera.schmitt@tu-berlin.de

Sebastian Möller
Quality and Usability Lab Technical University Berlin
Berlin, Germany
sebastian.moeller@tu-berlin.de

Maija Poikela
Fraunhofer Institute for Applied and Integrated Security AISEC
Berlin, Germany
maija.poikela@aisec.fraunhofer.de

ABSTRACT

The standard approach of *Notice and Choice* does not provide sufficient control over personal privacy preferences. A more granular analysis of privacy preferences is needed where the monetary valuation of different data types can contribute to the understanding of individual privacy concerns of personal information. The question of how much consumers value their privacy is still underexplored. Therefore, this study examines previous approaches of monetary valuation of different data types and analyses the monetary valuation for two different countries to verify earlier research results.

KEYWORDS

Privacy, Monetary Valuation, Country Comparison, Willingness to Pay, Privacy Concern, Data Types

1 INTRODUCTION

Invasive practices for the collection and use of personal information have shown that there is a widespread agreement on the need for privacy oversight and governmental regulations [31]. Incidents such as Facebook's Cambridge Analytica scandal [13] and high-profile data breaches (e.g., Equifax [32]) created a general unease about access to personal information and increasing privacy concern among governments and businesses alike [23].

However, as the collection and usage of personal information has grown drastically over the last decade, balancing privacy preferences and benefits has become a nontrivial task [15]. According to Parkins [20], personal information and behavioral data are the world's most valuable resources. The extensive sharing and usage of information online has a powerful impact on the development of new applications and the economy itself as various business models have been developed solely based on continuous data sharing and usage of online user information [21]. The amount of user data an IT company holds nowadays has a direct and profound contribution to its overall market valuation [25]. Digital advertising is one of the most important application to monetize user data [19] and, according to [12], global advertising spending has constantly increased since 2010 and is expected to grow to nearly 650 billion U.S. dollars in the end of 2021.

Mensch und Computer 2021, Workshopband, Workshop on 7. Usable Security und Privacy Workshop

https://doi.org/10.18420/muc2021-mci-ws14-390

To increase the market share of companies, targeted advertising has become an effective tool to reach targeted customer groups. Consequently, the collection of user data has become more aggressive [19] triggering a public debate around the trade-offs between innovation and civil rights such as personal data protection [14, 28, 29]. Surprisingly, there is not much empirical evidence on how individuals value their personal data and also different types of information such as location information, financial or medical data. Therefore, we will shed light on the monetary valuation of personal information of different data types and propose the following research questions:

> **RQ1**: *To what extent does the maximum amount people are willing to pay for the protection of their personal information diverge in different countries?*
> **RQ2**: *Based on the monetary valuation, which data types are the most relevant for the respective country?*

The estimation on how people value personal information and how these values vary across countries and different data types is still an underexplored problem. The monetary quantification of privacy can contribute to the evaluation and analysis of privacy policies in place, such as the General Data Protection Directive (GDPR) [16]. The monetary valuation of different data types from an economic perspective gives a more granular view on which data types are of most importance for consumers and, therefore, require special attention. Moreover, the monetary valuation can deviate from country to county [22] and even from region to region [23]. Thus, empirical evidence about the quantification of different data types needs to be collected from different countries. In sum, the contributions of this work are the following:

- Definition and evaluation of different data types and the corresponding value categories
- Comparison of the monetary valuation of different data types in two different countries.

The remainder of this paper is organized as follows. Section 2 discusses previous research on valuation of personal information. Section 3 describes the design of our experiment. In Section 4 the results are presented, and finally, Section 5 summarizes our discussion, outlines our conclusions, and presents future research directions.

2 BACKGROUND AND RELATED WORK

Notice and Choice is the predominant method to protect the consumers data. *Notice* gives the consumer information about what data are collected and used, while *Choice* enables users to choose whether or not their data can be collected or used in that way [4]. This approach has not led to major improvements in terms of protecting the users' privacy [4]. Social media platforms and digital

service providers still extract a great deal of data from users who are often nudged into less privacy friendly options [31].

The protection and disclosure of personal information often generate trade-offs which also have an economic dimension. The spread of mobile computing and sensor technologies facilitates the collection of personal information and turns mere consumers of information into public producers of often highly personal data [2]. Thus, the large amounts of collected personal information have an substantial economic value. Individual attributes, such as demographic information and online behavior, are increasingly regarded as business assets [16, 24, 30]. Improper collection and use of personal data have contributed to the development of governmental regulations such as the *General Data Protection Directive* (GDPR) in the EU or California's *Consumer Privacy Act* (CCPA). These efforts have been made to strengthen the consumer's control over personal information [16, 23].

Nevertheless, the tools and products made possible by the increased availability of personal data have yielded benefits for data subjects and data holders alike [2]. Whereas the economic value of data holders can be inferred more transparently from yearly revenues, there is not much empirical evidence about how data subjects value their personal information and if they value different types of information equally. Previous research has shown that the provision of economic valuations of personal data depend on the kind of information to be released [6, 8, 10, 23, 30]. Most of the previous studies explicitly or implicitly measure the amount of money consumers consider to be enough to share their personal information, namely the *Willingness to Accept* (WTA) [23, 28]. Less research has been done about tangible prices or intangible costs consumers are *Willing to Pay* (WTP) to protect their personal information [22, 26].

WTP and WTA are well established measures of how consumers value different kinds of products and is effective especially in scenarios where the consumer is familiar with the product, e.g. paying for the usage of a social media outlet they have experience with [3]. In scenarios, where the consumer is asked to pay for a good with ambiguous and unclear effects and consequences on the consumers life, it is questionable whether asking consumers for monetary valuation of their personal information reveals reliable insights [27]. However, in real markets involving data privacy, WTP and WTA are highly relevant [1, 31] and can be also used as proxy for the privacy concern [22]. Therefore, further investigation on WTP and WTA in scenarios where consumers are asked to value different types of data can facilitate the understanding which data types are of most concern [23]. Accordingly, protection mechanisms can be developed and improved from a governmental and commercial perspective. Hereby, the monetary valuation can shed some some light on the construct of privacy concern and foster an in-depth understanding.

Previous studies [1–3, 23, 31] have shown that using WTP and WTA to measure the individual valuation of privacy can provide insights to the question of how much is privacy worth. This can foster the discussion for developing different approaches of access to services and control over personal information than *Notice an Choice*.

3 EXPERIMENT DESIGN

In a recent study of Prince and Wallsten [23], the monetary valuation for different types of information was explored in a cross-country comparison. Hereby, the monetary values differed greatly among Mexico, Brazil, Colombia, Argentina, the United States and Germany, where Germany yielded the highest monetary values in exchange for personal information. Based on these findings, we employed discrete-choice surveys in English to compare the monetary valuation of different data types between Germany and Pakistan. Pakistan has not yet adopted data protection standards similar to the EU, Japan or the US and it is very little known about individual privacy perception of consumers in Pakistan. A recent publication of Imtiaz et al. [11] describes, that people from Pakistan mistrust the e-commerce environment in their country, and thus, often refuse to share their personal information. The special interest of comparing Pakistan and Germany resulted from the student group, who participated in the preparation of the study. Within the student group the nationalities from both countries were represented, and therefore, we run the online survey in both countries in order to shed further light on the valuation of different data categories.

The different data categories used in this study are related to the distinction made by Prince and Wallsten [23] but for our scope these have been simplified and mapped into the following data categories:

- **Personal Data**: e.g. name, age, date of birth and personal registration number.
- **Location Data**: e.g. location traces collected by navigation services which reveal the daily traveling routines and places where users spent time at.
- **Medical Records**: e.g. data stored by health insurance companies about the individual health situations and how often persons visit the doctor and for which purpose.
- **Web Activity Data**: data collected by search engines like Google to create a profile about personal interests to show relevant advertisements to users.
- **Financial Data**: data stored by financial services which reveal personal shopping habits and information about the financial situation of customers.

These different data types are taken as basis to ask the participants about their monetary valuation to explore which data types are of most importance and whether this holds true for different cultural contexts. In contrast to the approach of Prince and Wallsten [23], the survey focuses on the construct of WTP to retain privacy instead of WTA to give up various forms of privacy. The choice is driven by the circumstance that there is less empirical evidence for WTP for privacy in general and for the different data categories.

3.1 Value Categories

The estimation of WTP is based on discrete choice questions. The participants are given a scenario where they are asked how much they would pay on a monthly basis to keep the respective data type protected. As most companies are interested in continuous data sharing, the scenario of a subscription model for data protection is a more realistic setting. Furthermore, clarifying the time period of data sharing is crucial to provide more details about the sharing scenario to the participants. Hereby, the data requester was not

specified as we were aiming to study the general monetary monetary valuation of different data types by keeping other variables constant. Although previous research states that privacy perceptions and also monetary valuation is context dependent [18], the description of a general data sharing scenario was used to avoid the introduction of any additional biases [6]. In the hypothetical valuation scenarios, the participants could select the following discrete value categories:

Table 1: Discrete Value Categories in Euro and Rupee

Germany	Pakistan
0 € (nothing)	0 ₹ (nothing)
0-2 €	0-150 ₹
2-4 €	150-300 ₹
4-6 €	300-450 ₹
6 € and more	450 ₹ and more

Furthermore, the options *nothing* and *nothing, I cannot afford it* were added to give the participants extended options to state their preferences. These value categories were available for both currencies, Euro and Rupee. In order to find the respective value category in Rupee, we compared the prices of several items which are used and consumed in both countries on a daily basis. Thus, the value categories do not match the official exchange rate of both currencies. The validation of the different value categories based on the values of daily items resulted in lower value categories for Pakistan than the official exchange rate would yield. This also adjusts for the lower income in Pakistan compared to Germany.

As shown in previous research [7], discrete choice categories mitigate reporting inaccuracy of stated preferences. Even if the categories remain hypothetical, the estimation for changes in feature levels is statistically unbiased at least for WTP estimates [9, 17]. However, a reliable discrete-choice method requires careful design to motivate the participants to answer as truthfully as possible [5]. Therefore, the questionnaire contains three different sections.

The first sections entails questions to get demographic information from the participants such as age, ethnicity, gender and household income. The second part of the questionnaire is about general privacy concern and privacy awareness. In the third part, the participants were confronted with different data sharing scenarios where they were asked how much they would pay in order to protect different types of information from being shared with a unknown data requester.

The study was conducted in English in both countries and the participation was on a voluntary basis as the participants did not receive any payment. The online survey was distributed mostly among friends, families, colleagues and fellow students. Participants from Pakistan received the survey containing all value categories in Rupees. German participants received an identical survey only differing in the value categories, which were presented in Euros.

4 RESULTS AND DISCUSSION

4.1 Sample Demographics and General Privacy Concern

Overall, 85 participants took part in the survey where 40 participants were from Germany and 45 from Pakistan. As the participation in the survey was voluntarily and not rewarded with any payment or other incentives, a representative sample was difficult to achieve. The sample consists mostly of male participants, 84,6% male participants in Germany and 93% in Pakistan. Furthermore, 90% of German participants are students but only 31% of participants from Pakistan. 47% of the German participants have a background in IT but only 17% of participants from Pakistan.

Furthermore, 57,8% of participants from Pakistan are generally willing to pay for their data protection on a monthly basis compared to 62,5% of German participants. This might be due to the fact that the poverty rate in Pakistan is much higher than in Germany. On average, most of the participants are willing to pay between 5-10 Euros when asked for a monthly payment. Even though the sample is biased towards male IT-students in Germany and male professionals without IT background in Pakistan, the following analysis yields some interesting findings regarding the valuation of different data types.

In order to make the monetary valuation comparable, the value categories in Rupee where mapped to the value categories in Euros on basis of the above mentioned adjusted exchange rate. Thus, all the following monetary values are reported in Euro. The participants were asked to report the maximum value they would be willing to pay on a monthly basis for the protection of all their data.

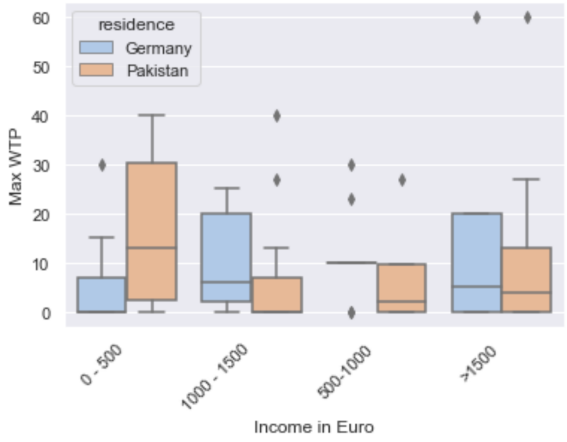

Figure 1: WTP Max Displayed with Income groups of Participants from Germany and Pakistan

Figure 1 gives insights about the relation of income and the maximum WTP for both countries. In Pakistan, the participants for the income group of 0-500€ report a significantly higher WTP compared to Germany. It can be inferred that people from low income groups in Pakistan are more willing to pay a higher price to protect their data. This can be due to the fact that career opportunities are highly dependent on the social status in Pakistan. By revealing

information about disadvantaged circumstances, it can have severe consequences on future career opportunities.

Similar insights can be drawn from Figure 2 where the average of the maximum WTP for the respective age group is displayed. Younger participants (between 18 and 25 years) in Germany are not willing to pay as much for their monthly data protection as participants from Pakistan which might confirm the assumption that participants from Pakistan are more careful regarding personal information, especially when it might influence their future career. This trend turns around for older age groups, where German participants seem to be willing to pay higher amounts, up to on 60€ on a monthly basis, for the protection of all their data, compared to up to 6€ in Pakistan.

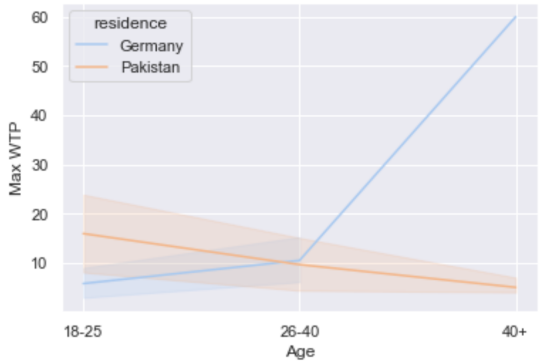

Figure 2: WTP Max Displayed with Age of Participants from Germany and Pakistan

When it comes to the maximum WTP in combination with general privacy concerns, major differences between Germany and Pakistan are visible from Figure 3. As the residuals of the data are not normally distributed we need to rely on the nonparametric alternative. We conducted a Mann-Whitney U test to find significant differences of the privacy concern categories between the two countries. For the category *very concerned* Pakistan yields significantly higher privacy concern than Germany ($X^2(1) = 133$, *p-value* = 0.004). German participants reported mostly that they know that they should be concerned, but actually are not. One explanation we found in later interviews with a subset of participants was, that enforcement of the GDPR makes the participants feel more protected than without any regulation in place.

Even though most of the German participants reported only a low privacy concern, they are still willing to pay on average up to 30€ per month for the protection of their data. Interestingly, participants from Pakistan tend to have a higher privacy concern, and thus, are willing to pay on average up to 40€ on a monthly basis for their data protection.

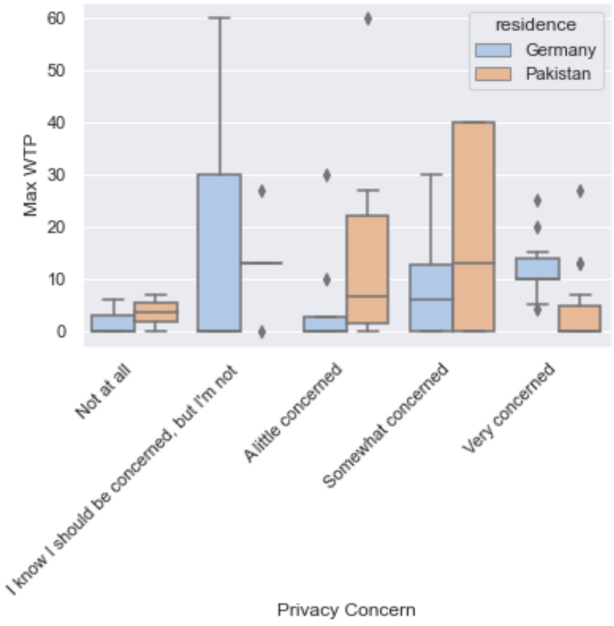

Figure 3: WTP Max and General Privacy Concern

4.2 Valuation of Different Data Types

The participants where asked how much they would spent to protect the different data categories described in Section 3. Almost half the participants in both countries were not willing to pay for data protection of any data category, as shown in Figure 4 and Figure 5.

Figure 4: WTP for Different Data Types and Value Categories (according to Table 1) for Germany

Figure 5: WTP for Different Data Types and Value Categories (according to Table 1) for Pakistan

For a more fine grain analysis we removed the participants who are not willing to pay for any data category, in order to compare the different value category for each data type, see Figure 6 for Germany and Figure 7 for Pakistan.

Figure 6: WTP for Different Data Categories without Non-Payers in Germany

We conducted a Kruskal-Wallis test in order to find differences between both countries with respect to the different data categories. The Kruskal-Wallis test yielded significant differences ($X^2(4) = 13.3$, *p-value*= 0.001). Further pairwise Mann-Whitney U-tests with Bonferroni correction (the new alpha level being $\alpha = 0.012$) showed only significant differences for the financial data type. Participants from Pakistan reported a significantly higher WTP for financial data compared to German participants, by selecting mostly the value category of "6 € and more".

The analysis of different data types within Germany reveals that after pairwise comparison with the Mann-Whitney U test with Bonferroni correction (the new alpha level being $\alpha = 0.005$) no significant differences between the monetary valuation of the different data types can be found. In contrary, the data types *Financial*

Figure 7: WTP for Different Data Categories without Non-Payers in Pakistan

and *Location*, and *Financial* and *Medical* differed significantly in Pakistan. Participants from Pakistan are willing to pay substantially more for the protection of financial data than location or medical data, by selecting the value category "6 € and more" most often for the financial data type.

Although no significant differences can be found for Germany, the Figures 4 and 6 show that in Germany data categories such as location information, financial data and personal information are valued the most, whereas in Pakistan the WTP for the protection of personal information, financial data and information about web activities are valued higher than location information and medical data. The comparison of the maximum WTP in both countries with a Mann-Whitney U test, did not result in significant differences ($X^2(1) = 875.0$, *p-value*= 0.82).

5 CONCLUSION

One major drawback of the study was the voluntary participation of all participants, which led to an biased sample, which might influence the results and might lead to false assumptions. Similar to the study of Prince and Wallsten [23] we found that German participants value financial data also quite high. Different values where achieved for location information and also medical records, which might be also caused by the biased sample. Furthermore, the importance of different data categories when comparing different countries can also be confirmed by our findings, even though they might be biased due to the skewed sample. Therefore, the research questions stated in the beginning, can be answered as follows:

RQ1: To what extent does the maximum amount people are willing to pay for protection of their personal information differs in different countries? There are no significant differences between Germany and Pakistan when participants were asked about the maximum amount of money to pay for the protection for their data.

RQ2: Based on the monetary valuation, which data types are the most relevant for the respective country? The WTP for different data categories differs significantly only for the *Financial* data type. Nevertheless, when comparing the other

data types, financial data, location information and medical records are valued the most in Germany, and in Pakistan financial data, personal information and web activity are perceived as more valuable.

However, as Solove [27] emphasizes, the approach of monetary valuation of an abstract concept such as privacy is a difficult endeavour. The concept of privacy is often not well understood and it is not clear, whether asking participants for monetary valuation might disclose the true value of privacy. Nevertheless, the WTP for different data categories reveals which data types are perceived as more important than others. The perception of which data categories are perceived as more valuable can be used as proxy to gain a more accurate insights of privacy concern. Moreover, more attention can be given to these data categories in future formulations of data protection regulations or for strategic advantages for companies, who focus on the design of products which entail data protection mechanisms.

6 ACKNOWLEDGEMENTS

The study was conducted with the support of Hamza Ahmed Khan, Hamza Ahmed Siddiqui, Paul Wille, and Felix Bublitz, who contributed to the construction of the survey and definition of the value categories.

REFERENCES

[1] Alessandro Acquisti, Leslie K John, and George Loewenstein. 2013. What is privacy worth? *The Journal of Legal Studies* 42, 2 (2013), 249–274.
[2] Alessandro Acquisti, Curtis Taylor, and Liad Wagman. 2016. The economics of privacy. *Journal of economic Literature* 54, 2 (2016), 442–92.
[3] Hunt Allcott, Luca Braghieri, Sarah Eichmeyer, and Matthew Gentzkow. 2020. The welfare effects of social media. *American Economic Review* 110, 3 (2020), 629–76.
[4] Susan Athey, Christian Catalini, and Catherine Tucker. 2017. *The digital privacy paradox: Small money, small costs, small talk.* Technical Report. National Bureau of Economic Research.
[5] Moshe Emanuel Ben-Akiva, Daniel McFadden, Kenneth Train, et al. 2019. Foundations of stated preference elicitation: Consumer behavior and choice-based conjoint analysis. *Foundations and Trends (R) in Econometrics* 10, 1-2 (2019), 1–144.
[6] Joseph R Buckman, Jesse C Bockstedt, and Matthew J Hashim. 2019. Relative privacy valuations under varying disclosure characteristics. *Information Systems Research* 30, 2 (2019), 375–388.
[7] Octavian Carare, Chris McGovern, Raquel Noriega, and Jay Schwarz. 2015. The willingness to pay for broadband of non-adopters in the US: Estimates from a multi-state survey. *Information Economics and Policy* 30 (2015), 19–35.
[8] George Danezis, Stephen Lewis, and Ross J Anderson. 2005. How much is location privacy worth?. In *Fourth workshop on the economics of information security*. WEIS, Harvard University, Cambridge, 1–13.
[9] Min Ding, Rajdeep Grewal, and John Liechty. 2005. Incentive-aligned conjoint analysis. *Journal of marketing research* 42, 1 (2005), 67–82.
[10] Bernardo A Huberman, Eytan Adar, and Leslie R Fine. 2005. Valuating privacy. *IEEE security & privacy* 3, 5 (2005), 22–25.
[11] Shoaib Imtiaz, Syed Hassan Ali, and Dong Jin Kim. 2020. E-Commerce Growth in Pakistan: Privacy, Security, and Trust as Potential Issues. *Culinary Science & Hospitality Research* 26, 2 (2020), 10–18.
[12] Statista Inc. 2021. Premium Digital advertising spending worldwide from 2019 to 2024 (in billion U.S. dollars). Retrieved June 7, 2021 from https://www.statista.com/statistics/237974/online-advertising-spending-worldwide/
[13] Jim Isaak and Mina J Hanna. 2018. User data privacy: Facebook, Cambridge Analytica, and privacy protection. *Computer* 51, 8 (2018), 56–59.
[14] A-Reum Jung. 2017. The influence of perceived ad relevance on social media advertising: An empirical examination of a mediating role of privacy concern. *Computers in Human Behavior* 70 (2017), 303–309.
[15] Michael Kummer and Patrick Schulte. 2019. When private information settles the bill: Money and privacy in Google's market for smartphone applications. *Management Science* 65, 8 (2019), 3470–3494.
[16] Gianclaudio Malgieri and Bart Custers. 2018. Pricing privacy–the right to know the value of your personal data. *Computer Law & Security Review* 34, 2 (2018), 289–303.
[17] Klaus M Miller, Reto Hofstetter, Harley Krohmer, and Z John Zhang. 2011. How should consumers' willingness to pay be measured? An empirical comparison of state-of-the-art approaches. *Journal of Marketing Research* 48, 1 (2011), 172–184.
[18] Helen Nissenbaum. 2004. Privacy as contextual integrity. *Wash. L. Rev.* 79 (2004), 119.
[19] Panagiotis Papadopoulos, Nicolas Kourtellis, Pablo Rodriguez Rodriguez, and Nikolaos Laoutaris. 2017. If you are not paying for it, you are the product: How much do advertisers pay to reach you?. In *Proceedings of the 2017 Internet Measurement Conference*. IMC, London, 142–156.
[20] David Parkins. 2017. *The world's most valuable resource is no longer oil, but data.* The economist. https://www.economist.com/leaders/2017/05/06/the-worlds-most-valuable-resource-is-no-longer-oil-but-data
[21] Jian Pei. 2020. A survey on data pricing: from economics to data science. *IEEE Transactions on Knowledge and Data Engineering* 32 (2020), 1–1. https://doi.org/10.1109/TKDE.2021.3073062
[22] Maija Poikela and Eran Toch. 2017. Understanding the valuation of location privacy: a crowdsourcing-based approach. In *Proceedings of the 50th Hawaii International Conference on System Sciences*. HICSS, Hawaii, 1985–1994.
[23] Jeffrey Prince and Scott Wallsten. 2020. How Much is Privacy Worth Around the World and Across Platforms?. In *The 48th Research Conference on Communication, Information and Internet Policy*. TPRC, Online, 1–44.
[24] Stephen Cory Robinson. 2017. What's your anonymity worth? Establishing a marketplace for the valuation and control of individuals' anonymity and personal data. *Digital Policy, Regulation and Governance* 19, 5 (2017), 353–366.
[25] Judy Selby. 2016. The impact of big data decisions on business valuations. Retrieved June 8, 2021 from https://bigdata-madesimple.com/how-big-data-decisions-impact-business-valuations/
[26] Anya Skatova, Rebecca Louise McDonald, Sinong Ma, and Carsten Maple. 2019. Unpacking Privacy: Willingness to pay to protect personal data.
[27] Daniel J Solove. 2021. The myth of the privacy paradox. *Geo. Wash. L. Rev.* 89 (2021), 1.
[28] Jacopo Staiano, Nuria Oliver, Bruno Lepri, Rodrigo de Oliveira, Michele Caraviello, and Nicu Sebe. 2014. Money walks: a human-centric study on the economics of personal mobile data. In *Proceedings of the 2014 ACM International Joint Conference on Pervasive and Ubiquitous Computing*. UbiComp, New York, 583–594.
[29] Catherine E Tucker. 2014. Social networks, personalized advertising, and privacy controls. *Journal of marketing research* 51, 5 (2014), 546–562.
[30] Jessica Vitak, Yuting Liao, Priya Kumar, Michael Zimmer, and Katherine Kritikos. 2018. Privacy attitudes and data valuation among fitness tracker users. In *International Conference on Information*. ICIS, San Francisco, 229–239.
[31] Angela G Winegar and Cass R Sunstein. 2019. How much is data privacy worth? a preliminary investigation. *Journal of Consumer Policy* 42, 3 (2019), 425–440.
[32] Yixin Zou, Abraham H Mhaidli, Austin McCall, and Florian Schaub. 2018. " I've Got Nothing to Lose": Consumers' Risk Perceptions and Protective Actions after the Equifax Data Breach. In *Fourteenth Symposium on Usable Privacy and Security ({SOUPS} 2018)*. SOUPS, Baltimore, 197–216.