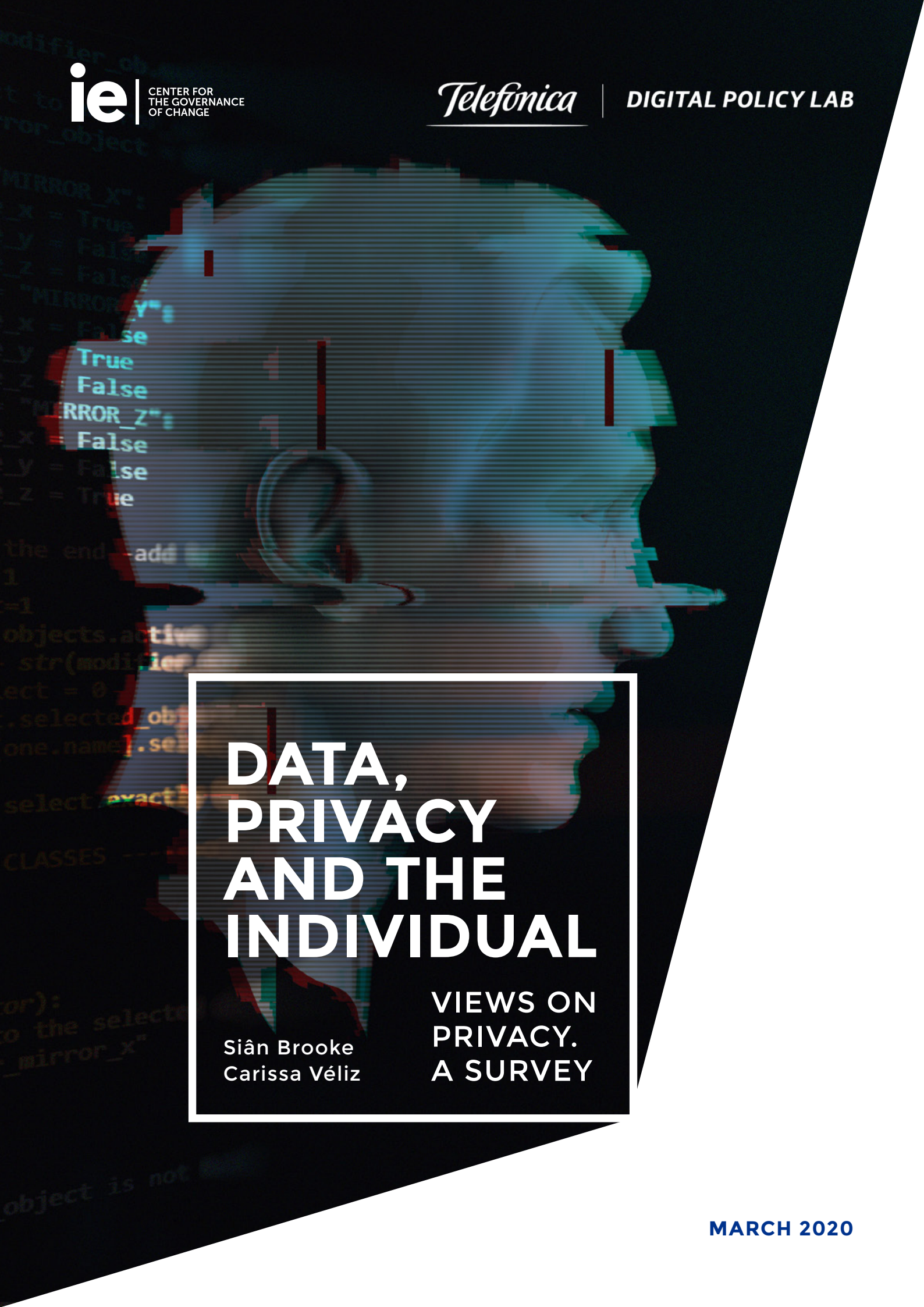
ie | CENTER FOR THE GOVERNANCE OF CHANGE

Telefónica | DIGITAL POLICY LAB



DATA, PRIVACY AND THE INDIVIDUAL

VIEWS ON PRIVACY. A SURVEY

Siân Brooke
Carissa Véliz

MARCH 2020

Background

The purpose of this survey was to gather individual's attitudes and feelings towards privacy and the selling of data. A total (N) of 1,107 people responded to the survey. We conducted an online survey through a distribution platform (Amazon's Mechanical Turk) which has been developed through Qualtrics software.

The survey covers:

(1) Experiences of online data and privacy breaches
(2) Concerns regarding privacy
(3) The use of personal data by companies
(4) The use of personal data by institutions
(5) Trust in companies and institutions
(6) Acceptability of monetising privacy
(7) Bulk data collection by governments

Author contributions:
Carissa Véliz initiated the project, wrote the first draft of the survey, and co-edited the survey report. Siân Brooke co-designed the survey instrument, distributed the survey, carried out the analysis of the data, and wrote up the results, including a comparison with previous work.

Contents

Survey Design

This section of the document will serve to outline the design of the survey and questions asked at each stage. Please note that later in the analysis multiple questions are used to formulate a singular measure, as is the norm with Likert scales. All questions had a "Prefer not to say" option, which allowed us to count purposefully unanswered questions, rather than just incomplete surveys. This option also meant that respondents could decline to answer any specific question, whilst still completing the majority of the survey. The survey was piloted before full release to ensure credible survey design. The pilot took place in two batches of 20 responses, a week apart, in early November 2019. The full survey was rolled out on the 12th of November 2019 and the final batch was completed on the 23rd of December 2019.

DEMOGRAPHICS

As is standard with a survey-based methodology, the first data that was collected was demographics. Responses to these questions were singular choice and participants selected the appropriate category that best described them. Note that non-binary and self-description options are available for gender identification, as is good practise when conducting surveys.

The demographic information collected was;

(1) gender identification
(2) age
(3) nationality
(4) highest level of education achieved to date
(5) employment
(6) income.

EXPERIENCES

In a similar manner to previous research, we first wished to ascertain frequency of privacy-related experiences among our respondents. This question was multiple choice, meaning that respondents were encouraged to select all experiences that applied to them. A free text option was also available, which is detailed in the results section.

(1) Unauthorised access to my online account.
(2) Credit card number stolen / bank fraud / unauthorised purchases from your account
(3) Being charged more for a product or service than other people
(4) Someone using spyware on me
(5) Someone impersonating me
(6) Private emails or messages posted online without my consent
(7) Public shaming online (people targeting me and shaming me for something I did or wrote, or for who I am)
(8) Private images or videos posted online without my consent
(9) Doxxing (private information posted online, such as my address)
(10) Other (Free Text)

CONCERN ABOUT PRIVACY

In the next section, respondents were presented with a series of 5-point Likert scales and were asked to what extent they agreed with each of the statements provided. The Likert scale points were labelled: Strongly Agree, Agree, Undecided/Neutral, Disagree, and Strongly Disagree.

7-Point scales were considered, but research has shown that the additional two options rarely add depth to findings (Dawes, 2008).

(1) My personal data could be used by others to steal money from me.
(2) My personal data could be used by others to impersonate me, which could affect my credit rating.
(3) My personal data could be used to badly affect my reputation.
(4) My personal data could be used by others to hurt me.
(5) My personal data could be used to unfairly discriminate against me.
(6) My personal data could be misused by governments.
(7) Not having privacy will lead me to change what I say online.
(8) Not having privacy will lead me to change my behaviour in negative ways.
(9) Not having privacy will lead other people to change their behaviour in negative ways.
(10) Privacy is a good in itself, above and beyond the consequences it may have.

RIGHTS VIOLATIONS

Respondents were asked to what extent they agreed with the statement "Violations to the right to privacy are one of the most important dangers that citizens face in the digital age". Responses were recorded as a 5-point Likert scale and were labelled: Strongly Agree, Agree, Undecided/Neutral, Disagree, and Strongly Disagree. Respondents were also asked "Do you think that privacy is a right?", which was a simple Yes or No answer.

PERSONAL DATA AND SENSITIVE INFORMATION

Prior to answering questions on personal data, respondents were presented with a definition of personal and sensitive data. This was included to address the ambiguity of the definition, and to account for different understandings between respondents. The text that was shown is below.

Take into consideration what personal data is and the sensitive information that can be inferred from it when you answer the following questions.
There are many things that count as personal data.

Examples include:

- Your name
- An identification number, such as your national insurance or passport number
- Your location data, such as your home address or mobile phone GPS data
- An online identifier, such as your IP or email address.

Personal data can be used to infer sensitive information, including:

- Racial or ethnic origin
- Political opinions
- Religious or philosophical beliefs
- Trade union membership
- Sex life or sexual orientation.

Following this information, respondents were presented with a series of 5-point Likert scales and were asked to what extent they agreed with each of the statements provided on companies uses of personal data. The Likert scale points were labelled: Strongly Agree, Agree, Undecided/Neutral, Disagree, and Strongly Disagree.

(1) Sell that data to third parties (insurance companies, governments, etc.) as part of their way of making money.
(2) Personalise ads to make them more relevant to individuals.
(3) Engage in price discrimination (charge different prices to different people for the same products and services).
(4) Research to develop new products.
(5) Investigate prospective employees (people who they want to hire).
(6) Investigate their current employees.
(7) Predict people's behaviour (e.g., where you are going to go, what you are going to buy, etc.).
(8) Try to influence what people will buy (try to get people to buy something they wouldn't otherwise buy).
(9) Try to influence how people will vote.

TRUST IN COMPANIES AND INSTITUTIONS

Respondents were shown a series of Variable Attribute Scales (VAS, 0-10) and asked to rate how much they trusted a selection of companies and institutions to protect privacy. The scales were rated from 0 – I don't trust them at all, to 10 – I trust them completely. The respondents moved sliders, and were not shown the number attributed to each company or institution.

The companies that respondents were asked to provide a rating of trust on were:

(1) Facebook
(2) Twitter
(3) Instagram
(4) Snapchat
(5) Google
(6) Amazon
(7) Apple

The institutions that respondents were asked to provide a rating of trust on were:

(1) My internet and telephone provider
(2) My bank
(3) My local neighbourhood shops
(4) My employer
(5) My government

Companies and institutions were grouped into two separate questions to avoid confusion.

GOVERNMENT DATA COLLECTION

Respondents were asked two questions on bulk collection of personal data by governments. The first question was: "Do you think it's okay for governments to bulk collect everyone's personal data?". The responses available were (1) Yes, there are some uses of this data that is necessary and acceptable, (2) No, Governments should not be about to collect everyone's data for any purpose, they should only be able to collect the data of criminal suspects, and (3) Prefer not to say.

The second question in this section asked respondents "Under what circumstances would you consider it acceptable for governments to collect everyone's data?". The answer was given in the form of multiple choice selection.

The options available were:

(1) Predict whether people will protest
(2) Predict how people will vote
(3) Try to influence how people will vote
(4) Make sure that people are paying their taxes
(5) Prevent petty (minor) crimes
(6) Prevent serious crimes
(7) Catch criminals of petty (minor) crimes
(8) Catch criminals of serious crimes

PRICE OF PRIVACY

The survey also examined respondents on the two most common payment models in regards to privacy online. The first model is *pay for access*. The question stated "It is known that most online platforms (e.g. Facebook, Google, and others) collect user personal data. For what amount (in USD) per month would you be willing to be paid to allow access to your personal data?".

The responses available were:

(1) The should pay me $ [Free text],
(2) Nothing, I'm not worried about online platforms having access to my personal data, and
(3) Nothing, privacy is a right and I don't think we need to pay for it.

The second question looked at the *pay for deletion* model. The questions stated: "It is known that most online platforms (e.g. Facebook, Google, and others) collect user personal data. What would you be willing to pay per month (in USD) to continuously delete all of your personal data from the parties that hold it?".

The responses available were:

(1) I would pay $ [Free text],
(2) Nothing. I'm not worried about online platforms holding my personal data, and
(3) Nothing. Privacy is a right and I don't think we should need to pay for it.

Respondents were limited to one response for each question. A comparison of the results of this measure and existing work by Winegar and Sunstein (2019) can be found in the last section

Ethical Procedure

Ethical approval for the study was granted by the University of Oxford in November 2019 (Ref: SSH_OII_CIA_19_065). Participants were provided with an information sheet and required to give written consent in the beginning of the survey in order to participate. They were also informed that they may withdraw at any time, and that questions were not mandatory, with "prefer not to say" options provided.

The confidentiality and anonymity of the subjects was guaranteed, and no personally identifiable data was collected. The remuneration offered to participants was 2.00 €. Such remuneration was calculated from the EU minimum living hourly wage, which is 10.03 €. This wage is also slightly above the average payment expected for such tasks on platforms such as Mechanical Turk, but is not high enough that it risks incentive affects and the validity of our data.

Sampling Frame

As previously outlined, respondents were collected through Mechanical Turk. We aimed to collect respondents from a range of nationalities, and split our sample between the United States and European countries to facilitate a regional comparison. The full demographics of the sample can be found in the results section.

NOTES ON MECHANICAL TURK

The survey was distributed through Amazon's crowd-sourcing Internet marketplace, Mechanical Turk (MTurk). MTurk is an increasingly prominent forum for digital social research, largely forming the basis of credibility in many online studies. Buhrmester, Kwang & Gosling (2011) evaluated the stability and quality of web-based data collection from samples drawn from Amazon's Mechanical Turk (MTurk).

Table 1: Breakdown of countries in sample

REGION	N	COUNTRIES	
EUROPE	630	Britain	303
		Germany	129
		Spain	84
		France	55
		Netherlands	23
		Italy	15
		Belgium	7
		Portugal	5
		Russia	5
		Czech Republic	2
		Ukraine	1
		Croatia	1
NORTH AMERICA	427	United States of America	427
SOUTH AMERICA	32	Mexico	16
		Colombia	8
		Brazil	5
		Venezuela	2
		Argentina	1
ASIA	18	Japan	10
		India	3
		Hong Kong	2
		Morocco	1
		China	1
		Singapore	1

QUESTION	ALTERATION
Please rate the degree to which you agree or disagree with each of the statements below. I am concerned about my privacy because.	Respondents were not shown Measure 6: My personal data could be misused by governments.
From 1 to 10, how much do you trust different institutions to protect your privacy?	Respondents were not shown Institution 6: My government.
Do you think it is okay for government agencies to bulk collect everyone's data?	Question not displayed
Under what circumstances would you consider it acceptable for governments to collect everyone's data?	Question not displayed

Table 2: Sensitive countries flow control measures.

The integrated compensation system, large sampling pool, ease of participant recruitment results in MTurk being an appealing platform for data collection in the social sciences (Casler, Bickel, & Hackett, 2013). However, concerns have been raised as to how MTurk compares with other samples, and the effects of task length and compensation/incentive amount.

Buhrmester, Kwang & Gosling (2011) administered around 500 personality tests to participants recruited through MTurk and a second large internet sample. The two tests were ad-ministered in two waves, three weeks apart (Buhrmester et al., 2011). By using the test-retest method, Buhrmester, Kwang & Gosling (2011) were able to conclude that stability of data collected through MTurk was very high, comparing favourably with correlations of traditional methods, resulting in a high level of reliability. A crucial factor in the reliability of a study is its stability over time, which is held to be high on MTurk.

Prior to using platforms such as MTurk, the predominant and most popular population from which research acquired samples was undergraduate students, of which the external validity as an unrepresentative sampling pool has been debated extensively (Buhrmester et al., 2011). Berinsky, Huber, & Lenz (2012) conducted an examination into the external and internal validity of MTurk as a promising source of subject recruitment.

In terms of internal validity with research conducted primarily on MTurk, the authors raise two concerns:

(1) "Do MTurk workers violate assignment by participating in experiments multiple times?" and

(2) "MTurk respondents may generally pay greater attention to experimental instruments and survey questions than do other subjects" (Berinsky et al., 2012, pp. 365-366).

Their study found that repeat survey taking was of minimal importance. Furthermore, sampling conducted through MTurk can be considered high quality due to demographic representativeness and high levels of diversity, largely negating concerns of external validity (Berinsky et al., 2012). In short, MTurk is a suitable platform to gather respondents.

NOTES ON COUNTRIES

Countries where asking questions on government data collection practises was considered sensitive information were shown an altered version of the survey, in order to comply with high ethical standards. Location was determined to be where the user identified their location at the start of the survey. China (n=1), Hong Kong (3), Russia (2) were deemed to be places in which questions about government data collection were too sensitive to ask.

Variables

As shown below, Table 3 provides a concise summary of each of the variables used in our analysis, as detailed in the survey design section.

Table 3: Summary of variables

VARIABLE	DESCRIPTION
Region	Geographic Region in which respondent is based. One of: Europe, South America, North America, Asia.
Age	Age brackets: 17 or younger, 18-20 years, 21-29 years, 30-39 years, 40-49 years, 50-59 years, 60 or older, Prefer not to say
Gender	Participant self-identified gender: Male, Female, Non Binary, Prefer to self-describe (free text)
Education	Highest level of education achieved: Less than high school degree, High school degree or equivalent (e.g., GED), Some college but no degree, Associate degree, Bachelor degree, Graduate degree (Masters/PhD), Prefer not to answer
Employment	Employment status: Full-time employed, Part-time employed, Not employed for pay (Unemployed), Caregiver (e.g., children, elderly) Homemaker, Full-time student, Part-time student, Other, Prefer not to say
Income	Income in USD: $0, $1 to $9 999, $10 000 to $24 999, $25 000 to 49 999, $50 000 to 74 999, $75 000 to 99 999, $100 000 to 149 999, $150 000 and greater
Experience	Privacy related experience, multiple choice.
Privacy Important	Likert scale (5-point) on the importance of privacy.
Pay for Access	How much should companies pay to access your personal data?
Pay to Delete	What would you be willing to pay to have your personal data deleted?
Privacy Measure	5-Point Likert scale. Strength of concerns regarding privacy. i.e. “My personal data could be used by others to steal money from me”.
Violations	5-point Likert scale. “Violations to the right to privacy are one of the most important dangers that citizens face in digital age”
Right	Single choice: Is privacy a right?
Companies Data Uses	5-Point Likert scale. Purposes for which companies can use personal data “Sell that data to third parties (insurance companies etc.) as part of their way of making money?”
Companies Protect	0-10 Variable Attribute Scale. To what extent are certain companies trusted to protect data
Institutions Protect	0-10 Variable Attribute Scale. To what extent are certain institutions trusted to protect data
Government Data: Collect	Select the statement that closest matches respondent’s own views. “Do you think it is okay for government agencies to bulk collect everyone’s personal data?”
Government Data: Circumstance	Under what circumstances is it okay for governments to collect everyone’s personal data. Multiple Choice. i.e.” To make sure people are paying their taxes”

Demographics

In the following section, the complete demographics of the sample are broken down into their respective categories and by region. The countries included in each region are detailed in the sampling frame section.

Figure 1: Gender of respondents

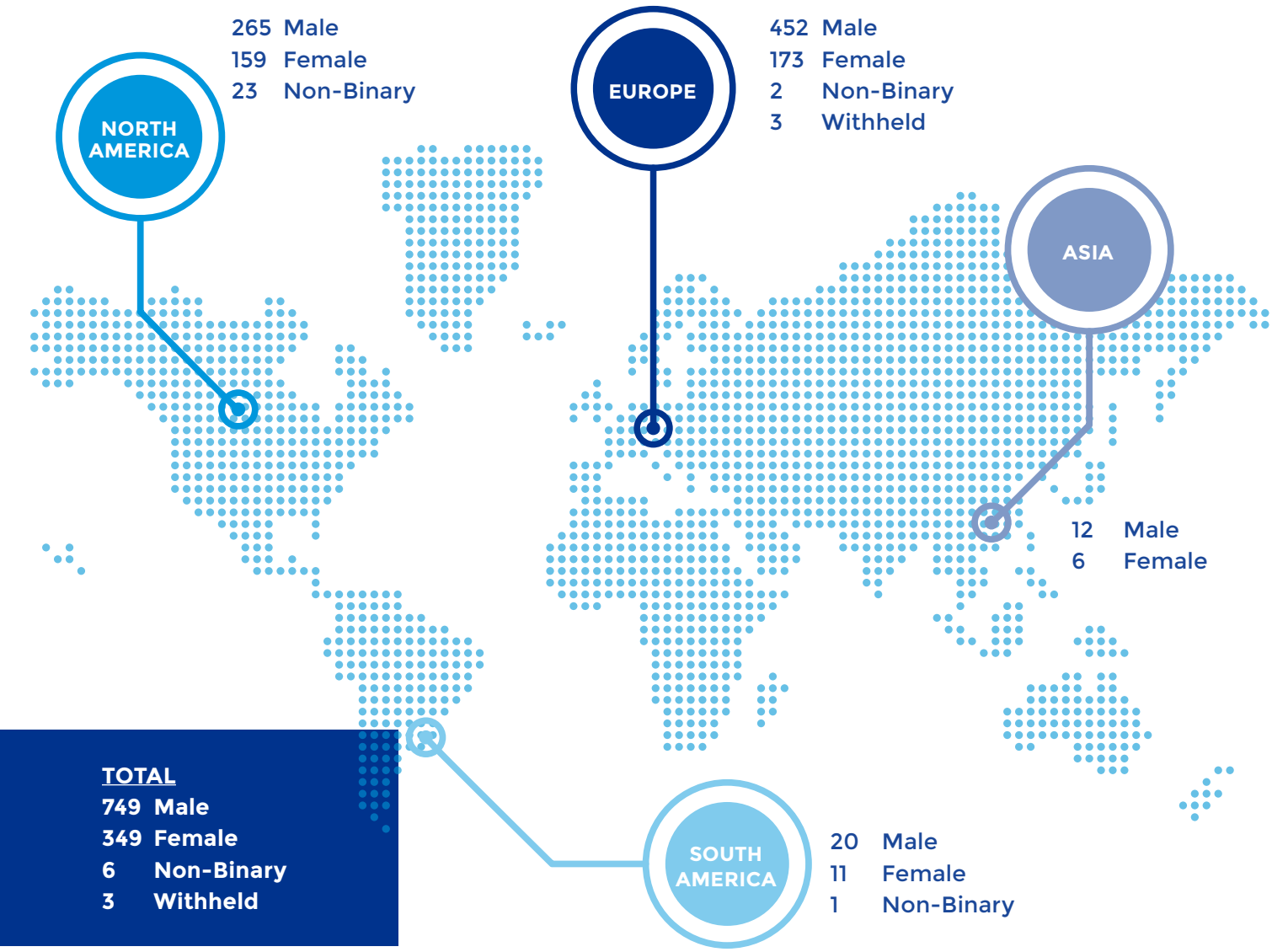


Table 4.1: Age of respondents

REGION	18-20 YEARS	21-29 YEARS	30-39 YEARS	40-49 YEARS	50-59 YEARS	60 YEARS OR OLDER
Europe	54	235	210	95	29	7
North America	2	131	180	57	35	22
South America	0	17	13	2	0	0
Asia	1	5	9	2	1	0
Total	**57**	**388**	**412**	**156**	**65**	**29**

Table 4.2: Highest education level achieved

REGION	LESS THAN HIGH SCHOOL	HIGH SCHOOL	SOME COLLEGE	ASSOCIATE DEGREE	BACHELOR'S DEGREE	GRADUATE DEGREE	WITHHELD
Europe	8	81	89	35	248	166	3
North America	2	51	64	38	222	50	0
South America	0	1	4	3	16	7	1
Asia	0	1	2	2	9	4	0
Total	**10**	**134**	**159**	**78**	**495**	**227**	**4**

Table 4.3: Current employment status

REGION	FT EMPLOYED	PT EMPLOYED	FT/PT STUDENT	HOMEMAKER/ CAREGIVER	UNEMPLOYED	OTHER	WITHHELD
Europe	357	120	87	18	24	21	3
North America	343	34	4	10	15	18	3
South America	16	9	3	3	1	0	0
Asia	9	1	4	0	1	3	0
Total	**725**	**164**	**98**	**31**	**41**	**42**	**6**

Table 4.4: Income level (USD)

REGION	0	1-9,999	10,000-24,999	25,000-49,000	50,000-74,999	75,000-99,999	100,000-149,999	150,000+	WITHHELD
Europe	27	129	147	179	66	40	7	1	34
North America	1	31	86	142	109	32	16	5	5
South America	0	5	11	10	3	1	0	0	2
Asia	2	3	5	6	2	0	0	0	0
Total	**30**	**168**	**249**	**337**	**180**	**73**	**23**	**6**	**41**

Experiences

This section breaks down the frequency of different negative experiences regarding privacy by region. There are two points to note here. Firstly, this is a self-reported measure, without a particular time limit, so depends on each respondent's ability to recall the event. The second is that the question was multiple choice, so the number of total recorded experiences will be larger than that of the total sample size. In total only 8% (n = 85) of respondents had no experience of their privacy being breached.

Across all regions, the average respondent had 1.86 bad experiences concerning privacy. The most common privacy breach was unauthorised access to an online account (n = 481), closely followed by the theft of credit card numbers, bank fraud, or unauthorised purchases from an account (450). In Other (Free Text), additional experiences listed included:

- Phishing/scam emails (Frequency: 3)
- Data Breach (4)
- Device Hack (6)
- Malware (1)
- False accusations (1)

Several respondents (3) referred specifically to being targeted using information that was leaked during the "Equifax Breach". In September 2017, Equifax (one of the three largest consumer credit reporting agencies) announced a cyber-security breach, which it claims to have occurred between mid-May and July 2017. The perpetrators accessed approximately 145.5 million U.S. Equifax consumer records, including their full names, Social Security numbers, birth dates, addresses, and driver license numbers. Equifax also confirmed at least 209,000 consumers' credit card credentials were stolen in the attack. On March 1, 2018, Equifax announced that 2.4 million additional U.S. customers were affected by the breach.

Europe followed the pattern of all regions collectively with the average European having 1.85 experiences of privacy breaches.

For North American respondents, the most common experience was the theft of credit card number, bank fraud, or purchases made from an online account (n = 217), followed by unauthorised access to an online account (169). The average North American had 1.90 experiences of privacy breaches.

We used an independent two-sample t-test (Welch) to assess if there was a significant difference between regions. We found that there is not a significant difference and the null hypothesis was accepted ($t = 1.08$ and $p = 0.29$. As $\alpha = 0.05$).

The results are broken down by gender in Table 5.1. We expected to see a significant difference in experiences of privacy breaches between men and women, as there is some literature claiming that women are more likely to be doxxed than men, or to have private images shared online without their consent (Mantilla, 2015; Phillips, 2015).

Table 5: Experiences regarding privacy: All Regions, Europe and North America

EXPERIENCE	ALL REGIONS		EUROPE		NORTH AMERICA	
	FREQUENCY	%	FREQUENCY	%	FREQUENCY	%
Unauthorised access to my online account	481	23%	290	25%	169	21%
Credit card number stolen / bank fraud / unauthorised purchases from your account	450	22%	210	18%	217	27%
Being charged more for a product or service than other people	210	10%	154	13%	48	6%
Someone using spyware on me	199	10%	114	10%	72	9%
Someone impersonating me	173	8%	93	8%	71	9%
Private emails or messages posted online without my consent	141	7%	77	7%	63	8%
Public shaming online (people targeting me and shaming me for something I did or wrote, or for who I am)	132	6%	68	6%	55	7%
Private images or videos posted online without my consent	129	6%	77	7%	60	7%
Doxxing (private information posted online, such as my address)	91	4%	46	4%	41	5%
Other (Free Text)	46	2%	27		17	2%
Total	**2052**		**1156**		**813**	

In the same manner as region, we ran a t-test between self-identified men and women looking for statistically significant differences in bad privacy experiences. Whilst we did collect data for non-binary and self-describing individuals, not enough data was collected in order to form a comparison. We found that $t = 2.49$ and $p = 0.02$. As p was below α, we reject the null hypothesis and concluded that there is a statistically significant difference between men and women's distribution of experiences across measures. We find that 97% of women have had one of the experiences listed, with an average of 1.02 experiences. For men, these figures are 96% and 1.02 experiences respectively. There was no statistically significant difference found for education, age, income, or employment.

Table 5.1 Experiences regarding privacy: All Regions, By Gender

EXPERIENCE	MALE		FEMALE	
	FREQUENCY	%	FREQUENCY	%
Unauthorised access to my online account	342	24%	136	22%
Credit card number stolen / bank fraud / unauthorised purchases from your account	285	20%	163	26%
Being charged more for a product or service than other people	156	11%	52	8%
Someone using spyware on me	151	11%	47	8%
Someone impersonating me	121	9%	52	8%
Private emails or messages posted online without my consent	103	7%	36	6%
Public shaming online (people targeting me and shaming me for something I did or wrote, or for who I am)	88	6%	43	7%
Private images or videos posted online without my consent	82	6%	46	7%
Doxxing (private information posted online, such as my address)	62	4%	29	5%
Other (Free Text)	31	2%	15	2%
Total	**1421**		**619**	

Importance of Privacy

The following tables reflect a breakdown of responses to the question "How important is privacy?". Results are broken down by region, gender, and education, as these are the characteristics thought to be most likely to have significant variations, according to the established literature.

In assessing if there are regional differences in concerns regarding privacy, the results of the t-test were $t = 1.08$ and $p = 0.04$. As p was below α, the null hypothesis was rejected, and we find that there are statistically significant differences between Europeans and North Americans in how they value privacy. By examining Figure 2.2, we can see that on average North Americans seem to place a higher value on privacy than Europeans.

HOW IMPORTANT IS PRIVACY?

Figure 2.1 Importance of Privacy (All respondents)

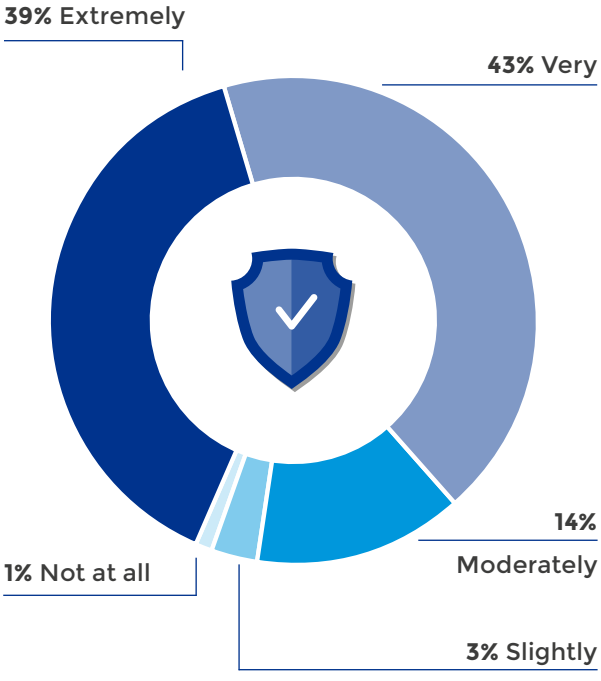
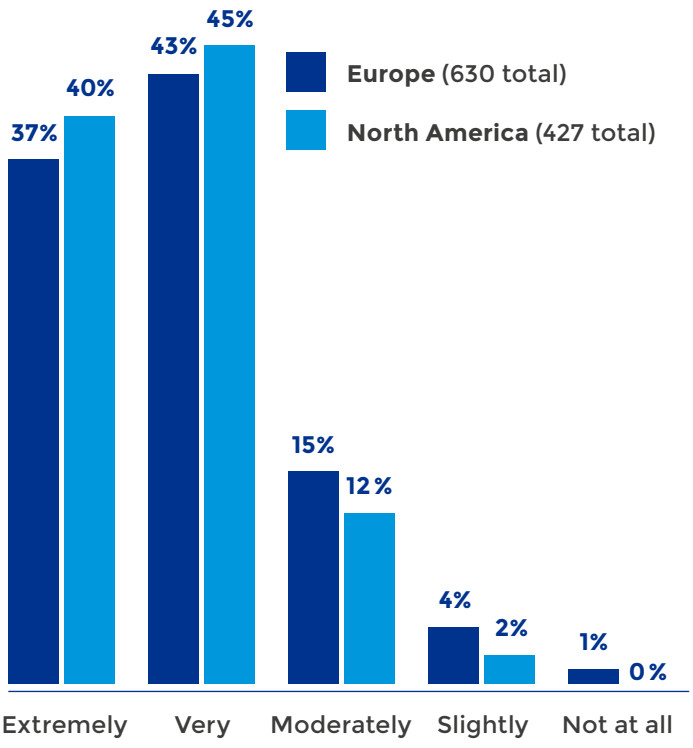


Figure 2.2 Importance of Privacy: By region



We found no significant difference between men and women ($p=0.39$) regarding how important they deem privacy. Both men and women, on average, think that privacy is very important. Similarly, a survey of the literature led us to examine if there is a significant relationship between highest level of education and the importance given to privacy. A test of association (Spearman) resulted in $p=0.46$, meaning that we accepted the null hypothesis that there is no statistically significant relationship between education level and belief in the importance of privacy.

Further testing revealed that there is no statistically significant association between belief that privacy is important and any of the demographic factors collected, including age of respondents.

Figure 2.3 Importance of Privacy: Gender

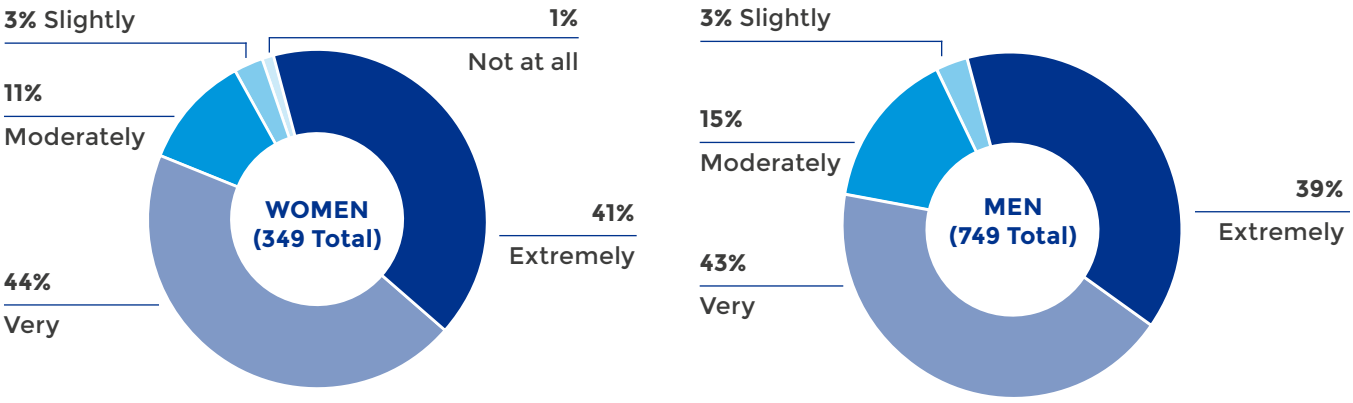


Table 6 Importance of Privacy: Education

EDUCATION	HOW IMPORTANT IS PRIVACY?				
	EXTREMELY	VERY	MODERATELY	SLIGHTLY	NOT AT ALL
Less than High School	4	3	2	0	1
High School	56	48	27	1	1
Some College	59	62	29	7	2
Associate Degree	34	32	8	4	0
Bachelor's Degree	192	224	61	16	1
Graduate Degree	87	107	27	5	1
Withheld	3	0	0	0	0

Concerns about Privacy

Concerns about privacy consisted of ten 5-point Likert scales. The full text of each of the scales is available in the Survey Design. Each statement here is an abbreviated version of the full statement displayed to participants.

Whilst there is some evidence of central tendency bias (an instance of social desirability[1] bias) here, overall the median and mode selection for every scale shows that in general respondents are very concerned with privacy across nearly all measures. However, change of behaviour in self averages at 3 – Undecided/Neutral; this result may be due to the statement being a bit unclear.

Following Welch's t-test, we conclude that there is no statistically significant difference between regions and reasons for concern about privacy. When we conduct the test with gender, however, we find that men and women differ significantly on every measure for concern about privacy.

Table 7.1 Privacy concern: All respondents and Regions

REASON FOR CONCERN ABOUT PRIVACY	ALL RESPONDENTS		EUROPE		NORTH AMERICA	
	MO (MODE)	MD (MEDIAN)	MO (MODE)	MD (MEDIAN)	MO (MODE)	MD (MEDIAN)
Theft of Money	5 (Strongly Agree)	4 (Agree)	5 (Strongly Agree)	4 (Agree)	4 (Agree)	4 (Agree)
Affect Credit Rating	4 (Agree)	4 (Agree)	4 (Agree)	4 (Agree)	4 (Agree)	4 (Agree)
Badly Affect Reputation	4 (Agree)	4 (Agree)	4 (Agree)	4 (Agree)	4 (Agree)	4 (Agree)
Used for Harm	4 (Agree)	4 (Agree)	4 (Agree)	4 (Agree)	4 (Agree)	4 (Agree)
Discrimination	4 (Agree)	4 (Agree)	4 (Agree)	4 (Agree)	4 (Agree)	4 (Agree)
Misused by Governments	4 (Agree)	4 (Agree)	4 (Agree)	4 (Agree)	4 (Agree)	4 (Agree)
Free Speech	4 (Agree)	4 (Agree)	4 (Agree)	4 (Agree)	4 (Agree)	4 (Agree)
Change Behaviour (Self)	3 (Undecided)	3 (Undecided)	3 (Undecided)	4 (Agree)	3 (Undecided)	3 (Undecided)
Change Behaviour (Others)	4 (Agree)	3 (Undecided)	4 (Agree)	3 (Undecided)	3 (Undecided)	3 (Undecided)
Good in Itself	4 (Agree)	4 (Agree)	4 (Agree)	4 (Agree)	4 (Agree)	4 (Agree)

1 The tendency of survey respondents to answer questions in a manner that will be viewed favourably by others. It can take the form of over emphasising or reporting 'good' behaviour, and underreporting undesirable (or 'bad') behaviour.

Looking at the mode values displayed on table 7.2, we can see that men are more often concerned with privacy as a protector against theft of money. We focus on mode here due to its ease of interpretation as the most common rating. Moreover, the results in table 7.2 also show that, compared to men, women disagree with the idea that a loss of privacy would lead them to change their own behaviour online in negative ways. This gendered difference could be explained by women largely seen as more social and communicative online, thus feeling that this behaviour is more constant and less likely to change (Papacharissi, 2010). No significant association was found with age, education, or income.

CONCERNS ABOUT PRIVACY BY AGE

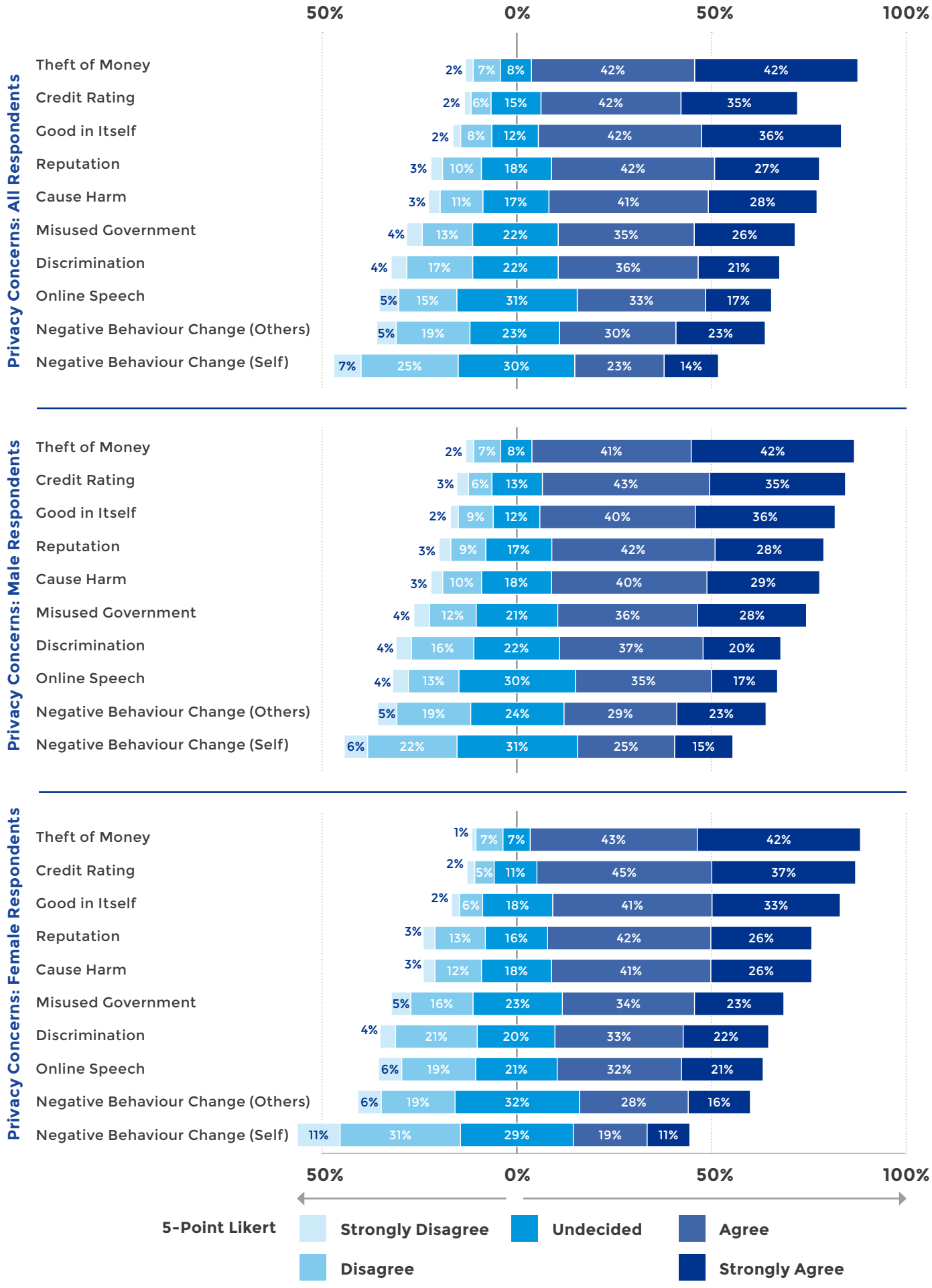
Analysing concerns about privacy by age, we first averaged the Privacy Concerns Likert scores into a singular Privacy Measure. We then used ANOVA (Analysis of Variance) to determine if privacy concerns vary significantly by age group. Overall, we found no significant variation between age groups.

Table 7.2 Privacy concern: Gender

REASON FOR CONCERN ABOUT PRIVACY	MEN		WOMEN		*t-stat*	*ß*
	MO (MODE)	MD (MEDIAN)	MO (MODE)	MD (MEDIAN)		
Theft of Money	5 (Strongly Agree)	4 (Agree)	4 (Agree)	4 (Agree)	-0.50***	-0.01
Affect Credit Rating	4 (Agree)	4 (Agree)	4 (Agree)	4 (Agree)	-1.76***	-0.06
Badly Affect Reputation	4 (Agree)	4 (Agree)	4 (Agree)	4 (Agree)	0.64***	0.2
Used for Harm	4 (Agree)	4 (Agree)	4 (Agree)	4 (Agree)	1.04***	0.05
Discrimination	4 (Agree)	4 (Agree)	4 (Agree)	4 (Agree)	0.40***	0.01
Misused by Governments	4 (Agree)	4 (Agree)	4 (Agree)	4 (Agree)	2.45***	0.07*
Free Speech	4 (Agree)	4 (Agree)	4 (Agree)	4 (Agree)	0.39***	0.01
Change Behaviour (Self)	3 (Undecided)	3 (Undecided)	2 (Disagree)	3 (Undecided)	4.60***	0.14***
Change Behaviour (Others)	4 (Agree)	4 (Agree)	3 (Undecided)	3 (Undecided)	2.44***	0.08**
Good in Itself	4 (Agree)	4 (Agree)	4 (Agree)	4 (Agree)	0.77***	0.02

Significance: p = >0.05 *, p = > 0.01 **, p = >0.001 *

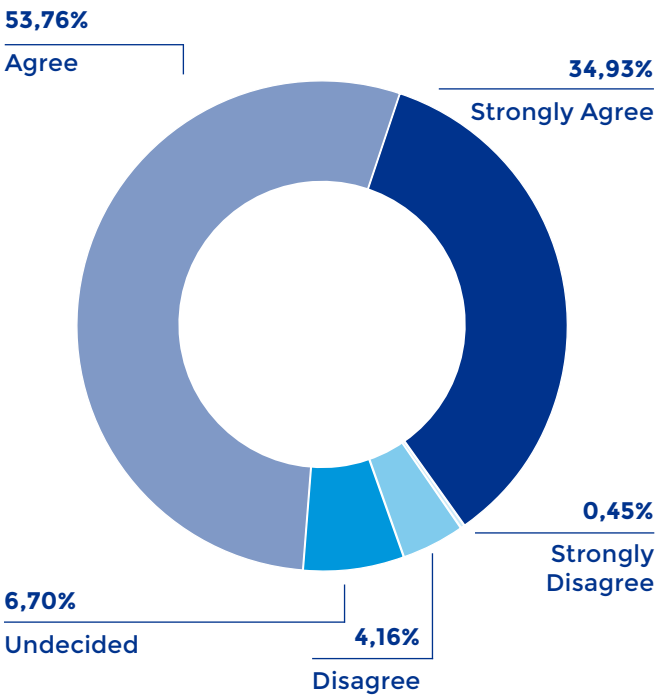
Figure 3 Privacy concerns



Violations to the Right to Privacy

This measure is a 5-point Likert scale, which states "violations to the right to privacy are one of the most important dangers that citizens face in the digital age". We found no statistically significant association between this measure and gender, region, age, income, education, or employment.

Figure 4 Violations to the right to privacy are one of the most important dangers that citizens face in the digital age



Is Privacy a Right?

This question simply asked respondents if they believe privacy was a right or not. As expected, a vast majority (97%) believed that privacy is a right. This pattern held across demographics.

Figure 5.2 Is privacy a right? By gender

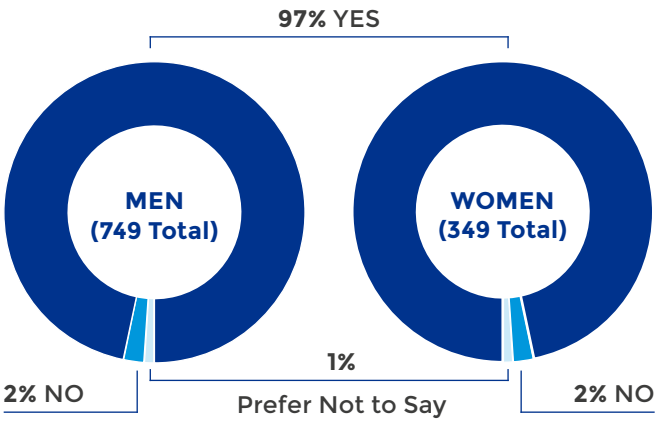


Figure 5.1 Is privacy a right? All respondents

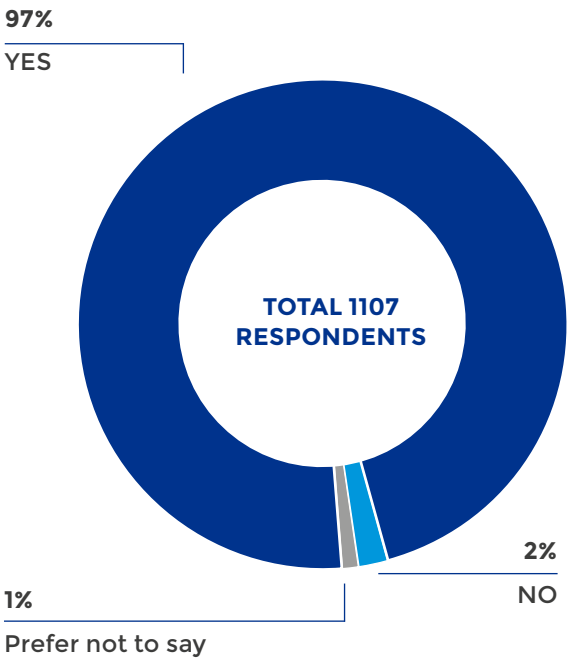
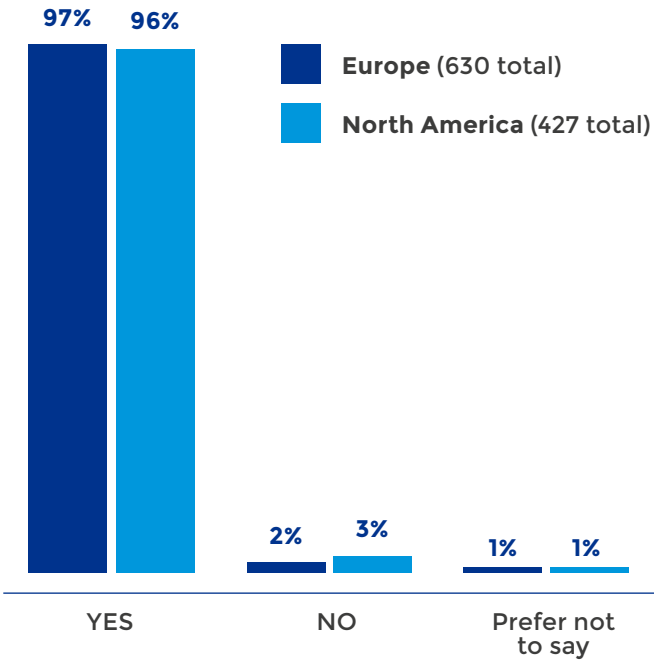


Figure 5.3 Is privacy a right? By region



The Use of Personal Data By Companies

Table 10.1 shows the average choices for all respondents on a collection of 5-point Likert scales regarding companies' use of personal data. The exact wording of these questions can be found in the survey design section.

The measures show some interesting dichotomies. For example, whilst it is generally agreed that it is acceptable to use personal data to personalise advertisements (43%), people tend to think it is unacceptable to use this data to influence purchases (57%). There are other interesting tensions, such as a 12% increase in dis-agreement with using personal data to investigate current employees (55% disagree and strongly disagree), compared to investigating prospective employees (43% disagree and strongly disagree).

Table 8.1 Companies' use of personal data: All respondents

USE OF PERSONAL DATA	MO (MODE)	MD (MEDIAN)
Sell to Third Parties	1 (Strongly Disagree)	2 (Disagree)
Personalise Ads	4 (Agree)	3 (Neutral/Undecided)
Price Discrimination	1 (Strongly Disagree)	1 (Strongly Disagree)
Develop New Products	4 (Agree)	4 (Agree)
Investigate Prospective Employees	4 (Agree)	3 (Neutral/Undecided)
Investigate Current Employees	1 (Strongly Disagree)	2 (Disagree)
Predict Behaviour	2 (Disagree)	2 (Disagree)
Influence Purchases	2 (Disagree)	2 (Disagree)
Influence Voting	1 (Strongly Disagree)	1 (Strongly Disagree)

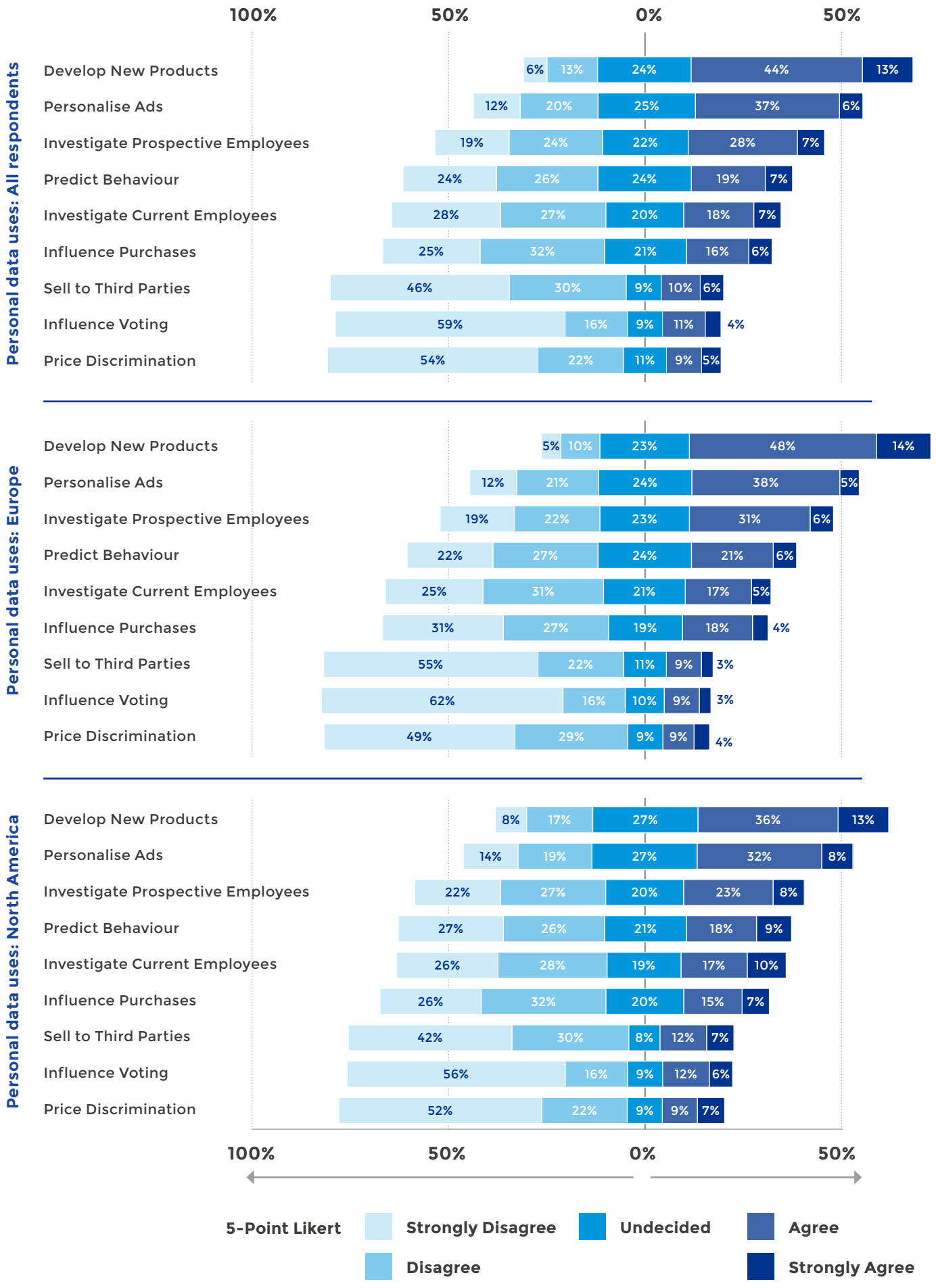
An independent t-test of Europe and North America (Table 8.2) shows that the continents differ significantly on each measure of companies' use of personal data. Below are divergent stacked bar chart of each of the measures, separated by region.

Table 8.2 Companies' use of personal data: By region

USE OF PERSONAL DATA	EUROPE		NORTH AMERICA		*t-stat*	*ß*
	MO (MODE)	MD (MEDIAN)	MO (MODE)	MD (MEDIAN)		
Sell to Third Parties	1 (Strongly Disagree)	2 (Disagree)	1 (Strongly Disagree)	2 (Disagree)	-3.08***	0.10**
Personalise Ads	4 (Agree)	3 (Undecided)	4 (Agree)	3 (Undecided)	0.43***	-0.01
Price Discrimination	1 (Strongly Disagree)	1 (Strongly Disagree)	1 (Strongly Disagree)	1 (Strongly Disagree)	-2.04***	0.08*
Develop New Products	4 (Agree)	4 (Agree)	4 (Agree)	3 (Undecided)	4.02***	-0.13***
Investigate Prospective Employees	4 (Agree)	3 (Undecided)	2 (Disagree)	3 (Undecided)	1.82***	-0.05
Investigate Current Employees	1 (Strongly Disagree)	2 (Disagree)	2 (Disagree)	2 (Disagree)	-2.36***	0.07*
Predict Behaviour	2 (Disagree)	3 (Undecided)	1 (Strongly Disagree)	2 (Disagree)	0.70***	-0.20
Influence Purchases	2 (Disagree)	2 (Disagree)	2 (Disagree)	2 (Disagree)	0.10***	-0.01
Influence Voting	1 (Strongly Disagree)	1 (Strongly Disagree)	1 (Strongly Disagree)	1 (Strongly Disagree)	-2.73***	0.09**

Significance: p = >0.05 *, p = > 0.01 **, p = >0.001 *

Figure 6 Personal data uses



Trust in Companies

The exact wording of these questions can be found in the survey design section. These questions consisted of Variable Attributed Scales (VAS) that were labeled from 0 – I don't trust them at all to 10 – I trust them completely.

Respondents were asked to rate each company. Std refers to the Standard Deviation, or a how much the members (95%) of a group differ from the mean value for the group. In table 9.1 we can see that the least trusted platform is Facebook, followed by Snapchat (which also had the most consensus), Instagram and Twitter. Google sits in about the middle of the group, with Amazon and Apple as the most trusted. The respondents surveyed rated Amazon almost twice as trustworthy, on average, than Facebook.

Figure 7.1 lists the statistically significant positive associations (Pearson's) between trust in different companies. The strongest association (0.78) is between Facebook and Instagram, which is likely due to the fact that Facebook is Instagram's parent company. All of the correlations are positive and relatively strong.

Figure 7.1 Heat Map of Trust in Companies: All Regions

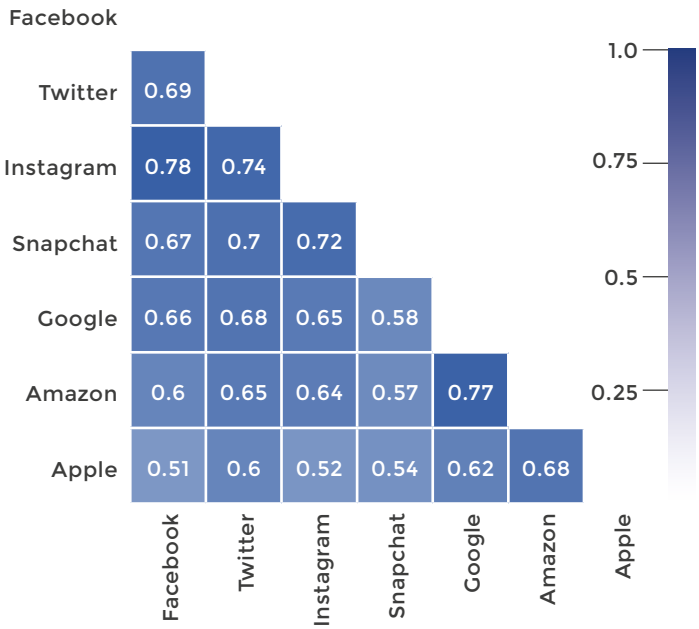


Table 9.1 Trust in Companies: All Respondents

COMPANY	RESPONSES	MEAN	std
Facebook	1101	2.75	3.01
Twitter	1101	3.84	2.81
Instagram	1102	3.39	2.92
Snapchat	1099	3.23	2.70
Google	1102	4.47	3.18
Amazon	1102	5.15	3.05
Apple	1099	5.03	3.11

Figure 7.2 Heat Map of Trust in Companies: Europe

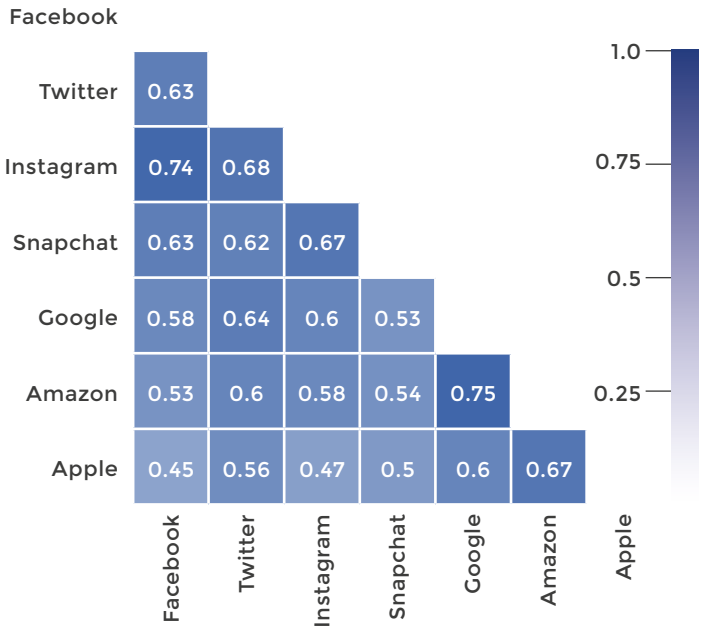


Table 9.2 Trust in Companies: Europe

COMPANY	RESPONSES	MEAN	std
Facebook	627	2.46	2.71
Twitter	627	3.84	2.64
Instagram	628	3.25	2.77
Snapchat	625	3.15	2.53
Google	628	4.36	3.06
Amazon	628	5.27	3.00
Apple	626	5.15	3.02

It is worth highlighting that even the companies with the highest rating for trust, still only reach the halfway mark of our scale for all respondents and regions. This finding is an indicative of low trust in companies. We also found statistically significant differences between Europe and North America, The HeatMaps and significant correlation coefficients are displayed in Figures 7.2 and 7.3. Table 9.2 shows that Europeans trust Facebook the least, and trust in Amazon and Apple is twice as high.

In North America, each company receives a lower average trust rating than in Europe, again with Facebook and Apple at opposite ends of the scale. For North America, not a single company passes over midpoint of the scale, which indicates low trust.

Figure 7.3 Heat Map of Trust in Companies: North America

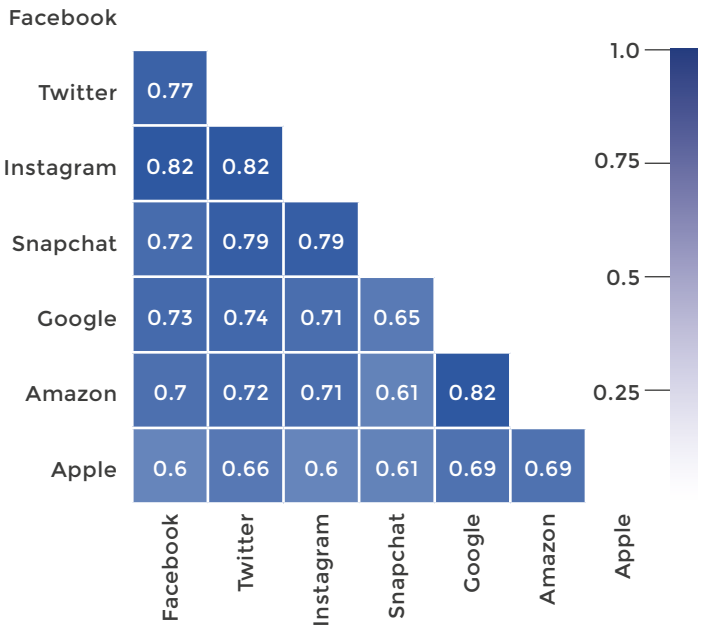


Table 9.3 Trust in Companies: North America

COMPANY	RESPONSES	MEAN	std
Facebook	424	3.15	3.36
Twitter	424	3.84	3.05
Instagram	424	3.55	3.10
Snapchat	424	3.29	2.94
Google	424	4.68	3.36
Amazon	424	4.93	3.15
Apple	423	4.73	3.25

Trust in Institutions

In the same manner as our question regarding companies, this question consisted of Variable Attributed Scales (VAS) that were labeled from 0 – I don't trust them at all to 10 – I trust them completely. The least trusted institution is the government, followed by Internet Service Providers (ISPs). Local shops are more trusted, relatively, followed by employers and banks, with healthcare providers as the most trusted.

On average, institutions rated as more trustworthy than companies. In comparing the two scales, we can see that in general Amazon and Apple are trusted more than the government. There was no significant difference (t-test) between men and women, or by region.

Figure 8 Heat Map of Trust in Institutions: All Regions

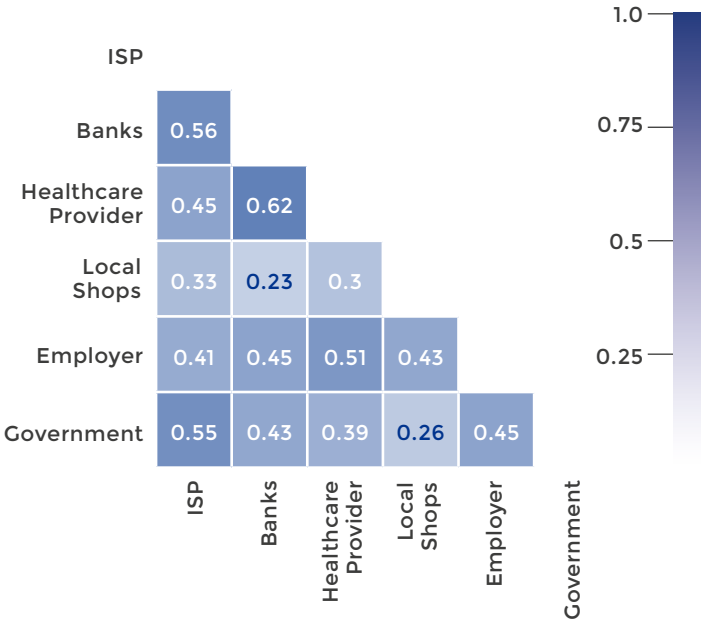


Table 10 Trust in Institutions: All Regions

INSTITUTION	COUNT	MEAN	std
ISP	1104	4.74	2.80
Banks	1103	6.69	2.57
Healthcare Provider	1103	6.71	2.47
Local Shops	1102	5.66	2.58
Employer	1102	6.36	2.49
Government	1098	4.50	2.95

Government Collection of Data

BULK COLLECTION OF PERSONAL DATA

In table 11.1, we can see that the majority of respondents believe that governments should not be allowed to collect everyone's personal data. Europeans are more likely than North Americans to find some uses of personal data collection acceptable.

CIRCUMSTANCES

This measure is a multiple choice question asking under what circumstances it is considered acceptable for the government to bulk collect everyone's personal data. The catching of criminals and prevention of serious crimes are generally considered to be the most acceptable use of personal data by governments. However, no use of personal data receives even 50% support from respondents. This is likely due to the low trust in governments that was outlined in Table 10.

Table 11.1 Bulk Collection of Personal Data: By Region

	ALL REGIONS		Europe		North America	
No, governments should not be allowed to collect everyone's data for any purpose, they should only be allowed to collect the data of criminal suspects.	603	55%	333	53%	247	58%
Yes, there are some uses of these data that are necessary and acceptable	450	41%	270	43%	162	38%
Prefer not to say	47	4%	24	4%	17	4%
TOTAL	**1100**		**627**		**426**	

Table 11.2 Governments' use of personal data: Region

	ALL REGIONS		Europe		North America	
Catch criminals of serious crimes	748	29%	471	28%	240	32%
Prevent serious crimes	625	24%	415	25%	175	23%
Make sure that people are paying their taxes	415	16%	279	17%	119	16%
Catch criminals of petty (minor) crimes	273	11%	193	12%	62	8%
Prevent petty (minor) crimes	226	9%	163	10%	45	6%
Predict how people will vote	123	5%	65	4%	53	7%
Predict whether people will protest	82	3%	44	3%	35	5%
Try to influence how people will vote	61	2%	28	2%	32	4%
TOTAL	**2553**		**1658**		**761**	

Price of Privacy

PAY FOR ACCESS

Participants were asked whether they would be willing to give companies access to their personal data in exchange for money. Table 12.1 shows that, on average, respondents are not willing to surrender their privacy for a fee[2]. Europeans are less likely than North Americans to surrender their personal data.

Table 12.2 shows the figure (USD) that respondents state they would have to be paid per month in order for companies to have access to their private data. In general, we can see that the figures here are very high. One way to interpret the very high values is to think that participants who entered very high values think privacy is so precious that they are not willing to consider it within a monetary framework. Even when we trim the mean, the amounts are reasonably high, in the hundreds, nearly a thousand, dollars.

Table 12.1 Pay for access to personal data: Region

	ALL REGIONS		Europe		North America	
I would surrender my personal data for a fee	473	43%	260	41%	192	45%
I would not surrender my personal data for a fee	631	58%	368	59%	234	55%
TOTAL	1104		628		427	

Table 12.2 Pay for access: Amount by region[3]

	COUNT	Mean	Median	MIN	25%	50%	75%	max	Trimmed mean (0.1)
Europe	258	9649	300	1	100	300	1000	500,000	958
America	189	4406	500	1	50	500	1000	100,000	768.07
TOTAL	468	7936	450	1	87.5	450	1000	500,000	954.40

2 Pay for Access is recoded into a binary variable for the sake of simplicity. The original form of the answer offered several justifications behind not considering payment for surrendering privacy or a fee for securing privacy.

3 Since the results of this table and the next are a fat-tailed distribution, standard deviation is not a meaningful way to calculate variance, which is the reason why we have omitted it in this table and the next (14.3, 14.4).

PAY FOR DELETION

This measure referred to respondents paying a monthly rate (USD) to have their personal data protected and deleted. The clear majority of participants would not pay to secure their personal data. Overall, North Americans were more likely to pay to secure their privacy than Europeans.

Unsurprisingly, the amounts that respondents gave for the figure they would be willing to pay to have their personal data is lower than the average figures for third parties to pay for access. Perhaps the most interesting figure here is that Americans are will to pay nearly twice as much (USD) as Europeans to have their personal data deleted. It is worth noting that the n for the t-test is relatively small, however, an a-priori analysis dictated that it was more than sufficient.

Table 12.3 Pay to delete personal data: Region

	ALL REGIONS		Europe		North America	
Nothing. Privacy is a right and I don't think we should need to pay for it	803	73%	477	76%	288	68%
I would pay a specified amount.	205	19%	109	17%	90	21%
Nothing. I'm not worried about online platforms holding my personal data	96	9%	42	7%	48	11%
TOTAL	**1104**		**628**		**426**	

Table 12.4 Pay to delete personal data: Amount by region

	COUNT	Mean	Median	MIN	25%	50%	75%	max	Trimmed mean (0.1)
Europe	108	278	10	0	5	10	46	15,000	20.25
America	88	1579	25	0	10	25	77.50	100,000	54.95
TOTAL	202	1579	14	0	5	14	50	15,000	29.54

Comparison with Previous Work

These measures were built from previous work by Winegar and Sunstein (2019) on the value placed on privacy among Americans. In their study the authors examine how much consumers value privacy by using metrics of willingness to pay (WTP) or willingness to accept (WTA) third party access. They find that the median participant is willing to pay $5 a month to maintain data privacy (to delete their data from all parties that have it), but demands a significant amount more ($80) to allow access to their personal data.[4]

Winegar and Sunstein (2019) argue that this disparity is indicative of a super-endowment effect, according to which individuals are much more likely to try and hold on to what they do have, rather than attempting to acquire it from other parties. In other words, individuals want to hold onto the privacy they do have (as a right), but are unwilling to make financial concessions to achieve more privacy.

Like Winegar and Sunstein (2019), our survey found that the figures submitted as an indicator of pay for access (WTA) were unlikely to be practical amounts that respondents were willing and able to pay. The entering of a value here is expressive, that is, a protest answer against the stipulation of the worth of privacy in the question. Our data also supported their finding that WTA is much higher than WTP for mean and median. Whilst Winegar and Sunstein (2019) found that 14% of respondents were not willing to pay anything for data privacy, our data shows that the vast majority of respondents were not willing, with only 19% prepared to pay for protecting their personal data. Winegar and Sunstein's (2019) result could be explained somewhat by the fact that their sample was completely American. Our data shows that when a comparison is drawn between Europe and North America, Europeans are more likely to consider their privacy a right and not be willing to pay for it. This difference is also noticeable in the average amount that North Americans and Europeans provide for WTA. Americans ask for 150% of what Europeans request for access to personal data.

4 While we also found that people ask more money to allow access to their personal data than what they are willing to pay to delete their data, the figures in our analysis are much higher than those of Winegar and Sunstein. The reason for this discrepancy is likely to be in how we handled outliers (i.e., people who responded with extremely high values). Winegar and Sunstein did the following to standardise responses: 'To determine a threshold at which to cut off responses, we took the 99th percentile of income in 2017, which IPUMS reported as roughly $300,000 per individual (IPUMS-USA). This equates to roughly $25,000 per month, and since it seems unlikely that participants would actually be willing and able to pay this amount (only 40 respondents of the 2,416 reported household income over $200,000 per year), we converted any amount of willingness to pay greater than $25,000 to $25,000.' They did the same with willingness to accept, for symmetry. Given that MTurk workers have an income lower than the national average, we had doubts about the meaningfulness of using IPUMS data. Instead, we trimmed the mean. It is likely that people who entered high numbers (even those much lower than $25,000) are not willing to pay that amount for privacy, and they might not realistically expect to receive that amount for giving up their privacy. Even then, however, the values entered are still a reflection of the value people place on privacy, albeit it might be a reflection that goes beyond their monetary framework.

EUROPEANS ARE MORE LIKELY THAN AMERICANS TO SEE PRIVACY AS A RIGHT.

Regarding paying to delete personal data (WTP), we found higher median amounts for all regions. Winegar and Sunstein (2019) found that the average (median) that respondents were willing to pay was $5 per month. However, our data shows that Europeans are willing to pay $10 and Americans are willing to pay $25. What is particularly interesting here is that Americans are willing to pay 250% the amount than Europeans to have their personal data deleted.

The overall finding here is that in general Americans are more likely to consider their personal data privacy for sale than Europeans. This results in a higher value placed on payment for access to their personal data, but also a higher value they are willing to pay in order to protect it. Our findings support Winegar and Sunstein's (2019) doubts that users are making trade-offs when exchanging personal data for free platform use, and that consumers lack clear information on what happens with their personal data on platforms. As speculated in their study, we find significant evidence of regional variation. In particular, Europeans are more likely than Americans to see privacy as a right.

BIBLIOGRAPHY

Berinsky, A. J., Huber, G. A., & Lenz, G. S. (2012). Evaluating online labor markets for experimental research: Amazon.com's mechanical turk. *Political Analysis,* 20(3), 351–368. https://doi.org/10.1093/pan/mpr057

Buhrmester, M., Kwang, T., & Gosling, S. D. (2011). Amazon's Mechanical Turk: A New Source of Inexpensive, Yet High-Quality, Data? *Perspectives on Psychological Science: A Journal of the Association for Psychological Science,* 6(1), 3–5. https://doi.org/10.1177/1745691610393980

Casler, K., Bickel, L., & Hackett, E. (2013). Separate but equal? A comparison of participants and data gathered via Amazon's MTurk, social media, and face-to-face behavioral testing. *Computers in Human Behavior,* 29(6), 2156–2160. https://doi.org/10.1016/j.chb.2013.05.009

Dawes, J. (2008). Do data characteristics change according to the number of scale points used? An experiment using 5-point, 7-point and 10-point scales. In *International Journal of Market Research* (Vol. 50).

Mantilla, K. (2015). *Gendertrolling: How Misogyny Went Viral.* Santa Barbara: ABC CLIO.

Papacharissi, Z. (2010). *A Networked Self.* https://doi.org/10.4324/9780203876527

Phillips, W. (2015). *This is why we can't have nice things: mapping the relationship between online trolling and mainstream culture* (Vol. 0777). https://doi.org/10.1080/14680777.2016.1213581

Winegar, A. G., & Sunstein, C. R. (2019). How Much Is Data Privacy Worth? A Preliminary Investigation. *Journal of Consumer Policy, 42*(3), 425–440. https://doi.org/10.1007/s10603-019-09419-y

AUTHORS:

Carissa Véliz, University of Oxford
Siân Brooke, University of Oxford

RECOMMENDED CITATION:

Brooke, Sian and Véliz, Carissa, *Data, Privacy & The Individual*. Madrid: Center for the Governance of Change, IE University, 2020