ERASMUS UNIVERSITY ROTTERDAM
Erasmus School of Economics
Bachelor Thesis Behavioural & Health Economics

The Perception of Legal Privacy Protection

Improving control over data-sharing decisions

Name student: Dylan van Oeveren
Student ID number: 434051

Supervisor: Francesco Capozza
Second assessor: Merel van Hulsen

Date final version: 31 January 2021

TABLE OF CONTENTS

1 SUMMARY

Privacy is a modern problem arising from the extensive use of the internet and the increasingly importance of data. This modern problem consists of people having concerns about their privacy and people not wanting to disclose personal information anymore because they are afraid for parties (accidentally) leaking their data. Prior research showed that lower privacy concerns is associated with higher perceived control over data (Hoadley, Xu, Lee, & Rosson, 2010). But, since other research found that the opposite is true, those findings are still open for debate (Brandimarte, Acquisti, & Loewenstein, 2012). This thesis tried to find an answer to the scientifically relevant question whether or not perceived control decreases privacy concerns by investigating the relation between perceived control and an individual's privacy concerns. This was done by employing an empirical study with a self-reported survey wherein one group gained knowledge on privacy laws, and one group not. By letting the respondents gain knowledge about privacy laws, the research tried to improve their perceived control. The results of the thesis showed that the group that gained knowledge on privacy legislation indeed reported a statistically significant higher perceived control than the group that had not gained knowledge. Therefore, it can be assumed that closing the knowledge gap (Hoofnagle & Urban, 2014) will improve an individual's perceived control. Furthermore, whether a respondent gained knowledge on privacy laws did also influence the effect of perceived control on privacy concerns. Perceived control did have a negative effect on privacy concerns, but the results were not significant. When the knowledge on legislation was included in the data analysis, it shows that this factor has an interaction effect with perceived control and privacy concerns, making privacy concerns decrease even more for the same increase in perceived control as the group that did not gain knowledge on privacy laws. In addition, the data analysis also showed age and prior privacy experience are important factors for privacy concerns to rise.

2 INTRODUCTION

Nowadays, most of our actions and interactions, such as e-mails, texts, dating, learning and shopping are internet based. This new dimension causes a public exposure of our personal information (Acquisti, Brandimarte, & Loewenstein, 2015). Despite firms, individuals and even the society as a whole (Tatonetti, Ye, Daneshjou, & Altman, 2012) are able to benefit from this exposure on the one hand (McAfee & Brynjolfsson, 2012), does the exposure on the other hand also come with a serious decrease in our privacy (Acquisti, Brandimarte, & Loewenstein, 2015).

Before moved on to the rest of the thesis, the term Privacy will be defined. This is difficult because privacy means different things to different people (Acquisti, Taylor, & Wagman, 2016). In this thesis, the following definition of privacy will be used: "Privacy is the control over and safeguarding of personal information" (Westin, 1967).

To warrant online privacy and protect data, the European Union (EU) adopted the General Data Protection Regulation (GDPR), which became directly applicable in 2018. The GDPR gives data controllers (i.e. website owners, e.g. firms) specific obligations in collecting and processing data. It also gives certain rights to data subjects (i.e. website visitors, e.g. customers), like the right of erasure which requires a data controller to delete all of the personal information that they hold about the requesting data subject (Lin, 2019). When the data controllers do not comply with the rules of the GDPR, they could be punished with fines. However, despite non-compliance with privacy laws can cause data controllers fines, online consumers still don't have full trust in firms and governments regarding to the warranty of their online privacy (Hoofnagle & Urban, 2014). Research even showed that two thirds of online customers would shop more online if they had more trust in retail sites with regard to their personal information (Acquisti A., 2004). This makes research that tries to improve individuals' perception of online privacy socially relevant because of the effect it could have on the economy.

A lot of research has been done in the field of improving individuals' perception of online privacy. A common found factor that influences privacy concerns, is the perceived control of individuals over their personal information. One finding is that lower perceived control over personal information is associated with higher privacy concerns (Hoadley, Xu, Lee, & Rosson, 2010). Other research found the same effect of perceived control on privacy concerns and reported that privacy concerns exist because individuals don't feel in control over the information they reveal (Acquisti & Gross, 2006). However, Brandimarte et al. found contradicting evidence that perceived control increases privacy concerns (Brandimarte, Acquisti, & Loewenstein, 2012). Hence, it is still unknown whether perceived control decreases or increases privacy concerns, which causes this matter to be still open for debate. By finding an answer to the question whether or not perceived control decreases privacy concerns, this thesis will contribute to the articles and will try to end the debate.

Perceived control over personal information is expected to have a positive correlation to privacy legislation. Reason for this expectation is that government legislation has a positive significant effect on individuals' perceived control over online privacy (Xu, 2007) (Armstrong & Culnan, 1999). However, online consumers are not fully aware of the privacy legislation. Hoofnagle and Urban (2014) call this the knowledge gap. When they asked individuals specific questions about privacy policies and basic privacy legislation, only 25% answered more than half of the questions correctly (Hoofnagle & Urban, 2014). With respect to these two experiments combined, it can be assumed that legislation has an effect on the perceived control over personal information, but not many people know about this legislation. Therefore, if those people gain more knowledge on legislation, the perceived control over their personal information is expected to improve. Furthermore, in 2016 Acquisti et al. found that consumers feel hindered in their control over their personal information because they are in a position of asymmetric information regarding the collecting and processing of their data (Acquisti, Taylor, & Wagman, 2016). Therefore, perceived control is also expected to improve if individuals know more about why and how their data is being tracked online. This thesis will examine the effect that knowledge on privacy laws could have on the perceived control of individuals, and how this knowledge relates to the individual's privacy concerns. Since the effect of knowledge on privacy laws has not been tested on perceived control or privacy concerns before, will this make the thesis scientifically relevant.

The most common critique on research that focuses on privacy holds that the surveys used in the research don't specify to a specific circumstance in the marketplace (Hoofnagle & Urban, 2014)**.** Therefore, this thesis will focus on the circumstance when individuals visit a website because almost everyone can imagine themselves being in the circumstance of browsing the internet and entering a webpage. By specifying this thesis to the circumstance of visiting a website, this research tries to improve prior research that is too general and abstract.

Summarizing all of the above, this thesis amounts the following. In this thesis the causal effect of knowledge on the GDPR on an individual's perceived control, and on privacy concerns will be estimated. This will be done by informing individuals about the privacy legislation (GDPR) with respect to the circumstance of visiting a website. The information provided in the experiment will close the knowledge gap and decrease asymmetric information, which in turn is likely to improve the individuals' perceived control. Additionally, the individuals are informed about the basics of websites' data tracking to understand the information about the legislation better. When the perceived control is improved, this thesis will test if this also decreased the individual's privacy concerns. To examine the different effects, this thesis will focus on the following research question:

How does knowledge on privacy legislation relate to an individual's privacy concerns?

The research question will be answered by rejecting or supporting the following hypotheses:

H1: *Knowledge on privacy legislation decreases an individual's privacy concerns*

H2: *Knowledge on privacy legislation improves individuals' perceived control over their online personal information*

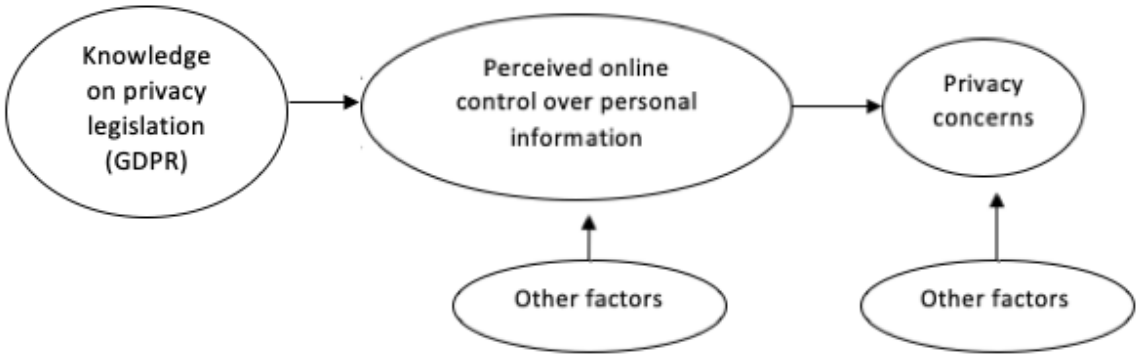
H3: *Improved individuals' perceived control over online personal information decreases their privacy concerns*

3 METHODOLOGY

3.1 Introduction

The expectations for this research, based on main findings in the literature about legislation on privacy, asymmetric-information in the online market place, perceived online control and privacy concerns, can be captured by the following arrow-diagram:

Figure 1
Arrow diagram of expected relations



The research question relates to the effect of 'knowledge on privacy legislation' on 'privacy concerns'. This will also be tested directly through the first hypothesis. However, based on main findings in the literature, the effect that knowledge on privacy legislation will have on privacy concerns is not direct but through (the variable) 'perceived online control over personal information'. This thesis will therefore also examine the effects that hypothesis 2 and 3 try to capture. Hypothesis 2 relates to the most left arrow. When hypothesis 2 can be supported with evidence, the perceived online control over personal information will improve when knowledge is improved. Hypothesis 3 will then find the (causal) effect of perceived online control on privacy concerns. As can be seen in the diagram, other factors could affect privacy concerns and/or perceived control. In the regression will be controlled for these factor to overcome the possible omitted variable bias and estimate the causal effect of perceived control. Based on findings in literature, knowledge on legislation is expected to have a positive correlation with perceived control. Perceived control could however have a positive or a negative correlation with privacy concerns based on contradicting evidence in literature.

The different expected relations will be examined through an experiment. This experiment consists of a survey that will be randomly distributed among different groups of people by issuing a hyperlink to the survey. The survey will be able to measure the different variables of interest and consists of questions and information about data-tracking, privacy laws, perceived control and privacy concerns. The survey is supposed to be distributed among respondents living in the European Union because the research contains legislation (GDPR) that is only applicable in the EU.

The survey randomly assigned the respondents to the treatment group or the control group. The treatment group had to complete a survey wherein they were given a quiz about privacy legislation (hereafter: "treatment-survey"), and the control group had to complete a survey containing a quiz about sports (hereafter: "control-survey"). The survey is included in the Appendix and will be discussed further in the section 'Survey'. The treatment-survey contains a quiz about privacy legislation in the EU with respect to the circumstance when visiting a website. Through this quiz there can be checked how much the respondents actually know about privacy laws that apply in the circumstance when an individual visits a website. After each question in the quiz, the right answer will be shown (regardless of whether the respondent gave the right or the wrong answer). The reason for giving the right information/answers after the question, is because people learn more after making a mistake in comparison with just reading information (Pressley, Tanenbaum, Mcdaniel, & Wood, 1990). After the quiz about legislation, the question was asked whether the respondent felt like he improved his knowledge on legislation that is applicable when he visits a website. Through this last question it can be checked whether the respondents felt like the quiz actually improved their knowledge. After the respondents completed their surveys, it can be checked whether the respondents that gained knowledge about privacy laws (treatment group) indeed had more perceived control or less privacy concerns than the respondents that hadn't gained knowledge about privacy laws (control group). How all variables were measured will be discussed in the section 'Survey'.

In order for the respondents to understand the legislation better and to gain more knowledge in a small amount of time, the respondents were also given information about some basic terms like "internet cookies" and "GDPR". Because this knowledge could also cause perceived control to grow, the results could be biased. Therefore, both the treatment and the control group were given the same information about these definitions. The effect that this knowledge could have on perceived control or privacy concerns, will therefore be filtered.

With the data derived from the completed surveys, all three hypotheses could be tested. The first hypothesis, – *Knowledge on privacy legislation decreases an individual's privacy concerns* – can be checked because the survey is able to measure the privacy concerns of the respondents, and is able to give one group knowledge on privacy laws, while the other is not given this information. Afterwards it could be tested whether the variable 'privacy concerns' is significantly higher or lower between the control group (the group that did not gain knowledge on privacy legislation) and the treatment group (the group that gained knowledge on privacy legislation).

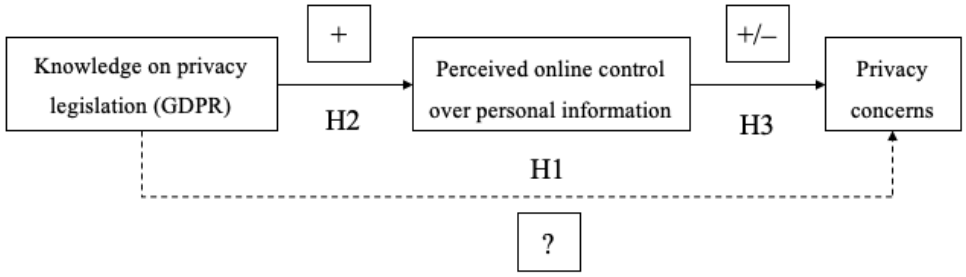
The second hypothesis, – *Knowledge on privacy legislation improves individuals' perceived control over their online personal information* – can again be checked by measuring the difference between groups, but this time, with perceived control as the outcome variable.

The third hypothesis – *Improved individuals' perceived control over their online personal information decreases privacy concerns* – can also be tested with the data derived from the completed surveys. Since perceived control is expected to improve after the respondents gained knowledge on privacy laws, the perceived control will differ among respondents. When this variable and the variable 'privacy concerns' are measured, it can be checked whether more 'control' caused more or less 'privacy concerns' with a regression.

Figure 2 summarizes the hypotheses in a model with expectations based on prior findings in the literature. The first hypothesis will hopefully contribute to the field of law and economics with determining a new effect on privacy concerns. This is the first test of the effect of knowledge on privacy legislation on privacy concerns. The second hypothesis is not expected to be rejected since closing the knowledge gap could contribute to the respondents feeling more in control. The third hypothesis is also not expected to be rejected since the main findings in research show a negative effect of perceived control on privacy concerns. However, since the effect is ambiguous, this is still unknown.

Figure 2

Relations between variables of interest



3.2 Survey

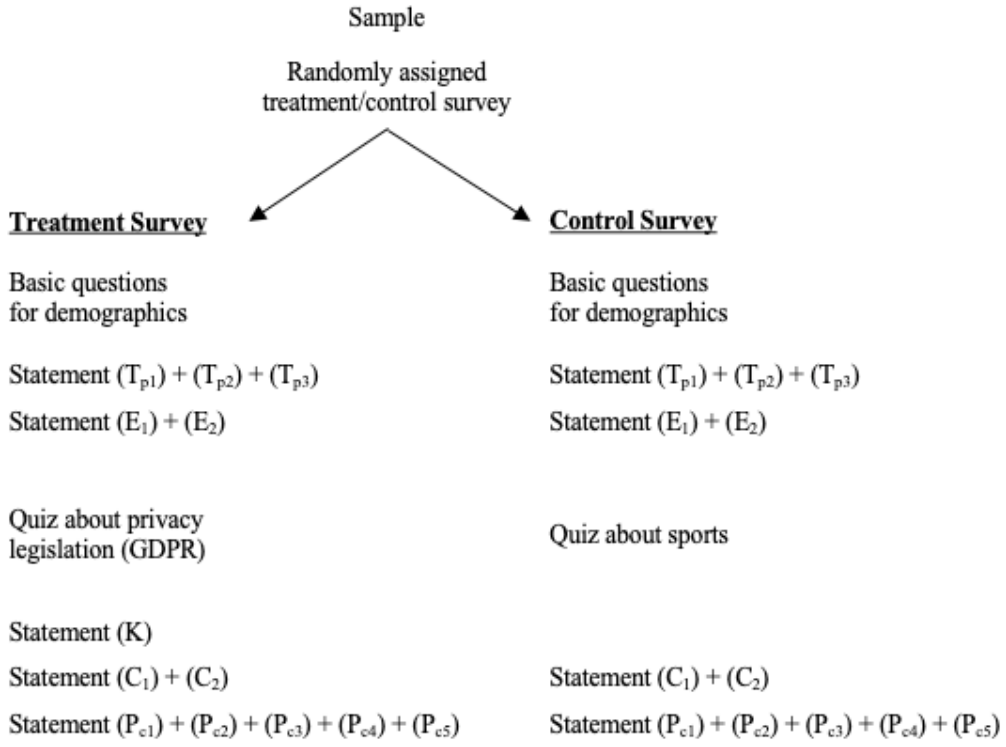
Both the treatment and the control group had to complete a survey. The survey is included in the Appendix (A). A respondent received either the treatment-survey or the control-survey. Both surveys were randomly distributed among the sample. One of the goals of the treatment-survey was to let respondents gain knowledge on privacy legislation and give basic information about data tracking to understand the information about the legislation better. In order to accomplish this, the survey contained a quiz with questions about legislation that is applicable in the circumstance when an individual visits a website. All these questions were based on the GDPR and rules that are derived from it.[1] Immediately after the respondent answered a question in the quiz, the survey provided the respondent with the correct

[1] Although parties are in violation when they don't comply with privacy laws, some legislation is still not applied on websites in practice. In the Appendix (B), one of the biggest Dutch newspapers is addressed (Het Algemeen Dagblad (AD)) about their violation of the GDPR. They replied with a statement of their Data Protection Office that they will review the privacy policy of their website.

answer to the question together with information about the subject of the question. The reason for giving the right information after the respondent had to answer the question, is because people learn more after being asked a question in comparison with just reading information (Pressley, Tanenbaum, Mcdaniel, & Wood, 1990). To optimize randomization, the survey of the control group also contained a quiz so that the respondents in the control group spend (about) the same amount of time on the survey as the treatment group. This quiz had on the other hand nothing to do with privacy or legislation but was about sports. The assumption that knowledge about sports had no effect on privacy concerns or perceived control, was therefore necessary. Figure 3 shows a scheme of the experimental flow of both surveys. The letters between brackets stand for the measured variables which are defined in table 1.

Figure 3

Experimental flow of both surveys



Although the quizzes between the treatment-survey and the control-survey differ, both surveys start with the same questions (as can be seen in figure 3). These questions ask basic information like the respondent's age, gender, marital status, highest education and employment status. This is done to check whether the sample is randomized. Another reason for incorporating basic questions in the survey is to use this information for control variables in the multiple regression. The last question of the beginning questions asked how much time the respondent spends on visiting websites on average per week.

After the basic questions, both surveys contain multiple statements that measure variables. In answering the statement-questions, the respondents could choose between: strongly disagree, disagree, somewhat disagree, neutral, somewhat agree, agree and strongly agree, which is based on the seven-point Likert scale. The seven-point Likert scale is used because this scale is more appropriate for electronic questionnaires that are unsupervised and is sensitive enough for an accurate evaluation while the answers remain compact (Finstad, 2010). The statements are based on renowned research that already tried to capture the effect of privacy legislation on perceived control and on privacy concerns. The data derived from the answers of the respondents to the statements, is able to measure the variables. Table 1 shows the different variables and the corresponding statements.

Table 1
Measuring the variables

Variable		**Measurement Items (Statements)**	**Reference Source**
Perceived Control (C)	C_1	I think I have control over my personal information when I visit a website.	Xu (2011)
	C_2	I think I have control over my personal information after providing my information to a website.	
Privacy Concerns (P_c)	P_{c1}	I am concerned that a website is collecting too much personal information about me.	Xu (2007)
	P_{c2}	I am concerned that a website may not take measures to prevent unauthorized access to my personal information.	
	P_{c3}	I am concerned that a website may keep my personal information in a non-accurate manner in their database.	
	P_{c4}	I am concerned that a website may share my personal information with other parties without getting my authorization.	
	P_{c5}	Overall, I feel unsafe about providing personal information to a website.	
Trust Propensity (T_p)	T_{p1}	Most people are honest.	McKnight et al. (2002)
	T_{p2}	I usually trust people until they give me a reason not to trust them.	
	T_{p3}	Most people keep their promises.	
Previous Privacy Experience (E)	E_1	In the past, I have had a negative experience regarding my online privacy.	Smith et al. (1995)
	E_2	I am aware of the privacy issues and practices in our society.	
Only for treatment group:			
Knowledge on Privacy Legislation (K)	K	After all this information, I feel like I gained knowledge about privacy legislation that is applicable to the circumstance of visiting a website.	

The answers to the statements of the respondents generate a score with a range from 1 to 7 because of the seven-point Likert scale. A score of 1 represents 'strongly disagree', a score of 2 represents 'disagree', the score 3 represents 'somewhat disagree', the score 4 represents 'neutral', the score 5 represents 'somewhat agree', a score of 6 means 'agree' and a score of 7 stands for 'strongly agree'. The variables Perceived Control, Privacy Concerns, Previous privacy experience and trust Propensity are measured by multiple statements. For these variables, the value of the variable for a respondent will be the mean of the scores of the corresponding statements. For instance, the amount of privacy concerns (P_c) a respondent has, is defined by the mean of the score of all the 5 corresponding statements together ($P_{c1} + P_{c2} + P_{c3} + P_{c4} + P_{c5}$).

The interest variable Perceived Control (C) will be measured by statements C_1 and C_2 , which are based on the statements used in the research of Y. Chang et al. (Chang, Wong, Libaque-Saenz, & Lee, 2018). These statements are in turn based on the psychological work of Reed et al. (Reed, Taylor, & Kemeny, 1993) and were able to measure perceived control. In this thesis, the statements of Y. Chang et al. are specified to the circumstance before and after releasing personal information to websites. The outcome variable P_c will be measured by five statements based on the research of Heng Xu (Xu, 2007), which in turn is based on the work of Smith et al. (1996). Again, the statements are specified to the circumstance of when an individual visits a website and were originally formulated as questions.

In most randomization experiments, the dependent variable is measured before and after the treatment to estimate the causal effect of the treatment. In this thesis the dependent variables will be measured only at the end of the survey. The reason for this is that the dependent variables are measured through questions/statements. If the dependent variables would have been asked before and after the treatment, the respondents could have had the feeling that they were being surveyed about those questions/statements. Hence, the effect of the treatment on control or privacy concerns could be biased. With this in mind other variables that are related with the dependent variables, were measured before the treatment. In this way, it could be checked through these measured variables before the treatment, whether the control group and the treatment group are comparable to each other. I.e. the treatment and the control group did not significantly differ from each other. This will in turn imply that the randomization worked without having measured the dependent variables before the treatment took place. The variables that have been measured before the treatment, and are related to privacy concerns and to perceived control are 'previous privacy experience' and 'trust propensity'. The variable 'previous privacy experience' (E) was measured with two questions based on the work of Smith et al. (1996). The other variable 'Trust Propensity' (T_p) was measured with three questions adapted from McKnight et al. (2002). Prior research found that these two variables could impact perceived control and privacy concerns. Therefore, these variables could also be included in the regression as control variables to downsize the standard error or to isolate the causal effect.

Only the treatment-survey contained the statement “After all this information, I feel like I gained knowledge about privacy legislation that is applicable to the circumstance of visiting a website”. This statement is incorporated in the survey to check whether the quiz about privacy legislation indeed caused respondents to gain more knowledge on privacy legislation.

The survey also asked whether the respondent lives in the European Union. This was relevant to ask since the GDPR is only applicable to European websites. When visiting websites from outside the European Union, most of the rules that were mentioned in the privacy-quiz do not apply. Respondents that live outside the EU visit more websites that are subjected to other rules than the GDPR than respondents living in the EU. Making the data of respondents living outside the EU less useful for the research in this thesis.

4 DATA

4.1 Data collection

To gather the data needed for a sufficient research, the survey has been randomly distributed among different groups of people by issuing a hyperlink to the survey. The respondents that clicked on the hyperlink were randomly assigned to either the treatment group or the control group. The treatment group had to complete the treatment-survey, and the control group had to complete the control-survey. The groups of people to which the hyperlink was distributed consisted of friends, family and fellow students. Some people of these groups distributed the survey subsequently in their social media or contacts causing the survey to be distributed multiple times across multiple groups. The survey was also distributed among websites that work as a platform where people can exchange surveys.

4.2 Sample

This section will discuss the original dataset and the reason behind adjusting the dataset to improve the internal validity. 208 respondents in total have started the survey from whom 167 finished it. The first adjustment that had been made to the sample is deleting the data of the respondents that have not finished the survey. The data of the respondents that haven't finished the survey is not useable because the final statements – wherein privacy concerns and perceived control is measured – were not answered. The second adjustment that had been made to the sample, is deleting the data of the three respondents that did not live in the European Union. The reason behind deleting this information is the following. The GDPR only applies to European websites because the GDPR is European legislation that only binds member states of the European Union. If the respondent lives in Europe there could still be a chance that the websites that the respondent visits are not bound by the GDPR because not every European country is a member state of the European Union. Hence, the survey asked the question if the respondent lives in the *European Union*, and not the question if the respondent lives in *Europe*.

The next adjustment of the original dataset is the deletion of the respondents that have completed the survey in a time that it would be impossible to have read all the given information. It is important that the respondents (in the treatment group) have read everything in the survey because they were supposed to gain knowledge on privacy legislation this way. If the respondents did not read the information about the legislation, it could be assumed that those respondents did not gain as much as knowledge as they could have gained when they have read everything. Their data will therefore not contribute to the internal validity. To remove the data of respondents who finished the survey 'too fast' is however not easily done because the plausible time a respondent would minimally need to complete the survey, is hard to detect. In order to check which respondent finished the survey too fast, the total time a respondent took to complete the quiz was measured. With this data it can be detected easier which respondents completed the quiz too fast. In the Appendix (C) two boxplots are shown. These boxplots show per group (treatment

or control group) how much time the respondents took to complete the quiz. Remarkably, there were no minimum outliers, only maximum outliers. The maximum outliers will however not be deleted because the more time a respondent spend on the quiz, does not imply that these respondents did not read everything. Moreover, they could be slow-readers or have dyslexia and took more time to still read everything. Hence, their data is useful. Another thing the boxplots show is that the time both groups spend on their quizzes was almost the same. The Privacy legislation quiz took slightly more time than the quiz about sports. This is also reflected in the means of the variable duration for both groups, 419 seconds for the treatment group versus 413 seconds for the control group.

After deleting the three people from outside the European Union, 164 respondents of the sample remain in the dataset. No data of respondents is removed because of the time they spend on the quiz. In total, 77 respondents have been assigned to the treatment group and 87 to the control group. A small sample size makes it unlikely that the randomization, such that the level of other possible influential variables is unequal among all groups, was properly done.

4.3 Descriptive statistics of the sample

Besides the two main key outcome variables perceived control and privacy concerns of the survey, other (background) variables were measured to check whether there were any significant differences between the treatment and control group before the treatment. Table 2 shows the demographics of the respondents. Due to the small sample size, the sample contained more females (N=101, 61.6%) than males (N= 63, 38.4%). If the sample was larger, this would probably go to a more equal distribution since gender is also equally distributed in the population. The youngest respondent was 18 years old while the oldest was 80 years old. Most of the respondents were 20 to 30 years old (Appendix C). The distribution in the variables education and average time visiting websites per week can also be seen in the Appendix (C).

4.4 Other relevant facts about the data

In the introduction of this thesis the 'Knowledge Gap' was mentioned in order to explain why information about legislation and data-tracking could affect perceived control and/or privacy concerns. The Knowledge Gap is actually also noticeable in this research. The treatment-survey contained a quiz wherein the respondent had to answer whether or not some statements were true. One statement stated that cookies were able to track the data of the respondents when they enter an online website and did not give permission for cookies. 56% Of the respondents answered wrong that the cookies were already able to track their data before giving permission. Another example of a statement wherein the knowledge gap was visible is: "If a website has a privacy policy, it means that the site cannot sell your address and

purchase history to other companies or the government". Only 48% of the respondents answered correctly that the statement was false. A privacy policy is a statement that discloses all the ways a party gathers, uses, discloses and manages data. When accepted, the party is allowed to do what their policy holds. The last but not least example of the knowledge gap is the percentage of people who know what the GDPR holds. In total 109 of the 164 respondents (66.4%) reported that they don't know what the GDPR holds. These percentages show that the knowledge gap is still relevant until now.

Table 2
Demographics of the respondents

Variable	Measure	Frequency	Percentage	Range
Duration of quiz				153-2668
Gender	Male	63	38.4	
	Female	101	61.6	
Age				18-80
Education	Elementary	1	0.6%	
	Secondary	19	11.6%	
	Diploma	51	31.1%	
	Bachelor	68	41.5%	
	Master	22	13.4%	
	Doctoral	0	0%	
	Other	3	1.8%	
Marital status	Single	63	38.4%	
	Married	63	38.4%	
	Divorced/separated	5	3%	
	Other	33	20.1%	
Employment status	Full-time	57	34.8%	
	Part-time	49	29.9%	
	Self-employed	6	3.7%	
	Retired	4	2.4%	
	Full-time student	35	21.3%	
	Unemployed	6	3.7%	
	Other	7	4.3%	
Visit times of websites per week	0-2 hours	50	30.5%	
	3-5 hours	65	39.6%	
	6-9 hours	26	15.9%	
	10-13 hours	8	4.9%	
	14+ hours	15	9.1%	
Trust propensity				2-7
Previous privacy experience				1-6.5

5 RESULTS

5.1 Randomization

In order to analyze the data and to be able to neglect other variables than the variable of the treatment effect, it will be needed to check whether the randomization worked properly. The following table shows descriptive statistics of the treatment and control group.

Table 3
Means of variables

Background Variable	Treatment	N	Mean	Std. Deviation	Std. Error Mean
Duration of Quiz	No treatment	87	413.49	301.146	32.286
	Treatment	77	418.56	183.203	20.878
Gender	No treatment	87	1.62	.488	.052
	Treatment	77	1.61	.491	.056
Age	No treatment	87	36.91	14.678	1.574
	Treatment	77	37.06	16.424	1.872
Education	No treatment	87	3.66	.986	.106
	Treatment	77	3.60	1.016	.116
Marital status	No treatment	87	2.20	1.445	.155
	Treatment	77	2.31	1.515	.173
Employment status	No treatment	87	2.55	1.790	.192
	Treatment	77	2.95	1.966	.224
Visit times of websites per week	No treatment	87	2.09	1.187	.127
	Treatment	77	2.38	1.203	.137
Trust Propensity	No treatment	87	4.7357	.93631	.10038
	Treatment	77	4.5675	1.04307	.11887
Previous Privacy Experience	No treatment	87	4.2126	1.10662	.11864
	Treatment	77	4.2792	1.05286	.11998

As can be seen in this table, the means of the different background variables did not differ much between the treatment and the control group, which could be an indicator for a randomized sample. However, this does not immediately imply that the sample is indeed randomized. To formally check whether the randomization worked properly, different tests were executed to test whether the two groups differed significantly on the different background characteristics.

First of all, the Mann-Whitney U test is used to test for differences between groups for the ordinal and continuous variables. These are the variables ‘Duration of quiz’, ‘Age’, ‘visit times of websites per week’, ‘trust propensity’, ‘previous privacy experience’ and ‘privacy concerns’. This test is chosen instead of the t-test since the data of those variables does not resemble a normal distribution (Appendix D). The Shapiro-Wilk and the Kolmogorov-Smirnov tests, which are used on small and large samples respectively, also confirmed that the ordinal and continuous background variables are not normally distributed. The H0 of the Shapiro-Wilk and the Kolmogorov-Smirnov test is that the data is normally

distributed. As the following table shows is every variable significant. Therefore, the H0 can be rejected and it can be concluded that the data of these variables is not normally distributed.

Table 4
Tests of Normality

	Kolmogorov-Smirnov[a]			Shapiro-Wilk		
	Statistic	**Df**	**Sig.**	**Statistic**	**Df**	**Sig.**
Duration	.175	164	.000***	.609	164	.000***
Age	.210	164	.000***	.868	164	.000***
Visit times of websites per week	.276	164	.000***	.822	164	.000***
Trust Propensity	.093	164	.001***	.974	164	.004***
Previous Privacy Experience	.169	164	.000***	.963	164	.000***

a. Lilliefors Significance Correction

***p < 0.01, **p < 0.05, *p < 0.10

Another reason for using the Mann-Whitney U test is that the data of the two groups is unpaired because the control-survey and the treatment-survey were randomly distributed. Hence, there is no relation between the respondents in the treatment group and the respondents in the control group. The null hypothesis of the Mann-Whitney U test is that both distributions are equal, which implies that an observation has the same probability of coming from the different samples. Table 5 shows the output of the Mann-Whitney U test.

Table 5
Output of Mann-Whitney U test

Variable	**Treatment**	**N**	**Mean rank**	**Sum of ranks**	**Mann-Whitney U test**	**Asymp. Sig. (2-tailed)**
Duration of quiz	No treatment	87	78.08	6793.00		
	Treatment	77	87.49	6737.00	2965.00	0.205
	Total	164				
Age	No treatment	87	81.89	7124.50		
	Treatment	77	83.19	6405.50	3296.50	0.861
	Total	164				
Visit times of websites per week	No treatment	87	76.62	6666.00		
	Treatment	77	89.14	6864.00	2838.00	0.076*
	Total	164				
Trust Propensity	No treatment	87	85.37	7427.00		
	Treatment	77	79.26	6103.00	3100.00	0.408
	Total	164				
Previous Privacy Experience	No treatment	87	80.82	7031.00		
	Treatment	77	84.40	6499.00	3203.00	0.624
	Total	164				

***p < 0.01, **p < 0.05, *p < 0.10

As can be seen in Table 5, none of the different variables are significant (with a 5% significance level). Therefore, H0 cannot be rejected. This means that the treatment and the control group do not significantly differ from each other regarding to the specific variables. Hence, it can be accepted with evidence that both groups are properly randomized regarding those variables. Nevertheless, the variable 'visit times of websites per week' does need attention. This variable is significant on a 10% significance level which implicitly means that the respondents in the treatment group are more likely to spend more time per week on visiting websites per week, than the respondents in the control group. This difference in groups could bias the results because people who spend more time visiting websites per week, ought to feel more in control over their actions online than people that spend less time visiting websites. To dig deeper in this possible bias, the regression could be controlled for this variable.

Secondly, differences between the treatment and the control group regarding to the remaining variables are tested with the Chi-square test or the Fisher's exact test. The reason for choosing those tests is that the variables are nominal (and the variable Gender is binary). Comparing two groups with a variable that is not continuous or ordinal, is not possible with a t-test or a Mann-Whitney U test, but is possible with the Chi-square and the Fisher's exact test. The McNemar test is also not possible since the treatment and the control group are unpaired. The choice between the Chi-square test and the Fisher's exact test is made based on whether the Chi-square test is allowed to be used. When the Chi-square test shows that more than 20% of the cells have an expected count of less than 5 (or when any expected frequency is less than 1), the Chi-square test is not allowed to be used and the Fisher's exact test will be used. Only the data of the variable Gender does have less than 20% of the cells with an expected count of less than 5 (Appendix E). Therefore, the differences between the treatment and control group were analyzed with a Chi-square test for the variable Gender, and measured with the Fisher's exact test for the variables Education, Marital status and Employment status.

The H0 of the Chi-square test is that there is no statistical relation between two variables. I.e. the variable Gender does not have a relation with the control group or the treatment group. As table 6 shows, the Chi-square test is not significant. Meaning, the H0 cannot be rejected and it can be accepted with evidence that there is no relation between the group a respondent is in and their gender. This positively contributes to the assumption that the two groups were properly randomized.

Table 6
Chi-Square Test of the variable Gender

	Value	**df**	**Asymptotic Significance (2-sided)**	**Exact Sig. (2-sided)**	**Exact Sig. (1-sided)**
Pearson Chi-Square	.018[a]	1	.892		
Fisher's Exact Test				1.000	.510

a. 0 cells (0%) have expected count less than 5. The minimum expected count is 29,58.

The H0 of the Fisher's Exact Test is like the Chi-square test, that the two classifications are not different from each other. I.e. in the case of the remaining variables; that employment status is not different between the treatment and control group, that marital status is not different between groups and that education is not different between groups. As table 7 shows none of the Fisher's Exact Tests came out with significant results. Therefore, the null hypothesis cannot be rejected and there can be concluded with evidence that the treatment group did not differ from the control group with regard to the variables Education, Marital status and Employment status.

Table 7

Fisher's Exact Tests

Variable	**Fisher's Exact Test Asymptotic Significance (2-sided)**
Education	0.919
Marital status	0.619
Employment status	0.190

***P < 0.01, **P < 0.05, *P < 0.10

Altogether, the Mann-Whitney U tests, the Chi-square test and the Fisher's Exact tests showed that the respondents in the treatment group did not differ from those in the control group with regard to the background variables before the treatment. Furthermore, because the respondents in both groups also did not differ from each other with regard to trust propensity and previous privacy experience, it could be concluded that the samples in both groups are properly randomized. In turn could therefore be assumed that the outcome variables, perceived control and privacy concerns, do not differ between the two groups before the respondents started the survey. Hence, the treatment effect will be causal.

5.2 Hypothesis 1

The first hypothesis – *Knowledge on privacy legislation decreases an individual's privacy concerns* – will be analyzed, as described in the methodology, with a with-and-without comparison. The first group is the treatment group that gained knowledge about privacy legislation. The other group, the control group, is the group without having gained knowledge. Before the data for the first hypothesis can be analyzed, it must be established that knowledge on privacy legislation has actually been gained through the quiz about legislation. To determine this, the following statement was incorporated in the treatment-survey: "After all this information, I feel like I gained knowledge about privacy legislation that is applicable to the circumstance of visiting a website". This statement could be answered on a seven-point Likert Scale. Table 8 shows the different responses of the 77 people in the treatment group.

Table 8
Frequencies of answers to the statement

	Frequency	Percent	Cumulative Percent
Totally disagree	0	0	0
Disagree	2	2.6	2.6
Somewhat disagree	2	2.6	5.2
Neutral	14	18.2	23.4
Somewhat agree	25	32.5	55.8
Agree	27	35.1	90.9
Totally agree	7	9.1	100.0
Total	77	100.0	100.0

0 respondents reported that they totally disagree with the statement. In total 4 respondents reported that they disagreed or somewhat disagreed with the statement. However, none of those 4 reported that they knew details about the GDPR which applies to the circumstance of visiting a website. Which is odd because the quiz provided exactly that information. The other 94.8 percent at least agreed with the statement or is neutral about it. Because of this, it will be concluded that the treatment group gained knowledge about privacy legislation that is applicable to the circumstance of visiting a website.

Because the sample is properly randomized, the with-and-without comparison is expected to give the causal effect of Knowledge on privacy legislation on Privacy concerns. The difference in means between groups of the self-reported Privacy Concerns on the seven-point Likert scale, will be formally tested with the Mann-Whitney U test since the data of Privacy concerns does not resemble a normal distribution (Appendix F).

Table 9
Output of Mann-Whitney U test for the variable Privacy Concerns

Treatment	N	Mean	Mean Rank	Sum of Ranks	Mann-Whitney U test	Asymp. Sig. (2-tailed)
No treatment	87	4.87	87.95	7652.00	2875.00	0.117
Treatment	77	4.67	76.34	5878.00		

***P < 0.01, **P < 0.05, *P < 0.10

Table 9 shows that the mean of Privacy concerns is higher for the control group than the treatment group. This implies that the respondents that completed the quiz about legislation - and therefore gained knowledge about this topic – reported to have less concerns about their online privacy than the respondents that had not completed the quiz about legislation. However, the Mann-Whitney U test shows that this difference in means is not significant. Therefore, there is no statistical evidence for the first hypothesis which means that it cannot be concluded that the knowledge on privacy legislation decreases an individuals' privacy concerns. The first hypothesis can thus be rejected.

5.3 Hypothesis 2

Hypothesis 2 – *Knowledge on privacy legislation improves individuals' perceived control over their online personal information* – is, like hypothesis 1, tested with a with-and-without comparison. The treatment effect is expected to have a causal effect on the dependent variable because the sample is properly randomized. Again, the difference in means of the variable Perceived Control between groups, will be formally tested with the Mann-Whitney U test since the data of Privacy concerns does not resemble a normal distribution (Appendix F).

Table 10

Output of Mann-Whitney U test for the variable Perceived Control

Treatment	N	Mean	Mean Rank	Sum of Ranks	Mann-Whitney U test	Asymp. Sig. (2-tailed)
No treatment	87	3.10	70.33	6118.50	2290.50	0.000***
Treatment	77	3.75	96.25	7411.50		

***P < 0.01, **P < 0.05, *P < 0.10

Table 10 shows that the mean of Perceived Control for the treatment group is 0.650 higher than the mean of the control group. This implies that the respondents that gained knowledge on privacy legislation felt like they are more in control over their online personal information than the respondents that did not gain knowledge. The Mann-Whitney U output shows that the difference in means between groups is statistically significant ($p < 0.05$). Therefore, the second hypothesis cannot be rejected which in turn means that it can be assumed with evidence that the quiz about legislation improves an individual's perceived control.

5.4 Hypothesis 3

The last hypothesis – *Improved individuals' perceived control over online personal information decreases their privacy concerns* – will determine the (causal) effect of perceived control on an individual's privacy concerns. In order to run the regression for the last hypothesis, a scatterplot has been computed to check for linearity.

Figure 4 shows the scatterplot for perceived control on privacy concerns per group. As can be seen, both lines are not parallel with each other. This points out that there is a moderating effect of knowledge on privacy legislation. Moreover, since the lines are both decreasing, knowledge causes an interaction effect. To account for this effect, the possible regression will include the interaction term C*T, which is calculated by the value for Perceived control times the value of whether or not someone is in the treatment group. Furthermore, the assumption of linearity will be made based on the scatterplot since no other model fitted the scatterplot better which would decrease the R^2.

Figure 4

Scatterplot of Perceived control on Privacy concerns per group

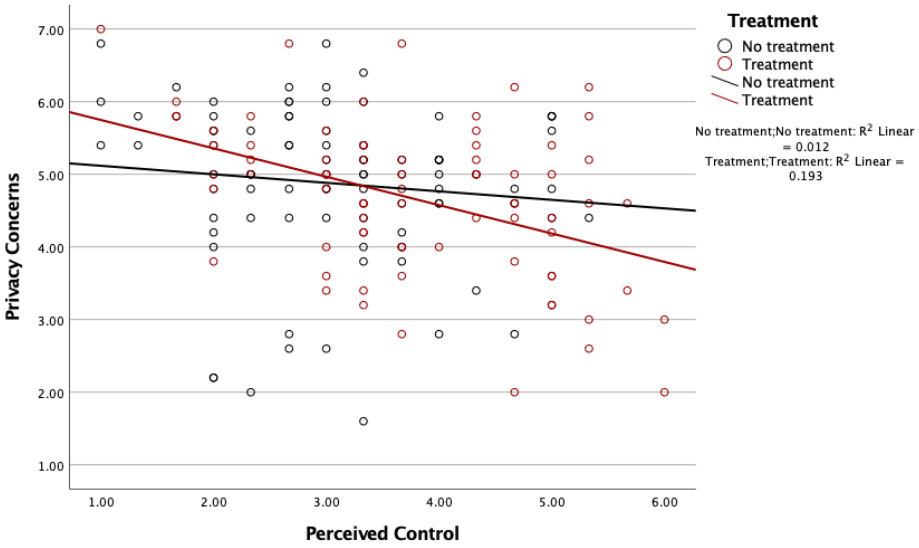
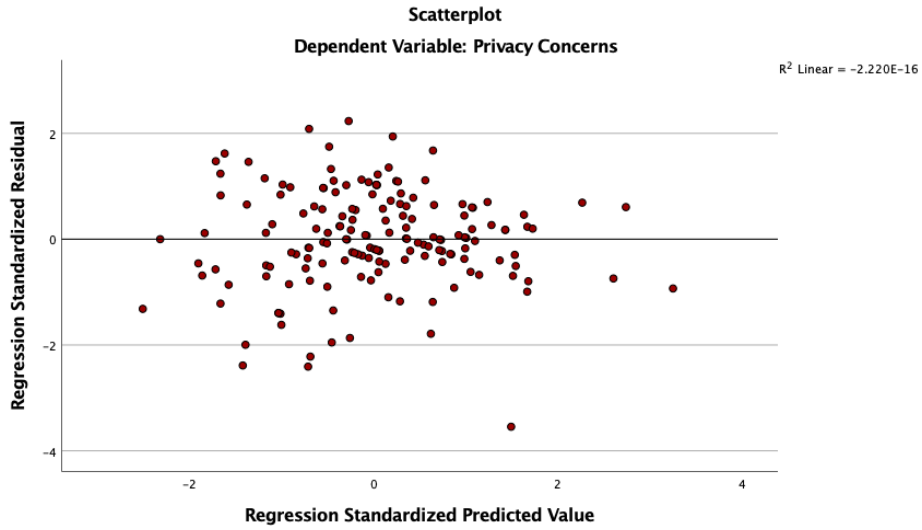


Figure 5

Scatterplot of Perceived control on the residuals Privacy concerns



The second and third assumption are a constant variance of the residuals, also known as homoscedasticity, and independence of the residuals respectively. The dots in figure 5 show that the variance in the residuals is equal among different levels of perceived control. This indicates no heteroscedasticity and that the assumption is probably met. The third assumption is also met because the residuals were independent as assessed by a Durbin-Watson statistic of 2.005.

The fourth assumption is no multicollinearity. This means that the independent variables aren't allowed to be related themselves. There was no evidence of multicollinearity, since all tolerance values were greater than 0.1. Of course, the variables 'treatment' and the interaction term, did have a tolerance value lower than 0.1, but this will not form a problem for the regression because the interaction term is defined as the product of the variable 'Treatment' and 'Perceived control'.

In the regression, multiple dummies were included for the categorical variables 'Marital status', 'Employment status', 'Educational level' and 'Time visiting websites per week'. To prevent multicollinearity issues, every variable excluded 1 dummy. The variable 'marital status' excluded a dummy for 'Single', the variable 'Employment status' excluded a dummy for 'Full-time' and the variable 'Educational level' excluded a dummy for 'Elementary'. Working with dummies means that the observation is in the excluded dummies if all dummies included in the regression are 0.

The last assumption is exogeneity. This assumes that the error term of the population is independent of the independent variable 'Perceived control' and the dependent variable 'Privacy concerns'. Although this assumption cannot be formally tested because only the residual (of the sample) is known, it can be assumed that the assumption is met because evidence is found for a properly randomized sample. However, since the sample is relatively small, it might be useful to control for other factors than only Perceived control to overcome the omitted variable bias. Another cause for the assumption of exogeneity to not be met, is reversed causality. However, this will not form a threat for the regression because it is not logical that privacy concerns are able to make a respondent feel more in control over their online personal information.

To obtain the highest possible R^2 and adjusted R^2, a hierarchical regression was run with including and excluding different variables. Table 11 shows the model that explains 30% of the variation in Privacy concerns. Although including the interaction term in the model did not result in a significant improvement of the R^2, the interaction term was still included in the model since the plot (figure 4) showed evidence for an interaction effect. In table 11, the coefficients for the regression with the highest possible R^2 and adjusted R^2 can be found. The multiple linear regression was computed with the following equation.

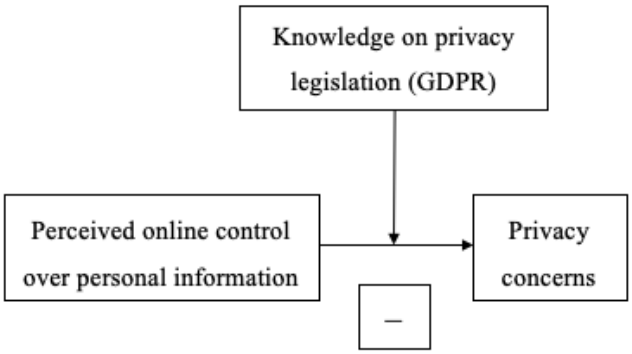
$$P_c = \alpha + \beta_1 C + \beta_2 T + \beta_3 (T * C) + \beta_4 G + \beta_5 A + \beta_6 T_p + \beta_7 E + \beta_8 Edu_2 + \beta_9 Edu_3 + \beta_{10} Edu_4 + \beta_{11} Edu_5 + \beta_{12} Edu_6 + \beta_{13} Mar_2 + \beta_{14} Mar_3 + \beta_{15} Mar_4 + \beta_{16} Emp_2 + \beta_{17} Emp_3 + \beta_{18} Emp_4 + \beta_{19} Emp_5 + \beta_{20} Emp_6 + \beta_{21} Emp_7 + \beta_{22} Time_2 + \beta_{23} Time_3 + \beta_{24} Time_4 + \beta_{25} Time_5 + \varepsilon$$

This is the regression Perceived control (C) on Privacy concerns (P_c); With (α) being the constant, (β_x) being the different parameters for the different variables, (T) being the treatment effect (whether someone completed the treatment-survey and thus gained knowledge on privacy legislation), (G) as the variable Gender, (A) being the variable Age, (T_p) being the variable Trust propensity, (E) as Previous privacy experience, (Edu_x) as the different educational levels, (Mar_x) as the different Marital statuses, (Emp_x) as the different employment statuses, ($Time_x$) as the different average amounts of time spend on visiting websites per week and (ε) as the residual.

The first thing to notice from the results of the regression in table 11 is that Perceived control does not have a statistically significant effect on privacy concerns ($\beta = -0.086$, $p>0.1$). This is probably because the interaction term is also included in the model. The interaction term and treatment are only significant on a 10% level ($\beta = -0.274$, $p< 0.1$ & $\beta = 0.931$, $p< 0.1$). Therefore, the hypothesis will be rejected. Because of the small sample size and the fact that all the variables were self-reported through a survey, the hypothesis could also not be easily rejected. In both groups privacy concerns decreased as perceived control increased, this would imply that there still is an effect of perceived control on privacy concerns, but that effect cannot be proved through this research. A negative effect of perceived control on privacy concerns is also in line with most of the literature on perceived control and privacy concerns (Hoadley, Xu, Lee, & Rosson, 2010). When is assumed that due to the small sample size, the coefficients were not significant, it could be concluded that perceived control does decrease an individuals' privacy concerns, but this effect will be bigger when that individual gained knowledge on privacy legislation.

Figure 6

visualization of moderator/interaction effect



Besides the variables of perceived control and treatment, three other variables are noticeable. Those are the variables Previous privacy experience ($\beta = 0.169$, $p< 0.05$), Age ($\beta = 0.024$, $p< 0.05$) and Retired ($\beta = -1.040$, $p< 0.1$). Controlling for the variable Previous privacy experience was based on the work of Xu (2007). Nevertheless, the other control variable based on the work of Xu, Trust propensity, did not have a significant parameter. Therefore, trust propensity does not affect privacy concerns. The small effect (0.169) of previous privacy experience could be explained by the logic that people who have had a negative experience with their online privacy, do have more concerns because they think it will happen again. Whether someone gained knowledge on privacy legislation did not change this. The positive coefficient of Age means that when respondent A is 1 year older than respondent B, with all other characteristics equal, respondent A has 0.024 more concerns about their online privacy than respondent B. This looks like a small effect but when the age difference is for example 50 the privacy concerns would increase or decrease with 1 (keeping all other variables equal). This would result in a respondent going from "Agree" with having privacy concerns to being "neutral" about having them. The last

noticeable variable, whether or not an individual is retired, represents a big effect on privacy concerns. Keeping all other characteristics equal, will a retirement decrease the privacy concerns of an individual with 1 (on a scale from 1 to 7). However, since the effect is not significant and the sample is quite small, this effect will not be accounted for. Nevertheless, the coefficient of 'retired' being negative is contradictory to the effect of 'age'. When an individual will get older they are also more likely to retire from their work, but 'age' has a positive effect on privacy concerns while 'retired' does not.

Altogether, when the significance level is enlarged because of the small sample size, the following formula could predict the value of an individual's privacy concerns:

$$P_c = 3.979 - 0.086 * C + 0.931 * T - 0.274 * T * C + 0.169 * E + 0.024 * A - 1.040 * Emp_4 + \varepsilon$$

Again with Perceived control (C), Privacy concerns (P_c), (T) as the treatment effect (whether someone completed the treatment-survey and thus gained knowledge on privacy legislation), (A) being the variable Age, (E) as Previous privacy experience, (Emp_4) as the dummy for being 'retired', a constant of 3.979 and (ε) as the residual.

Table 11
Output of multiple linear regression on Privacy Concerns

Independent Variable	**Coefficient**	**Std. Error**
(Constant)	3.979***	(1.297)
Perceived Control	-.086	(.109)
Treatment	.931*	(.555)
Interaction T*C	-.274*	(.153)
Gender	.104	(.178)
Age	.024**	(.009)
Trust Propensity	-.123	(.087)
Previous Privacy Experience	.169**	(.075)
Educational Level		
Elementary	-	-
Secondary	.411	(1.113)
Diploma	.274	(1.110)
Bachelor	.269	(1.115)
Master	.492	(1.136)
Other Level Of Education	.684	(1.246)
Marital Status		
Single	-	-
Married	-.159	(.283)
Divorced/Separated	-.813	(.555)
Other Marital Status	-.195	(.239)
Employment Status		
Full-Time	-	-
Part-Time	-.110	(.211)
Self-Employed	-.513	(.480)
Retired	-1.040*	(.601)
Full-Time Student	-.127	(.263)
Unemployed	-.334	(.449)
Other Employment Status	.353	(.418)
Time Visiting Websites Per Week		
0-2 Hours	-	-
3-5 Hours	-.328	(.216)
6-9 Hours	-.146	(.264)
10-13 Hours	.028	(.394)
14+ Hours	.055	(.329)
Observations	164	
R^2 (adjusted R^2)	0.304 (0.178)	
Residual std. error	0.983 (df = 138)	
F statistic	2.410*** (df = 25;138)	

***P < 0.01, **P < 0.05, *P < 0.10

6 CONCLUSION

This thesis investigated the relation between knowledge on privacy legislation and an individual's privacy concerns. This was done to find an answer to the scientifically relevant question whether or not perceived control decreases privacy concerns, and to find an answer to the socially relevant question whether or not knowledge on privacy laws (or closing the 'knowledge gap') could improve perceived control or decrease privacy concerns. To contribute to those matters and in order to formulate an answer to the research question, three hypotheses needed to be evaluated.

The first hypothesis was rejected based on the result of the with-and-without comparison. This comparison proved that there was no significant difference in self-reported privacy concerns between the group that gained knowledge on privacy knowledge and the group that didn't gain knowledge on privacy laws. This implies that there was no direct effect of knowledge on privacy concerns. The research question could therefore be answered (based on this finding only) with "there is no relationship between knowledge on privacy legislation and privacy concerns". However, hypothesis 2 and 3 show contradicting evidence that there is actually an indirect relation between knowledge on privacy legislation and privacy concerns.

Hypothesis 2, which had been incorporated in the thesis to check whether knowledge on privacy legislation positively affects perceived control, could not be rejected. The with-and-without comparison showed that the respondents in the treatment group reported that they significantly feel more in control than the respondents in the control group ($\beta = 0.650$, $p < 0.01$). Therefore, it could be assumed with evidence that knowledge on privacy legislation positively affects an individual's perceived online control over personal information. This finding is in line with the work from Xu (2007) and Armstrong and Culnan (1999).

The third hypothesis was (not easily) rejected. A multiple linear regression analysis has been conducted to investigate the relationship between perceived control and privacy concerns. The coefficients of the regression showed that perceived control indeed negatively affects privacy concerns but the effect depended heavily on whether or not an individual gained knowledge on privacy legislation. However, since this effect was not significant, it may not be concluded that perceived control decreases privacy concerns based on this research.

Concluding, this research improved perceived control by letting individuals gain knowledge on privacy legislation through a quiz. Perceived control does in turn negatively relate to an individual's privacy concerns but this was not statistically significant. The quiz about privacy legislation contributed to this effect as a mediator. Based on the findings in this research, it may not be concluded with significant evidence that perceived control or knowledge on privacy legislation, decreases privacy concerns.

7 DISCUSSION AND LIMITATIONS

7.1 Discussion

It could be relevant for (but not limited to) the policy of NGO's, the government and webstores/companies when perceived control does indeed decrease privacy concerns. A policy wherein the entity causes individuals to feel like they are in control over their online personal information could result in improved results for the entity. Elaborating on the research of Acquisti (2004), if the websites of companies/webstores have a user-friendly interface and it is very clear how the visitors could keep control over their personal information, it could result in more sales because the visitors are less concerned.

When the results of the thesis are placed in a broader perspective and it is assumed that knowledge on legislation does indeed reduce privacy concerns, it could be argued that the government (with the help of NGO's) must give more attention to the knowledge of inhabitants on privacy matters. Because of the knowledge gap, individuals lack knowledge on the basics of privacy laws like the GDPR and the basics of online data tracking. When the government would provide the inhabitants proactive with knowledge on privacy legislation (privacy matters), i.e. by teaching extra courses at schools, privacy concerns may decrease on a great level. Moreover, since the thesis showed that age also played a role in the privacy concerns of individuals, NGO's could teach the elderly about privacy matters in community centers, retirement homes, sports centers etcetera.

7.2 Limitations

The results are not satisfying and do not correspond with main findings in the literature. The finding that perceived control negatively relates to privacy concerns is in line with the research of Xu et al. (2010) and with the research of Acquisti and Gross (2006). However, since the results were not significant and the effect was only evidently visible when an individual gained knowledge, i.e. when an individual was in the treatment group, the debate about whether or not perceived control decreases privacy concerns, could not be ended. Furthermore, the regression did not show convincing evidence for the research of Brandimarte et al. (2012) to be indefensible.

Besides the results not being significant (without the effect of age and previous privacy experience), the research contained other imperfections. The first point of criticism could be that the design of this research was not able to find the causal effect of knowledge on privacy legislation. Based on how the survey was put together, it could be argued that not the knowledge on privacy legislation increased perceived control, but the knowledge on data tracking. In the quiz of the treatment-survey, information was also given about some basics about data tracking to understand the legislation better. By gaining knowledge about this information, respondents could also perceive their online control over their

personal information to improve. Although this has been done to let respondents understand the privacy laws better, it could have biased the results and not give the effect of knowledge on privacy laws only.

The effects of the coefficients in the multiple linear regression of perceived control on privacy concerns are quite small. This could be due to the survey measuring the answers to the statements on a seven-point Likert scale. For future research, it could make the effects more noticeable when the answers to the statements are measured with a five-point Likert scale. This would imply less possibilities, so the data will be more sensitive.

Another criticism concerns the internal and external validity of the research. Although all the hypotheses leaned on more or less perfect randomization, the assumption of randomization could not be assumed because of the small sample size. This makes the research immediately less internally valid. Moreover, the sample was not a perfect reflection of the population. The average respondent was for example younger than the average respondent of the population. The sample size was too small to be representative for the population, making it less externally valid to conclude and make assumptions for the population.

The quiz about legislation being limited in size was also a limitation. A shorter quiz has been chosen to let the respondents spend less time on the survey. Otherwise, there would be a chance that the respondents no longer wanted to complete the quiz. Because the quiz was limited in the amount of questions, it allowed the quiz to appear as if a lot was protected by legislation, while there are actually plenty of things that are not protected. The legislation may therefore appear more positive than it actually is.

The survey was self-administered. Therefore, it could not be checked whether the respondents answered the questions in the survey with optimized care. Furthermore, most questions in the survey contained a subjective element which could not be objectively answered. Statements that were answered on a Seven-point Likert scale, could differ between respondents although they both feel the same way towards the statement. This is because a respondent is limited in his ability and sensitivity to his emotions. Since the sample was not perfectly randomized, the subjectively answered statements could bias the results.

Future research could improve the limitations by not collecting quantitative data, but qualitative data. This could be done by measuring the variables with objective criteria. Knowledge on privacy laws could be measured by the amount of correctly given answers to the questions in the quiz, perceived control could be measured with different groups giving different control over their online personal information when they apply for the research, and privacy concerns could be measured by the outcome whether or not an individual wanted to disclose their personal information for example.

8 REFERENCE LIST

Acquisti, A. (2004, May). Privacy in Electronic Commerce and the Economics of Immediate Gratification. *In Proceedings of the 5th ACM conference on Electronic commerce* (pp. 22-29). New York: ACM.

Acquisti, A., & Gross, R. (2006). Imagined communities: awareness, information sharing, and privacy on the Facebook. *Proceedings of the Sixth Workshop on Privacy Enhancing Technologies*.

Acquisti, A., Brandimarte, L., & Loewenstein, G. (2015, January 30). Privacy and human behavior in the age of information. *Science, 347*(6221), 509.

Acquisti, A., Taylor, C., & Wagman, L. (2016). The Economics of Privacy. *Journal of Economic Literature, 54*(2), 442-492.

Armstrong, P. K., & Culnan, M. J. (1999). Information Privacy Concerns, Procedural Fairness and Impersonal Trust: An Empirical Investigation. *Organization Science, 10*(1), 104-115.

Brandimarte, L., Acquisti, A., & Loewenstein, G. (2012). Misplaced Confidences: Privacy and the Control Paradox. *Social Psychological and Personality Science, 4*(3), 340-347.

Chang, Y., Wong, S. F., Libaque-Saenz, C. F., & Lee, H. (2018). The role of privacy policy on consumers' perceived privacy. *Government Information Quarterly, 35*, 445-459.

Culnan, M. J. (1995). Consumer Awareness of Name Removal Procedures: Implications for Direct Marketing. *Journal of Interactive Marketing, 9*, 10-19.

ePrivacy company Cybot. (2019, August 21). *https://www.cookiebot.com/en/tracking-cookies/*. Retrieved April 21, 2020, from https://www.cookiebot.com/en/: https://www.cookiebot.com/en/tracking-cookies/

Finstad, K. (2010, May). Response interpolation and scale sensitivity: Evidence against 5-point scales. *Journal of usability studies, 5*(3), 104-110.

Hoadley, C., Xu, H., Lee, J., & Rosson, M. (2010). Privacy as information access and illusory control: The case of the Facebook News Feed privacy outcry. *Electronic Commerce Research and Applications, 9*, 50-60.

Hoofnagle, C. J., & Urban, J. M. (2014). Alan Wetsin's Privacy Homo Economicus. *Wake Forest Law Review, 49*(2), 261-318.

Lin, T. (2019, August 30). Valuing Intrinsic and Instrumental Preferences for Privacy. *The University of Chicago Booth School of Business*.

McAfee, A., & Brynjolfsson, E. (2012, October). Big Data: The Management Revolution. *Harvard Bussiness review, 68*(128).

McKnight, D. H., Choudhury, V., & Kacmar, C. (2002). Developing and Validating Trust Measures for E-Commerce: An Integrative Typology. *Information Systems Research, 13*(3), 334-359.

Pressley, M., Tanenbaum, R., Mcdaniel, M. A., & Wood, E. (1990). What happens when university students try to answer prequestions that accompany textbook material? *Contemporary Educational Psychology, 15*, 27-35.

Reed, G. M., Taylor, S. E., & Kemeny, M. E. (1993). Perceived Control and Psychological Adjustment in Gay Men with AIDS. *Journal of Applied Social Psychology, 23*(10), 791-824.

Smith, H. J., Milberg, J. S., & Burke, J. S. (June 1996). Information Privacy: Measuring Individuals' Concerns About Organizational Practices. *MIS Quarterly, 20*(2), 167-196.

Tatonetti, N., Ye, P., Daneshjou, R., & Altman, R. (2012, March 14). Data-Driven Prediction of Drug Effects and Interactions. *Science Translational Medicine, 4*(125).

Westin, A. (1967). Privacy and Freedom. New York: Atheneum Publishers.

Xu, H. (2007, December). The effects of self-construal and perceived control on privacy concerns. *Proceedings of 28th Annual International Conference on Information Systems*.

9 APPENDIX

A: Survey

To obtain plausible results, it was necessary that the respondents understood the information and questions in the survey. Therefore, the survey was written in English and Dutch because most of the respondents were Dutch. If the respondent wanted to take the survey in Dutch, they could click on the table in the upper right corner of their screen and switch to "Dutch". Besides this feature, every part of the survey was timed so that the data could be analyzed and may be eliminated from the data.

Link to the survey:
https://erasmusuniversity.eu.qualtrics.com/jfe/preview/SV_3wmilaVDpjC31Ot?Q_CHL=preview&Q_SurveyVersionID=current

English

Dear respondent,

This survey will take a maximum of 5 minutes.

It is necessary to read the instructions carefully and to complete the survey in one session.

Thanks in advance for completing this survey!

Survey Completion
0% 100%

What is your gender?

- Male
- Female

What is your age?

0 10 20 30 40 50 60 70 80 90 100

Age

What is your highest level of education?

- Elementary
- Secondary
- Diploma
- Bachelor
- Master
- Doctoral
- Other

What is your marital status?

- ◯ Single
- ◯ Married
- ◯ Divorced/seperated
- ◯ Widowed
- ◯ Other

What is your employment status?

- ◯ Full-time
- ◯ Part-time
- ◯ Self-employed
- ◯ Retired
- ◯ Full-time student
- ◯ Unemployed
- ◯ Other

On average, how much time per week do you spend on visiting websites (without social-media timelines)?

- ◯ 0-2 hours
- ◯ 3-5 hours
- ◯ 6-9 hours
- ◯ 10-13 hours
- ◯ 14+ hours

To what extent do you agree or disagree with each of the following statements?

	Strongly disagree	Disagree	Somewhat disagree	Neutral	Somewhat agree	Agree	Strongly agree
Most people are honest	◯	◯	◯	◯	◯	◯	◯
I usually trust people until they give me a reason not to trust them	◯	◯	◯	◯	◯	◯	◯
Most people keep their promises	◯	◯	◯	◯	◯	◯	◯

To what extent do you agree or disagree with each of the following statements?

	Strongly disagree	Disagree	Somewhat disagree	Neutral	Somewhat agree	Agree	Strongly agree
In the past, I have had a negative experience regarding my online privacy	◯	◯	◯	◯	◯	◯	◯
I am aware of the privacy issues and practices in our society	◯	◯	◯	◯	◯	◯	◯

Do you live in the European Union (EU)?

- ◯ Yes
- ◯ No

→ *If no, the survey was closed for the respondent.*

→ *If yes, the respondent could continue the survey.*

Do you know what internet cookies are?

- ◯ Yes
- ◯ No

→ *If no, the respondent was showed the following information*

An internet cookie (also known as HTTP cookie, web cookie, or simply cookie) is a small piece of data sent from a website and stored on the computer of the visitor. The cookies are able to remember information about the visitor and can provide this information to the website's operator.

→ *If yes, the respondent was not showed information this information.*

Do you know what the General Data Protection Regulation (GDPR) holds in general?

- ◯ Yes
- ◯ No

→ *If no, the respondent was showed the following information.*

The General Data Protection Regulation is a regulation in EU law on data protection and privacy in the European Union (EU). The GDPR aims primarily to give control to individuals over their personal data and privacy.

→ *If yes, the respondent was showed the following*

Do you know specific details about the GDPR with regard to cookies, privacy policies and the processing of personal information when visiting a website?

- ◯ Yes
- ◯ No

**If the respondent answered this question with "Yes", their data could be eliminated from the dataset because these respondent couldn't gain more knowledge because they already knew specific details about the GDPR.*

After these questions, the respondents were divided into a control group and a treatment group. The following part of the survey (a quiz about sports) applies to the ***control group****. The right answers of the quiz were given after the respondent answered the question with "true" or "false". In the pictures below every question is answered with true for clarity.*

English

You are now going to make a quiz about sports. You have to decide whether 10 statements are true or false. The purpose of this quiz is to let you gain knowledge about sports. It is therefore important that you understand the explanation of the correct answers to the statements, even if you have given the correct answer!

Survey Completion
0% 100%

→

France has won the last (FIFA) World Cup in football, while they are also the record holder in most World Cup titles.

- (●) True
- ( ) False

False, although France has won the last World Cup in 2018, they do not have the most World Cup titles. Brazil has won the FIFA World Cup the most often. With 5 titles, winning the last in 2002, they are ahead of every other country.

The white colour makes the first move in a game of chess.

- (●) True
- ( ) False

True. In chess, the white colour makes the first move. However, white hasn't always been the colour that makes the first move. Multiple sources found historical evidence that black also made the first move in the past. Making the first move is important because the player that makes the first move has a slight advantage, called the first-move advantage. Researchers have found significant evidence that the maker of the first move wins between 52 - 56 % of all games,

In tennis, deuce means that the score is 30-30.

- (●) True
- ( ) False

False, scores from zero to three points are described as "love", "15", "30", and "40", respectively. If at least three points have been scored by each player, making the player's scores equal at 40 apiece, the score is not called out as "40–40", but rather as "deuce". If at least three points have been scored by each side and a player has one more point than his opponent, the score of the game is "advantage" for the player in the lead.

The last winner of the Toure de France is Egan Bernal

- (●) True
- ( ) False

True, the Colombian cyclist won the Toure de France in 2019.

Michael Phelps is the record holder of most gold Olympic medals.

- (●) True
- ( ) False

True, Michael Phelps has won 23 gold Olympic medal in his career. He even holds the record for most gold Olympic medals (8) in one tournament.

A regular soccer (football) match without extra time lasts 100 minutes.

- (●) True
- ( ) False

False, a regular soccer (football) match lasts 90 minutes.

Peter Wright has defeated Michael van Gerwen in the last PDC World Darts Championship with 7-3.

- (●) True
- ( ) False

True, the tournament is being held in England, and an Englishman (named Phil Taylor) also holds the record for most PDC World Darts Championship victories.

In golf, a birdie means that a player scored 1 under par, while an eagle means that a player scored 2 under par.

- (●) True
- ( ) False

True, other terms used for scoring under par are Albatross (-3) and Condor (-4).

In basketball each team can have a maximum of 4 players on the court. Substitutions can be made as many times as a team wishes during the game.

- (●) True
- ( ) False

False, Although substitutions can be made as many times as a team wishes during the game, each team can have only have a maximum of 5 players on the court.

The world record for the bench press lift is 487.6 kilograms / 1,075 pounds.

- (●) True
- ( ) False

True, the record is being held by the American Ryan Kennely.

After the quiz, the last questions were asked.

English

To what extent do you agree or disagree with each of the following statements?

	Strongly disagree	Disagree	Somewhat disagree	Neutral	Somewhat agree	Agree	Strongly agree
I think I have control over my personal information when I visit a website.	○	○	○	○	○	○	○
I think I have control over my personal information after providing my information to a website.	○	○	○	○	○	○	○
I feel like I am being protected by privacy legislation when visiting websites.	○	○	○	○	○	○	○

To what extent do you agree or disagree with each of the following statements?

	Strongly disagree	Disagree	Somewhat disagree	Neutral	Somewhat agree	Agree	Strongly agree
I am concerned that a website is collecting too much personal information about me.	○	○	○	○	○	○	○
I am concerned that a website may not take measures to prevent unauthorized access to my personal information.	○	○	○	○	○	○	○
I am concerned that a website may keep my personal information in a non-accurate manner in their database.	○	○	○	○	○	○	○
I am concerned that a website may share my personal information with other parties without getting my authorization.	○	○	○	○	○	○	○
Overall, I feel unsafe about providing personal information to a website.	○	○	○	○	○	○	○

Survey Completion

0% 100%

*The next part applies to the **treatment group**. The right answers of the quiz were given after the respondent answered the question with "true" or "false". In the pictures below every question is answered with true for clarity.*

English

You are now going to make a quiz about privacy legislation. You have to decide whether 10 statements about online privacy are true or false. The purpose of this quiz is to let you gain knowledge about the GDPR. It is therefore very important that you read and understand the explanation of the correct answers to the statements, even if you have given the correct answer!

Survey Completion
0% 100%

→

A website does not have to give information about why and how it processes your personal information.

- (●) True
- ( ) False

False, the collector of the personal information has to inform the owner of the personal information about:

1. **Who the collector is and how he or she can be reached**
2. **For what reason the personal data will be processed**
3. **The type of personal data that will be processed**
4. **Details about the lawful basis for processing data**
5. **For how long the data will be stored**
6. **The types of organizations the data will be shared with (eCommerce platforms, email providers, etc.)**

If a website has a privacy policy, it means that the site cannot sell your address and purchase history to other companies or the government.

- (●) True
- ( ) False

False, a privacy policy is not legislation like the GDPR which can give you rights. A privacy policy is a statement that discloses some or all of the ways a party gathers, uses, discloses, and manages a customer's or client's data. The privacy policy can therefore hold that the operator of a website is able to sell your personal information. The only condition is you have consented to the privacy policy. A privacy policy often contains the information mentioned in the previous question.

If a website has a privacy policy, it means that the site cannot share information about you with other companies, unless you give the website your permission.

- (●) True
- ( ) False

True. It is important that you always read the privacy policy so that you know what for you are giving your permission. Article 12 of the GDPR states that information given in a privacy policy needs to be provided in "a concise, transparent, intelligible and easily accessible form, using clear and plain language". The privacy policy cannot therefore consist of long texts and difficult words which are hard to understand.

If a company wants to follow your internet use across multiple sites on the internet, it must first obtain your permission.

- (●) True
- ( ) False

True, following your internet use can be done by tracking/third-party/marketing cookies but the GDPR sets strict requirements for them. When you give permission for tracking cookies, the website places tracking cookies that are able to follow you on all websites that you visit. This is also the main reason that the clothing you were searching for online, appears on your social media timeline(s).

It is allowed for website operators to deny access to their websites if visitors do not accept their cookies.

- (●) True
- ( ) False

False, this is not allowed. So called "cookiewalls" where operators deny access to their website with a blockage, are forbidden according to the GDPR. Article 7 of the GDPR states that consent must be given freely via a clear affirmative act (opt in). When a website uses a cookiewall, the consent in line with the GDPR cannot be given.

When you enter a website, the cookies are already able to gather personal information.

- (●) True
- ( ) False

False, The GDPR states that data subjects have to give their consent before cookies can be placed.

You can request deletion of all the personal data that a company has about you. If a company does not comply with your request it would be illegal.

- (●) True
- ( ) False

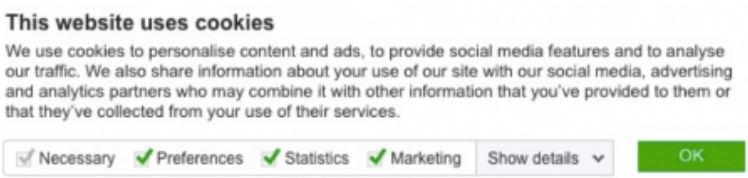
True, article 17 GDPR states the right to erasure. The right to be forgotten is in some circumstances possible, for instance when the data is no longer required for its intended purposes or the data subject has withdrawn its consent.

Pursuant to the GDPR, you can obtain access to all personal information a party has about you.

- (●) True
- ( ) False

True. Under Article 15 of the GDPR, data subjects are allowed to request information about any of their personal data that a party is processing. This is known as a Subject Access Request.

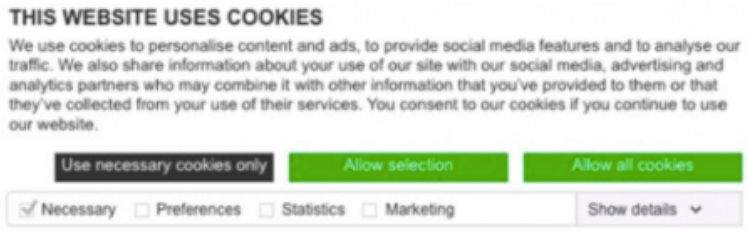
If you visit a website, and the website asks for your consent with all the checkboxes already ticked like the following picture:



The usage of cookies by this website would be legal if you press "OK".

- (●) True
- ( ) False

False, this would be illegal. A website that has the consent checkboxes for its cookies already ticked, as in the previous picture, cannot get a valid consent of the visitor. If the company would nevertheless use the cookies, the usage of the cookies would be illegal. In the previous picture, all cookies are already ticked and a visitor would have to "opt-out" all the boxes if he or she does not want to give permission for all cookies. According to the GDPR only consent given on a basis of "opt-in" consent, is legal. Like the following picture:



This means that the old cookie consent popups that featured only an OK-button without details of cookie types, durations, purposes, and which third parties' personal data is shared with, are illegal in the EU.

If a party violates the GDPR they can be fined.

- (●) True
- ( ) False

True, a fine for a minor violation of the GDPR could be around €100,000. Greater violations can go up to €20,000,000 or 4% of the worldwide annual turnover. Since the fines are really high, companies are extra careful with data. Nevertheless, some companies did already receive huge fines.

After the quiz, the last questions were asked.

English

To what extent do you agree or disagree with the following statement?

	Strongly disagree	Disagree	Somewhat disagree	Neutral	Somewhat agree	Agree	Strongly agree
After all this information, I feel like I gained knowledge about privacy legislation that is applicable to the circumstance of visiting a website.	○	○	○	○	○	○	○

To what extent do you agree or disagree with each of the following statements?

	Strongly disagree	Disagree	Somewhat disagree	Neutral	Somewhat agree	Agree	Strongly agree
I think I have control over my personal information when I visit a website.	○	○	○	○	○	○	○
I think I have control over my personal information after providing my information to a website.	○	○	○	○	○	○	○
I feel like I am being protected by privacy legislation when visiting websites.	○	○	○	○	○	○	○

To what extent do you agree or disagree with each of the following statements?

	Strongly disagree	Disagree	Somewhat disagree	Neutral	Somewhat agree	Agree	Strongly agree
I am concerned that a website is collecting too much personal information about me.	○	○	○	○	○	○	○
I am concerned that a website may not take measures to prevent unauthorized access to my personal information.	○	○	○	○	○	○	○
I am concerned that a website may keep my personal information in a non-accurate manner in their database.	○	○	○	○	○	○	○
I am concerned that a website may share my personal information with other parties without getting my authorization.	○	○	○	○	○	○	○
Overall, I feel unsafe about providing personal information to a website.	○	○	○	○	○	○	○

Survey Completion

0% 100%

Final page for both the control and the treatment group.

We thank you for your time spent taking this survey.
Your response has been recorded.

Survey Completion

0% 100%

B: Reply of the Dutch newspaper “het Algemeen Dagblad (AD)”

DE PERSGROEP.

Geachte
heer Van Oeveren,

Hartelijk dank voor uw e-mail met betrekking tot
onze cookiewall. Graag lichten wij die toe.
DPG Media gebruikt onder meer cookies om gebruikers
van haar websites de meest relevante artikelen en advertenties te tonen. Om gebruik te kunnen maken van de websites van DPG Media dienen gebruikers daarom toestemming te geven middels de cookiemuur voor het plaatsen van cookies.

De Telecommunicatiewet bevat de regels omtrent
deze cookies, welke voortkomen uit de Europese ePrivacy richtlijn. Op grond van deze regels mogen bepaalde cookies niet zonder toestemming worden geplaatst. Volgens de wetgever is een cookiemuur een rechtmatige manier om deze toestemming te verkrijgen. Het
principe van 'voorwaardelijk toegang' staat in de ePrivacy richtlijn.

De Autoriteit Persoonsgegevens heeft onlangs uitgesproken
dat een cookiemuur op grond van de Algemene Verordening Gegevensbescherming (AVG) niet zou zijn toegestaan. Naar onze mening verandert de AVG de regels omtrent cookies echter niet. DPG Media beraadt zich op dit moment verder over de zienswijze van de Autoriteit
Persoonsgegevens. Mocht op basis van deze uitkomsten aanpassingen benodigd zijn in onze cookiepolicy, dan zullen wij deze aanpassingen doorvoeren.

Wij
vertrouwen erop u hiermee voor dit moment afdoende te hebben geïnformeerd.

Met
vriendelijke groet,
Addy,
Data Protection Office DPG Media

C: Descriptive plots of the data

Figure 1

Boxplots of the total duration of the quiz per group

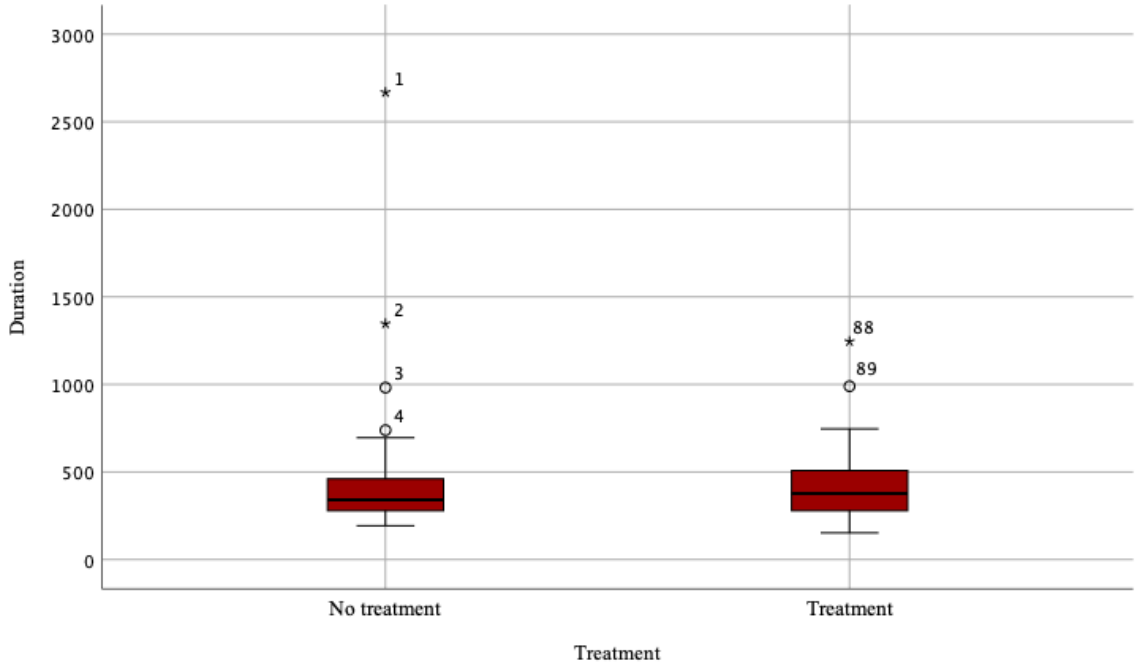


Figure 2

Distribution of the frequency of age per group

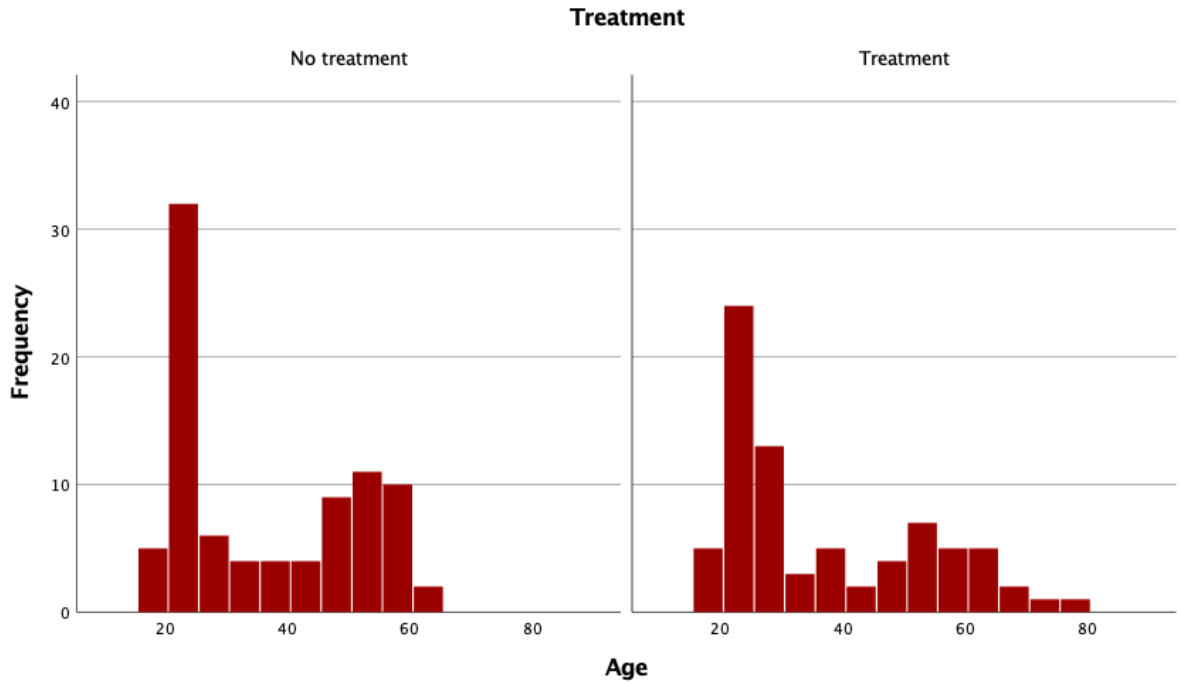


Figure 3

Pie chart of the different educations of the sample

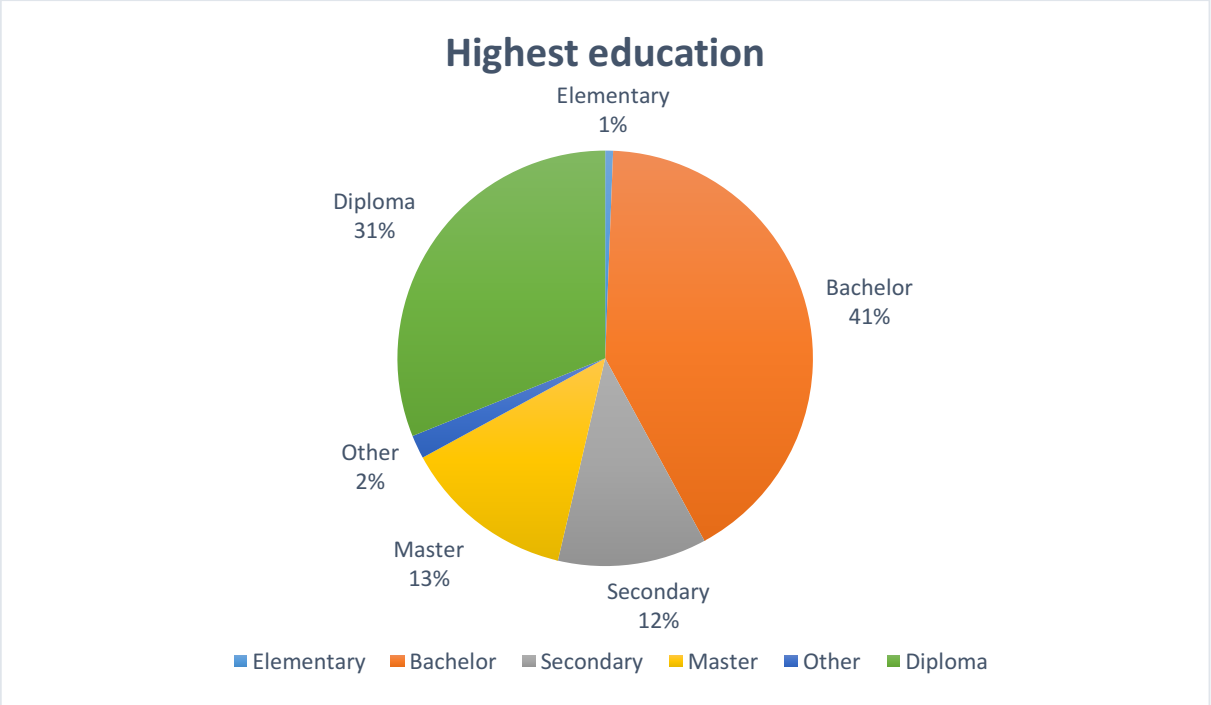
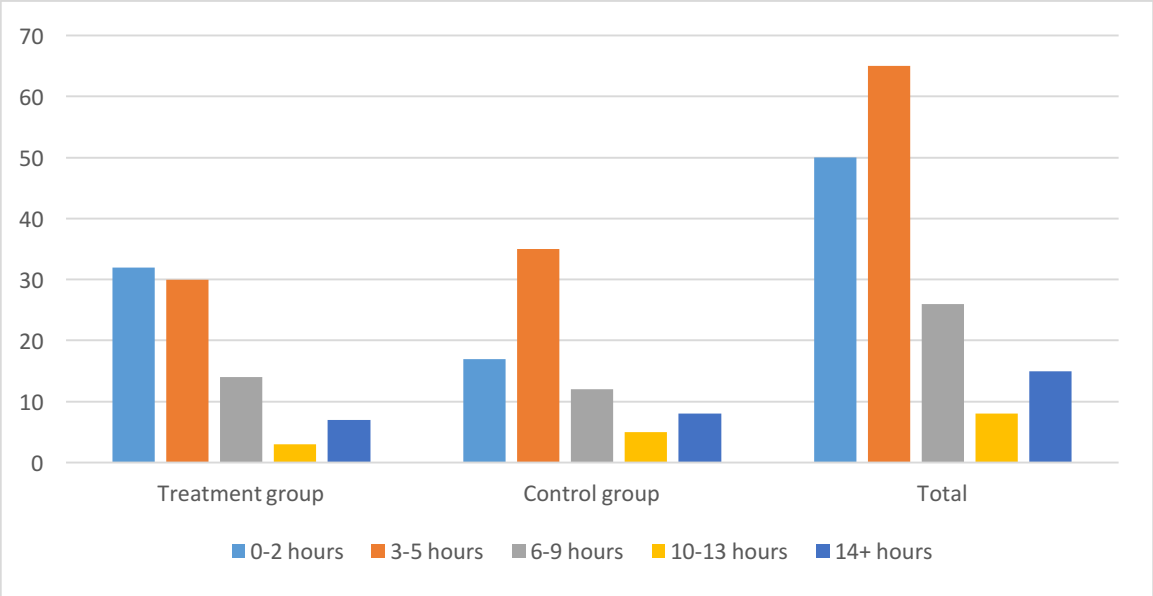


Figure 4

Histograms of the respondents' visiting time of websites per week



D: Histograms

Figure 1

Histogram of the distribution of the variable Duration of quiz

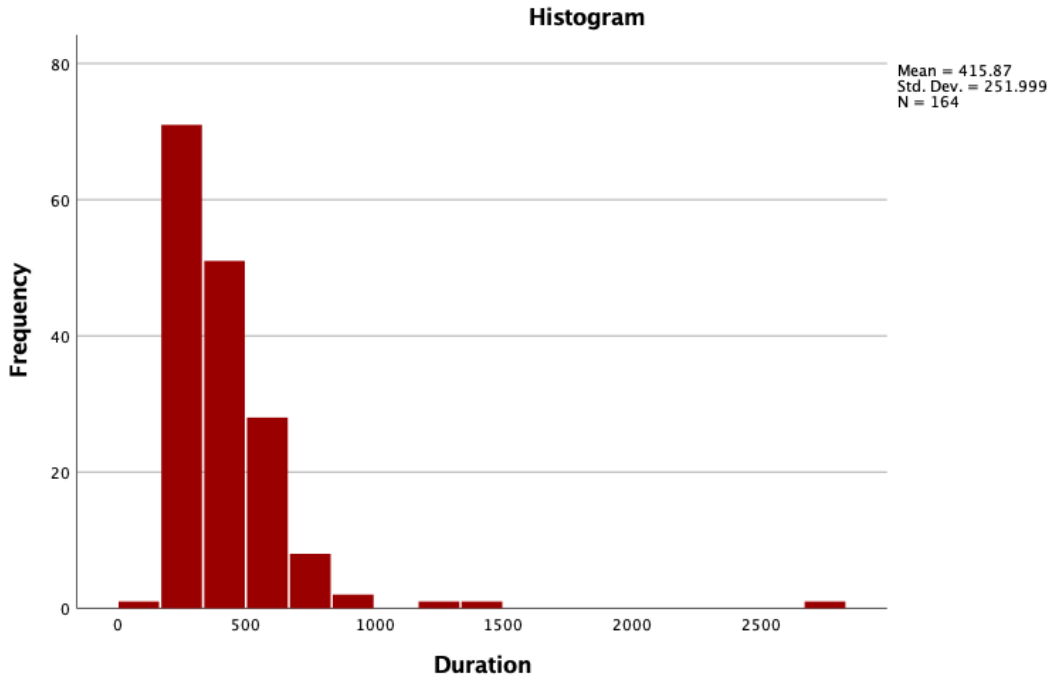


Figure 2

Histogram of the distribution of the variable Age

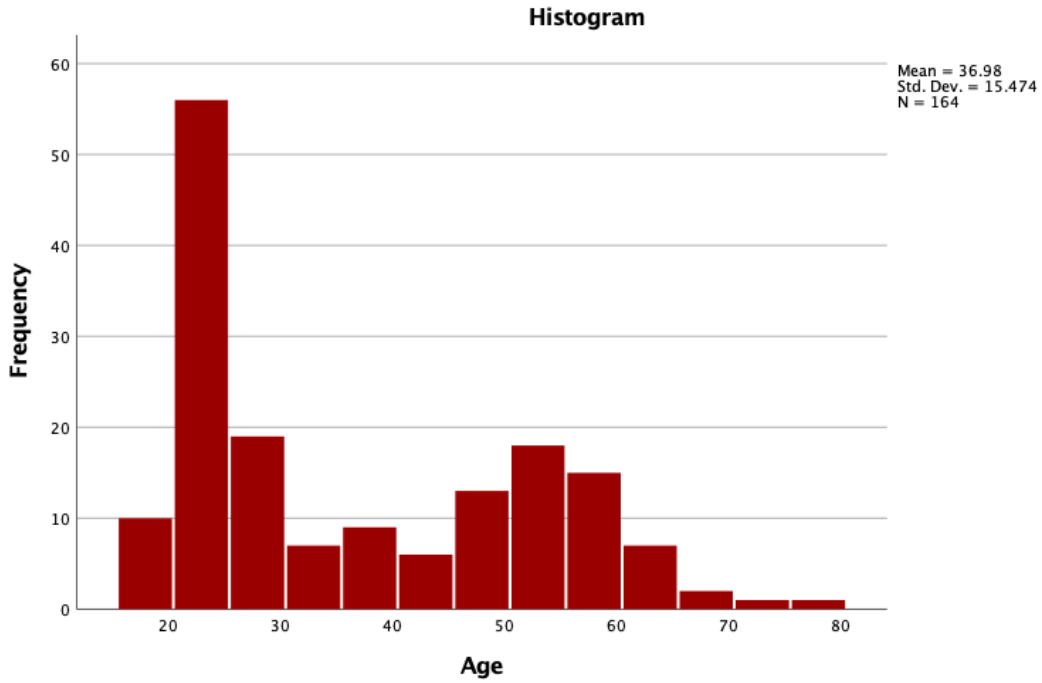


Figure 3

Histogram of the distribution of the variable Time visiting websites per week

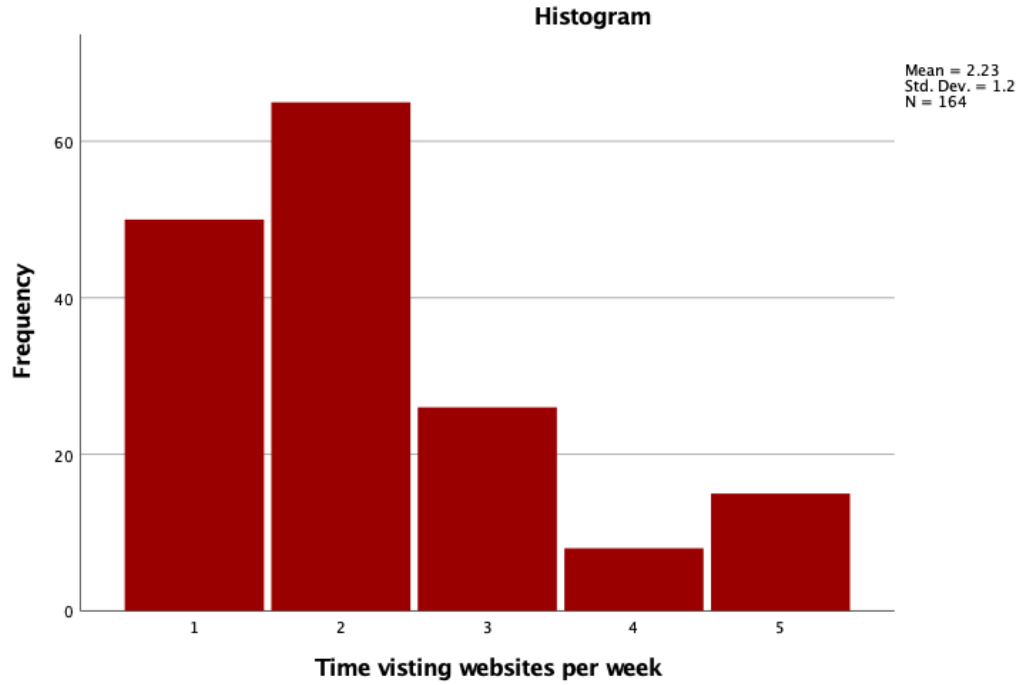


Figure 4

Histogram of the distribution of the variable Trust propensity

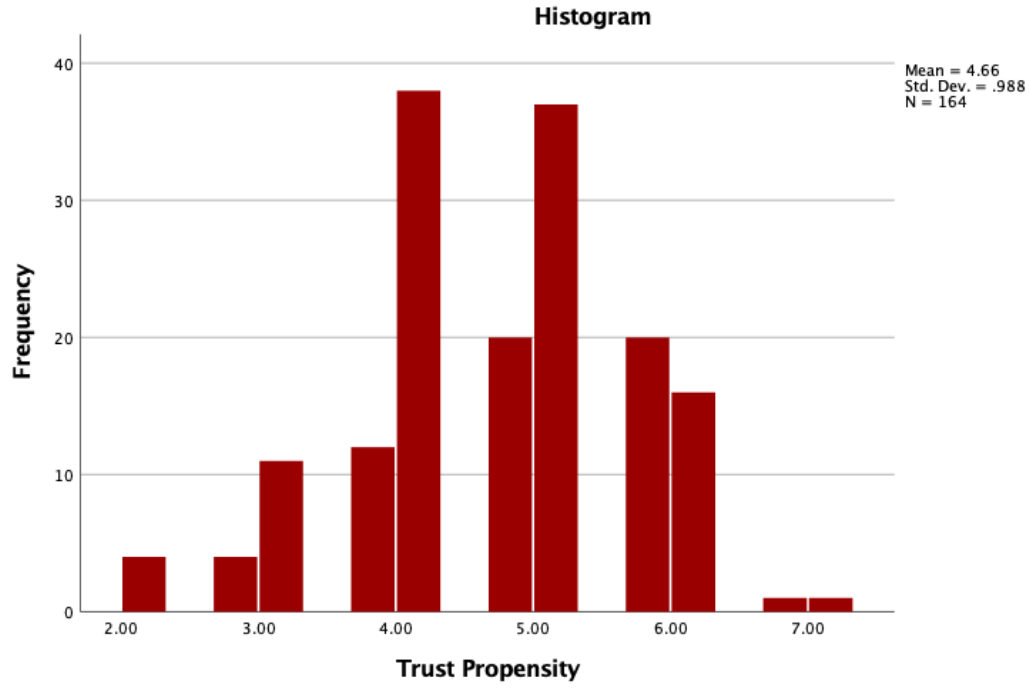
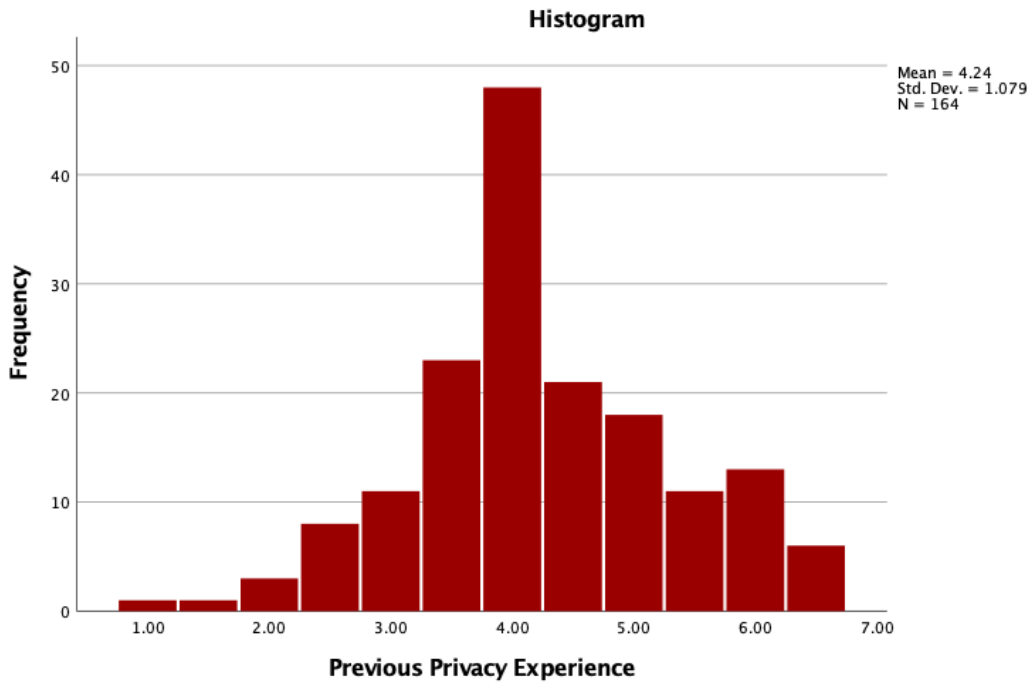


Figure 5

Histogram of the distribution of the variable Previous privacy experience



E: Chi-square tests

Table 1

Chi-square test of the variable Gender per group

	Value	Df	Asymptotic Significance (2-Sided)
Pearson Chi-Square	.018[a]	1	.892

a. 0 Cells (0.0%) Have Expected Count Less Than 5. The Minimum Expected Count Is 29.58.

Table 2

Chi-square test of the variable Education per group

	Value	Df	Asymptotic Significance (2-Sided)
Pearson Chi-Square	1.980[a]	5	.852

a. 4 Cells (33.3%) Have Expected Count Less Than 5. The Minimum Expected Count Is .47.

Table 3

Chi-square test of the variable Marital status per group

	Value	Df	Asymptotic Significance (2-Sided)
Pearson Chi-Square	2.022[a]	3	.568

a. 2 Cells (25.0%) Have Expected Count Less Than 5. The Minimum Expected Count Is 2.35.

Table 4

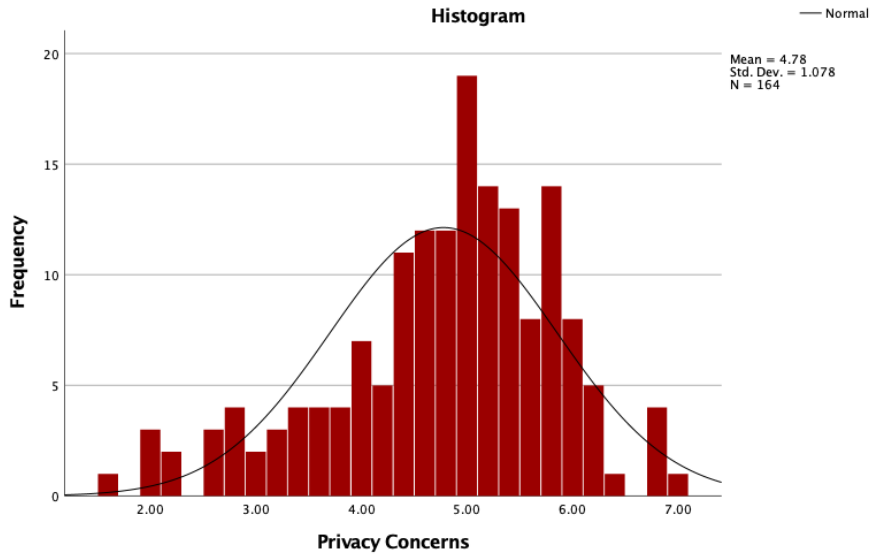
Chi-square test of the variable Employment status per group

	Value	Df	Asymptotic Significance (2-Sided)
Pearson Chi-Square	8.815[a]	6	.184

a. 8 Cells (57.1%) Have Expected Count Less Than 5. The Minimum Expected Count Is 1.88.

F: Normality tests for hypothesis 1 and 2

Hypothesis 1

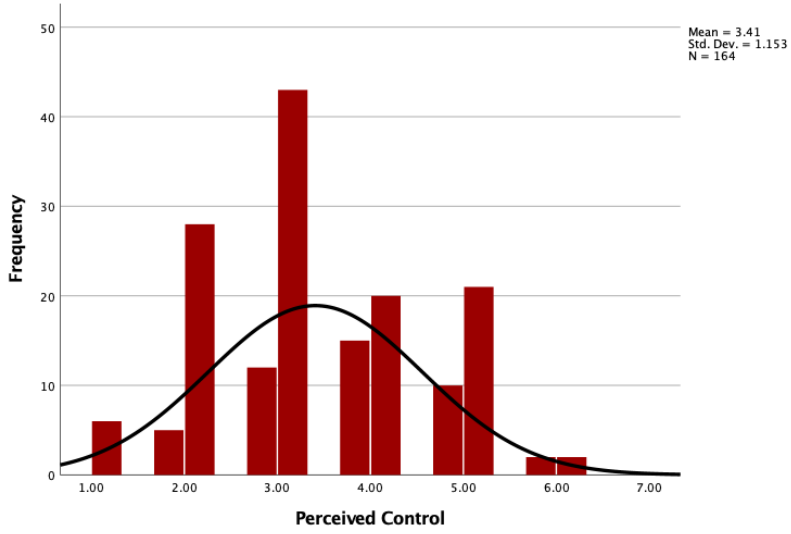


	Kolmogorov-Smirnov[a]			Shapiro-Wilk		
	Statistic	**Df**	**Sig.**	**Statistic**	**Df**	**Sig.**
Privacy Concerns	.112	164	.000***	.959	164	.000***

a. Lilliefors Significance Correction

***p < 0.01, **p < 0.05, *p < 0.10

Hypothesis 2



	Kolmogorov-Smirnov[a]			Shapiro-Wilk		
	Statistic	**Df**	**Sig.**	**Statistic**	**Df**	**Sig.**
Perceived Control	.100	164	.000***	.976	164	.007**

a. Lilliefors Significance Correction

***p < 0.01, **p < 0.05, *p < 0.10