Information is Power: The effect of the GDPR on digital privacy concern and trust in Google and Facebook

ERASMUS UNIVERSITY ROTTERDAM

ERASMUS SCHOOL OF ECONOMICS

Bachelor Thesis: International Bachelor of Economics and Business Economics

Name Student: Jimmy Farrell

Student ID number: 507227

Supervisor: Robert Dur

Second Assessor: Rubén Poblete Cazenave

Date final version: 1/07/2021

Abstract

This paper attempts to uncover the relationship between an individual's knowledge of the EU's General Data Protection Regulation (GDPR) and their resulting trust for data-collecting organisations and concern for digital privacy. This is interesting in the context of the composition of the regulation as anti-paternalistic and necessitating agency from the individual. Using an experimental survey from Germany, and building on prior research, I find weakly significant evidence that providing information regarding the GDPR can increase trust in specifically Google and not Facebook. Additionally I find robust evidence that the same provision of information regarding the GDPR increases digital privacy concern specifically for Facebook's commercially-purposed data collection practices. Finally, I find that awareness of the GDPR does not influence user account creation or deletion for either Google or Facebook.

Table of Contents

Introduction

The inability of regulation to keep up with advancements in technology has been an ever-present reality throughout history. The relatively recent rise of so called "Big Data" collection in both the private and public sector is no different. This logical progression of the information age has heralded significant increases in efficiency in the digital landscape. However, it is becoming clear that these benefits are coupled with a steadily increasing non-monetary cost of our individual privacy *(Puaschunder, 2020)*. Worldwide debate surrounding the issue of data privacy has increased amongst academics, concerned users and regulators themselves. However, as of yet little has been done by way of policy implementation intended to curb the increasingly pervasive invasions of privacy by companies relying most significantly on a business model of data collection.

Whilst exorbitant data collection is the future for almost all sectors of the economy, the most significant concern for individual privacy lies within companies and institutions that interact most frequently with their targets for data collection: their users. Thus, the current state of infancy of the information age has necessitated that the spotlight of data privacy concern shines most brightly on a few technology companies; specifically those that have been the first to adopt these practices en masse. The most salient of these companies are Google and Facebook *(Zuboff, 2019)*. These tech giants, amongst others, have found themselves increasingly scrutinized by the public and regulating bodies in recent years, with the CEO's of both companies appearing in front of the US Congress to discuss both matters of user privacy as well as anti-competitive business practises; issues that appear to be inexorably tangled in a causal loop *(Chiou & Tucker, 2017)*.

This perpetual circuit of data collection and unabated market conquering is facilitated by the emergence of an entirely new economic framework; one in which services such as Google Search and Facebook Newsfeed are free to use and revenues are sourced almost entirely from highly targeted and personalised advertising. In the case of Facebook for example, advertising made up 97% of revenue from the first quarter of 2021 *(Facebook, 2021)*. This has led to the infamous internet axiom: if something is free, *you* are the product. Whilst this simplification is not entirely accurate[1], it correctly suggests that participation in the digital milieu does not come without a cost. Our demographic, preferential and behavioural data, constantly manifesting from our internet activities, serves as raw material only accessible by a handful of companies who wield the sprawling capital infrastructure, economies of scale and network effects that are required to create complex user profiles for targeted advertising. This new market structure, in which buyers and sellers operate in a multi-billion dollar exchange of "behavioural futures", entirely

[1] Whilst our own bodies or minds are not the products of the market, increasingly accurate psychographic profiles of our personalities and interests are *(Zuboff, 2019)*.

removed from the users of the mentioned internet services, has been named "Surveillance Capitalism" *(Zuboff, 2019).*

The European Union's *General Data Protection Regulation* or GDPR of May 2018 marked the first meaningful government-level push back against this practice of unbounded mass personal data collection. Amongst numerous other protections, the regulation ensures the right to knowing who is processing your data, what data is being processed and why it is being processed. It also requires that organisations ask for consent to collect personal data; therefore giving the individual the agency to opt in or out. This is why websites operating on European IP addresses must now ask for permission before collecting non-essential digital cookies *(Dabrowski et al, 2019)*. The introduction of the GDPR appears to have caught the attention of regulators across the Atlantic with the *California Consumer Privacy Act* (CCPA) coming into effect on January 1st 2020 and the *Consumer Data Protection Act* (CDPA) of Virginia coming into effect in 2023. This is welcome news for concerned internet users in the United States; a country that has famously lagged behind the EU in the matter of broadly scoped data privacy regulation[2].

Although the promise of data privacy appears to be comprehensively outlined in the GDPR, the jury is still out regarding its practical efficacy. This is due to the largely unfamiliar and entirely unprecedented nature of the problem it is attempting to manage. This idea is supported by a large strand of literature which has attempted to uncover the vast informational asymmetries that exist between internet users and data collecting surveillance capitalists *(Bashir et al, 2015; Bartsch and Dienlinn, 2016; Epstein and Quinn, 2020; Livingstone et al, 2020; Pingo and Narayan, 2019)*. The relevance of this informational asymmetry becomes apparent when compared to the underlying mechanism through which an important part of the GDPR operates: informed consent *(Van Ooijen and Vrabec, 2018)*. Instead of a paternalistic data privacy protection policy - which would prohibit the collection and monetisation of personal data altogether - the GDPR represents a framework in which individuals must actively pursue their right to privacy by way of informed consent, an action that is significantly distorted by this asymmetry of information *(Van Ooijen and Vrabec, 2018)*.

Thus, the efficacy of the policy relies heavily on both the public's knowledge and hence concern of issues surrounding data privacy as well as the detailed rights guaranteed by the policy itself. The latter is of particular significance as it can be communicated completely and impartially. A 2019 Special Eurobaromoter conducted by the EU surveying over 27,000 randomly sampled Europeans found that only 36% had heard of the GDPR and understood what it was *(European Commission, 2019)*. Thus follows the central purpose of this paper: to investigate *whether the provision of detailed information*

[2] US data privacy law tends to be highly specific depending on the economic sector with for example separate regulations for the protection of health care data, consumer financial services data and the data of children. This segregated regulatory structure is in direct contrast to the encompassing nature of the GDPR *(Rustad and Koenig, 2019)*.

regarding the GDPR influences individuals trust and data privacy concern with regard to specific data-driven organisations and specific data-collection purposes. Subsequent additional questions of interest include whether GDPR awareness effects trends in user account creation and deletion for big tech platforms such as Google and Facebook.

To answer these questions I build on a recently published paper about whether knowledge of the GDPR increased trust in data-collecting organisations *(Bauer et al, 2021)*[3]. This paper made use of multiple waves of a non-probability online survey of over 2000 observations from Germany. The three waves, conducted in April, July and October 2018 respectively, allow empirical analysis regarding public opinion of data privacy in the context of the implementation of the GDPR (May 2018). The third wave consists of an experiment in which detailed information regarding the GDPR and its most important guarantees is provided randomly to the survey participants. This is followed by numerous questions regarding trust in Google and Facebook, as well as concern for research or commercially purposed data collection specific to Google, Facebook and the Federal Statistical Office (FSO). Further, questions regarding the ownership of accounts with various internet companies over time provide insight into whether the GDPR has had influence on actual behaviour (through account creation or deletion) rather than simply digital privacy concern. This is interesting in the context of the "privacy paradox" which describes the pervasive phenomenon in which individuals report high concern for digital privacy but paradoxically do very little to protect it *(Kokolakis, 2017)*.

I conceptualise concern in the realm of data privacy as the process of increasing vigilance towards a potential danger that is not fully understood. This increased vigilance in the form of concern is particularly important in the context of ensuring the maximum efficacy of the GDPR. As will be explained in further detail, the regulation relies on the user taking affirmative action, something that must be set in motion by concern for the danger of losing digital privacy. This concept of concern is operationalised by asking the respondents about their levels of comfort in various data collecting organisations sharing their personal data for first research and then for commercial purposes. It is assumed that an increase (decrease) in comfort reflects a decrease (increase) in concern and therefore a decreased likelihood of enacting the rights guaranteed by the GDPR. I hypothesise that showing an individual comprehensive information about the GDPR will increase their concern for data-privacy on average. This is because the regulation represents a blueprint for individuals to take back control of their personal data and no longer leave the issue of data-privacy in the hands of the data collectors

[3] Originally, this thesis was to be an original examination of the effect of GDPR knowledge on trust however this related paper was published during the writing process (23/05/2021). For the purposes of originality I will attempt to replicate the authors' results whilst also providing an original method regarding trust, as well as additional empirical analysis regarding data-privacy concern. This paper was not initially intended to be building off existing work however due to the unfortunate timing of the publication of this paper I decided to incorporate it into this thesis.

themselves. When individuals understand the risks and the accompanying rights they are entitled to to protect from those risks, they are more likely to be concerned about the issue and act on that concern.

I find that the provision of information regarding the GDPR increases trust in Google not to share data to third-parties, however these results are weakly significant. No effect is found for trust in Facebook. Additionally, I find that the same provision of GDPR information increases concern only for commercially-purposed data sharing specifically by Facebook to the magnitude of 7.6%. No effects are found in the case of Google or the Federal Statistics Office, or in the case of research-purposed data collection for all three data-collecting organisations. Finally, becoming aware of the GDPR was found to have no effects on account creation or deletion for either Google or Facebook.

Data privacy regulation is currently in its infancy, however, these issues will only increase in scope and complexity as the information age progresses. For this reason it is essential for regulators and academics to understand the importance of effectively communicating the GDPR to the European public to improve its effectiveness; thus cementing the social relevance of this paper. Simultaneously, the scientific relevance of this study arises from the originality of the experiment in the context of data-privacy concern. As mentioned previously a paper has been recently released concerning the outcome of trust *(Bauer et al, 2021)* however this paper will build on these results with a nuanced methodology. This will feature the use of lagged dependant variables as controls in order to account for possible omitted variable bias due to the unbalanced assignment of the experimental treatment. The only other publication to use the data from this survey makes use of experiments in waves 1 and 2 *(Bauer et al, 2019)*. Hence, this original empirical analysis into trust in data-collecting organisations and data-privacy concern is adding to strands of literature relating to both data-privacy concern in general *(Wirtz et al, 2007; Youn, 2009; Wu et al, 2012; Xie & Karan, 2019)* and the efficacy of the GDPR in the context of its anti-paternalistic composition *(Van Ooijen and Vrabec, 2018; Škrinjarić, 2019)*.

This paper will proceed with an overview of data privacy and the GDPR designed to protect it as well as a description of the data and setting. This will be followed by a brief review of the paper my research builds on, a revised method and critique of the original paper for the question of trust, and my own empirical analysis for the question of concern featuring a multiple linear regression. Finally I will make use of panel data to determine trends in account creation and deletion with regard to GDPR awareness. Each claim will be supported by robustness checks followed by a discussion of the results as well as policy implications and suggestions for further research.

Background

A New Marketplace

Before evaluating the efficacy of a government regulation, it is important to understand the issue at hand. Personal data-privacy in its modern form is an entirely unique societal concern thanks to its history being fairly recent. Unsurprisingly, the rise of this issue has coincided with the emergence of the internet as the platform through which our world connects.

Internet companies such as Facebook and Google primarily rely on a business model characterised by selling user data to advertising companies who are subsequently able to systematically segment their desired audiences with increasing precision. This market dynamic can be traced back to October 2000 when an emerging Google Inc. introduced personalised ads in the form of *AdWords*. This was following immense pressure from early investors to set up a revenue stream for the company amidst the chaotic dotcom bubble *(Zuboff, 2019)*. In this initial case, Google Search was the supply line of personal data in the form of keywords, phrasing, dwell times and click patterns. However, the introduction of Gmail in 2004 allowed Google to access the content of personal emails for the same purpose of targeted advertising *(Anandam et al, 2005)*. Similarly, Facebook extracts personal data explicitly through user-provided information such as demographics and the identities of friends. Newsfeed, the now recognisable home screen of an individual's Facebook experience since 2006, allowed Facebook to create complex personal profiling regarding implicitly provided information such as likes, posts, comments and ad clicks *(Andreou et al, 2018)*.

Additionally, in the case of Facebook, a controversial initiative launched in 2007 called Beacon was introduced as a feature of the company's advertising infrastructure. The program shared the behaviour and purchases of Facebook users on certain third-party websites with their friends without their consent *(Fletcher, 2010)*. It was thereafter scrapped in its original form due to a class-action lawsuit, however, it appeared to resurface under the name Facebook Connect only the year after *(Kane, 2010)*. This is the familiar program which allows users to log-in via Facebook on third-party websites giving Facebook the ability to track users across the internet,.

These revolutionary data-collection supply chains have only increased in both scale and scope, largely unchecked by regulators and supported by swift technical adjustments to quell public dissatisfaction when feathers are ruffled. Aside from the creation of an external market of user-prediction products traded between internet companies and advertisers, the immense data-collection allows the companies to personalize the digital experience to increase overall engagement. This combination of a highly personalized experience as well as significant network effects is considered to be the driving force behind the monopolistic market shares that these companies enjoy *(von Briel and Davidson, 2019)*.

Increased market share allows increased investment into data-mining infrastructure which leads to more market share, ever concentrating informational power in the hands of the few. This fundamental feedback loop in part forms the basis for modern data-privacy concern. Individuals are not economically compensated for the exchange in value between tech companies and advertisers, made possible only by their personal information *(Zuboff, 2019)*. Data-collection and privacy, in relation to targeted advertising, has typically raised concerns ranging from reducing consumer autonomy to predatory manipulation of psychological weaknesses and insecurities. These concerns are not new for advertising in general, however, the depth of access to our personal information has amplified these issues *(Susser et al, 2019)*.

Mainstream data-privacy concern shifted to the realm of democracy with the Cambridge Analytica scandal of 2016 that threatened to bring Facebook to its knees. Through the previously mentioned Facebook Connect feature, a political consultancy firm hired by the Trump campaign was able to create psychographic profiling of over 50 million Americans. This information was then used to deliver targeted, inflammatory campaigning to swing-voters in order to potentially influence the 2016 presidential election *(Issak & Hanna, 2018)*. The resulting major public backlash elevated concern regarding data-privacy to new heights, as once futuristic dangers, such as manipulation of democracy, began to crystalize.

Debate about the *future* of data privacy is primarily inhabited by concern for a growing tech phenomena labelled the "Internet of Things" (IoT). The IoT refers to a network of physical objects that constantly share data in the form of internet-connected "smart" devices. This has serious implications for digital-privacy as information will potentially be trawled not only from our internet activity, but also from our bodies in the form of wearable technology, and our immediate environments in the form of "smart" homes and cities *(Wachter, 2018)*. Despite such technology being in its early stages, major IoT programs such as the "smart" city Google Sidewalk Labs have already been rejected due to privacy concerns amongst other difficulties *(Peel and Tretter, 2019)*.

As such, data privacy is an emerging social issue that is difficult to comprehend due to informational asymmetry by design. Whilst mass data collection has brought significant increases in efficiency, manipulation of both consumers and democracy ensure that it remains unclear that sacrificing personal privacy is the price we should be willing to pay.

The General Data Protection Regulation

The progression of data-privacy over the past few decades as an existential threat to human autonomy has spawned multiple regulation efforts, the most comprehensive of which is the EU's *General Data Protection Regulation* (GDPR). Completed in 2016 and fully implemented in 2018, it replaced the by-then outdated *EU Data Protection Directive* of 1995. Some coverage changes included increasing the geographical scope so that the data of individuals who reside in European Union member states are protected regardless of where the company operates. The change also increased the minimum age of person subject to data collection from 13 to 16. Structural changes include each nation having a single office responsible for receiving complaints and large data-controllers requiring to appoint a data-protection officer. Fines can be up to 20 million euros or 4% of global annual turnover from the most recent financial year *(European Union, 2016)*.

The main changes regarding data-collection itself revolve around increased transparency and communication between data-collectors and users of internet services. Firstly, detailed information must be readily available about the type of data that is collected and what it is used for. Additionally, individuals must give consent for data-processing through so-called affirmative action also known as "opt-in" consent *(Maldoff, 2016)*. This is why most websites feature banners with the option to either accept or reject cookies[4]. Finally, the regulation ensures the "right to be forgotten", by having your data deleted upon request or if the data is not being used as it was originally purposed *(European Union, 2016)*.

Whilst this introduction of "opt-in" consent gives individuals the personal agency to protect their data in theory, many argue that in practice, informational asymmetry between data-collectors and data-subjects as well as a lack of knowledge about the GDPR discredits the validity of this consent. Article 4 of the GDPR defines consent as 'freely given, specific, informed and unambiguous' *(European, Union, 2016)*, however several experimental studies have shown that patterns of consent regarding data-collection on websites are substantially influenced by convenience *(Utz et al, 2019; Nouwens et al, 2020)*. It can thus be argued that consent is an inappropriate mechanism to protect data-privacy due to the malleability and ignorance of data-subjects. To increase the protection of the personal data of individuals in the information age, many have argued for a more paternalistic regulation in order to compensate for the informational gap *(Reviglio, 2019; Shahizam, 2021)*.

[4] Cookies refer to files used to store information about a user's activity on a website.

Data and Setting

To provide insight into my research question I have made use of an online panel survey operated by the market research company Respondi. The survey featured individuals living in Germany and over the age of 18. It consists of three waves all from the year 2018, with the first wave being conducted between the 14th and 22nd of April, the second between July 24th and August 2nd and the third between October 29th and November11th. Thus, the GDPR was officially enacted in between waves 1 and 2 on May 25th of the same year. Although the sampling was non-probabilistic as completion of the survey is both voluntary and monetarily incentivised, each sample was trimmed to fit known national quotas of gender, age and smart-phone ownership in order to replicate random sampling. Each wave was open to new respondents and as such the sample did not remain constant over the three waves. The resulting sample sizes were 2095, 2046 and 2117 for each wave respectively with a total of 1269 respondents participating in all three waves.

Whilst this sample has been trimmed to fit various aggregate characteristics of the German population, it must be noted that attitudes towards digital-privacy are generally stronger in Germany compared to other parts of the world. This has been shown by comparing the average willingness to pay for privacy in separate categories of data, from health and credit card information to web browsing behaviour. Germany was found to have a significantly higher average willingness to pay than the UK, US, China and India for most categories *(Morey et al, 2015)*. This is supported by Germany's historically harsh stance towards Google Street View, which was suspended in 2011 due to privacy concerns *(Geissler, 2011)*. Thus, findings from this sample cannot be generalised to the global population in which digital-privacy concern differs greatly from country to country.

GDPR on Trust in Facebook and Google: Building on Prior Research

A paper, published on the 23rd of May 2021 by researchers from the University of Mannheim, University of Maryland and University of California investigated the effect of awareness of the GDPR on individuals trust in data collectors using both observational and experimental data *(Bauer et al, 2021)*. Using the same survey as mentioned above and amongst other empirical objectives, the researchers attempted to find a causal relationship between providing detailed information regarding the GDPR and an individual's self-reported trust in specifically Google and Facebook.

The observational portion of the empirical analysis utilised waves 1 and 2 of the survey and found no evidence that trust in data-collecting organisations is higher among those who are or became aware of the GDPR over the course of the survey. This was the result of first a cross-sectional linear regression model of wave 1 that used a dummy for GDPR awareness as independent variable and multiple questions relating to trust in specific organisations as outcome variables. This was followed by a cross-sectional linear regression constructed with both waves 1 and 2. It used a dummy for whether an individual became aware of the GDPR in between the waves or not as independent variable and the difference in reported trust in the same outcome questions as the dependent variable. Each model included controls for socio-demographics (age, sex and education) as well as Google and Facebook account ownership, general trust, general privacy concern and device ownership (Smart-phone, non-smartphone, PC, tablet and Ebook) which has been shown to be a reliable predictor of digital literacy *(Hargittai et al, 2019)*. Balance tests performed on the mentioned controls revealed unbalanced assignment of treatment with regards to both sex and PC ownership (Appendix C & D). These controls are therefore included in later sections. Ultimately, the results of these regressions revealed no significant effect of GDPR awareness on trust in the aforementioned data collecting organisations. Using the same data and replicating the same method, I found the same results.

The experiment used to uncover this relationship was featured in the third wave of the survey for which data was collected in October and November of 2018. The experiment divided the sample randomly into three groups with 50% in the treatment group and 25% each in separate control groups. In the second half of the survey, the treatment group received a half-page of information regarding the main rights afforded to citizens by the GDPR (Appendix A). Briefly, this included the right to information on who processes your data, what data is processed and for what purpose is it processed, as well as the right to have your data moved between organisations or "forgotten" at your discretion. The information also included a description of the "opt-in" consent mechanism through which data processing must occur under the GDPR, as explained previously. Finally, the text included a stipulation of the responsibility data-collecting organisations have to their users in the event of security breaches. The

first control group received a similar length text about airline passenger rights. This was chosen as a placebo control that had no relation to the outcome variables. The final 25% of the respondents received no treatment.

The outcome variable for the linear regression model that followed was the difference between responses to identical questions regarding trust in Google and Facebook that were asked before and after the experiment. These questions appeared as follows:

- On a scale from 0 "not at all" to 10 "completely", how much do you trust that **Google** uses your personal data only internally, so does not share them with third parties?
- On a scale from 0 "not at all" to 10 "completely", how much do you trust that **Facebook** uses your personal data only internally, so does not share them with third parties?

These linear regressions featured the same controls as the observational empirical analysis with the addition of controlling for GDPR awareness and reducing the sample to only those who reported the information they were provided as comprehensible. In accordance with the experimental results, the authors found no evidence that providing additional information about the GDPR effected individual trust in data-collecting organisations. I also replicated this experiment and found the same results.

However, I believe this method could be improved by removing the potential for experimenter demand effects. This refers to a situation in which respondents subject to an experiment change their behaviour towards what they perceive to be appropriate based on cues from the experiment itself *(Zizzo, 2010)*. In this case, an identical question appearing twice in the same wave of a survey, before and after the experiment, may bias the response. This could happen either through the respondent being less willing to change his previous answer in order to not seem indecisive, or more willing to change his previous answer to conform to the perceived purpose of the repeated question. Thus, the methodology may cause bias in the results.

A solution to this problem could be to only ask the questions regarding trust in Google and Facebook after the experiment so that the answers are not prone to experimenter demand effects. Instead of recording the difference in trust pre and post-experiment, the outcome variable of this regression would simply be the raw answer to the question post-experiment. In this scenario the questions appear for the first time already having been influenced by the experiment for the treatment group and not for the control group. This would ensure unbiased random assignment of the treatment and would potentially result in a causal effect as the treatment and control groups would only differ on aggregate by exposure to the experiment itself, mitigating the potential for omitted variable bias. Thus, simply using the second responses would not remove the potential for experimenter demand effects, as the second responses are already biased.

Additionally, a second problem with this method arises due to the unbalanced nature of the questions regarding trust pre-experiment. Balance tests for these lagged dependant variables reveal that random assignment of treatment was not achieved in the case of individuals trust for Facebook pre-experiment, with a significant coefficient at the 10% level. This was not found for trust in Google. This unbalance in the assignment of treatment through the variable of pre-treatment trust could cause omitted variable bias. This could erroneously influence the results of the linear regression model that uses the difference between pre and post experiment trust as outcome. The results of these balance tests can be found in appendix B.

In order to account for this, an alternative method is proposed. This will use the lagged pre-treatment dependant variable as a control in a linear regression model that uses the post-treatment response as the outcome variable. As mentioned previously, this does not solve the problem of experimenter demand effects, however it corrects for the potential omitted variable bias caused by the imbalance in treatment assignment. Thus, I construct a linear regression with the following properties:

$$y_{i,t} = \beta_0 + \beta_1 T_i + \beta_2 y_{i,t-1} + \varepsilon_i$$

In this case, $y_{i,t}$ represents the outcome variable of trust in either Google or Facebook, measured after the experiment at time t. The questions are answered on an 11 point scale with 10 representing "completely" and 0 representing "not at all". T_i represents a dummy for treatment which takes the value of 1 if the individual received the GDPR information and a 0 if they did not. β_1 represents the coefficient of the effect of treatment on the outcome variable. In other words, by how much will trust change on average if an individual receives information about the GDPR. $y_{i,t-1}$ represents the relevant lagged dependant variable and thus the responses to the trust questions regarding Google and Facebook respectively, at time $t-1$ (pre-treatment). A quadratic of the lagged dependant variable is also used as a control to establish the possibility of a non-linear relationship. β_2 represents the coefficient of the effect of the lagged dependant variable on the outcome variable. β_0 represents the constant and hence the average value of the outcome if treatment equals 0. Finally, ε_i represents the error term and as such the distance between the observed outcome and the predicted outcome.

Table 1 – TRUST IN GOOGLE AND FACEBOOK

	Trust in Google t			Trust in Facebook t		
	1	2	3	1	2	3
GDPR Treatment	0.114	0.083*	0.081*	0.257**	0.069	0.066
	(0.122)	(0.048)	(0.048)	(0.118)	(0.046)	(0.046)
Trust in Google/Facebook $t-1$		0.923***	1.014***		0.916***	1.066***
		(0.009)	(0.026)		(0.011)	(0.029)
(Trust in Google/Facebook)^2 $t-1$			-0.011***			-0.020***
			(0.003)			(0.005)
Constant	3.681	0.277	0.183	2.573	0.188	0.083
	(0.086)	(0.043)	(0.047)	(0.082)	(0.034)	(0.030)
Observations	2110	2101	2101	2110	2107	2107

Notes: This table displays results from six linear regression models featuring the experimental treatment and the outcome variables of trust in Google and Facebook to not share data with third-parties. Model 1 of each organisations features the treatment coefficient without controls, model 2 includes the control of the lagged dependant variable (Trust in Google or Facebook pre-experiment), and model 3 includes both the lagged dependant variable and the quadratic lagged dependant variable. Robust standard errors for the coefficients appear in parentheses. * represents significance of the coefficient at 10%, ** significance at 5% and *** significance at 1%.

Table 1 displays the results of the mentioned regression models. The only result significant at a 5% level is that of Facebook without controlling for the lagged dependant variable. As explained previously, this significance is caused by an unbalance in treatment assignment and can therefore not be interpreted. Other than this, significant effects at the 10% level are found for Google in both models that include the lagged dependant variable. Thus it can be argued that providing an individual with detailed information about the GDPR can possibly increase an individual's trust in the case of Google however the weak significance renders these inferences inconclusive. No such results are found for Facebook.

Unsurprisingly, the coefficients for the linear lagged dependant variables for both data-collecting organisations are significant at 1% and relatively close to 1 indicating that previous measures of this outcome variable are predictive of future measures. Additionally, the coefficients are smaller than 1, indicating slight mean reversion in which relatively low values of the lagged dependant move upwards towards the mean on average and relatively high values move down on average. This is reflected by the spread of the variable which decreases slightly from lagged dependant to dependant. In the case of Facebook, the large decrease in the coefficient for the variable of interest from 0.257 in model 1 to 0.069 in model 2 indicates that the inclusion of the lagged dependant variable somewhat accounted for the imbalance in treatment assignment and hence potentially removed confounding unobserved effects.

GDPR on Digital Privacy Concern

Descriptive Statistics

An important feature of the GDPR is the anti-paternalistic mechanism through which individuals must take steps themselves to protect their personal data, rather than this being automatically done. Thus, the regulation can only be put into effect if an individual's digital privacy concern is significant enough to produce the aforementioned agency that is required. For this reason, descriptive statistics are useful for gaining an understanding of how digital privacy concerns differ between certain data collecting organisations and between different data collecting purposes.

In order to measure this data privacy concern specific to organisation and data collecting purpose, I use responses to six different questions regarding comfort for data collecting practices. Comfort is then operationalised as the opposite of concern as justified previously. Each respondent is only asked one of the six questions resulting in a sample size of approximately 350 respondents per model. This is called random selection and is used for the purpose of minimising the length of the survey. The six questions are as follows:

- How comfortable or uncomfortable would you feel if **Google** would share your data for research purposes (e.g., with university researchers)?
- How comfortable or uncomfortable would you feel if **Facebook** would share your data for research purposes (e.g., with university researchers)?
- How comfortable or uncomfortable would you feel if the **Federal Statistical Office** would share your data for research purposes (e.g., with university researchers)
- How comfortable or uncomfortable would you feel if **Google** would share your data for commercial purposes (e.g., with advertising or marketing companies)?
- How comfortable or uncomfortable would you feel if **Facebook** would share your data for commercial purposes (e.g., with advertising or marketing companies)?
- How comfortable or uncomfortable would you feel if the **Federal Statistical Office** would share your data for commercial purposes (e.g., with advertising or marketing companies)?

Each question is answered on a 5 point scale where 1 represents "very uncomfortable" and 5 represents "very comfortable". Table 2 displays the descriptive statistics for responses to each of the six questions.

Table 2 – DESCRIPTIVE STATISTICS

	Research-Purposed			Commercially-Purposed		
	Google	Facebook	Federal Statistics Office	Google	Facebook	Federal Statistics Office
Mean	2.34	2.1	2.57	1.97	1.82	1.88
Std. Dev.	1.02	0.98	1.03	0.96	0.95	1.03
Observations	370	382	322	343	355	341

Notes: This table displays descriptive statistics for responses relating to how comfortable an individual would be with personal data being shared about them on a 5 point scale with 5 representing total comfort. This question is divided into the three different data-collecting organisations of Google, Facebook and the Federal Statistics Office and two different data-collecting purposes: research and commercial. For each category the means, standard deviations and sample number are shown.

The most obvious trend from this descriptive analysis is that individuals appear to be more concerned about data collection for commercial purposes than for research purposes. This is reflected by the higher mean levels of comfort in research-purposed data collection for all three data-collecting organisations. The level of comfort for research-purposed data sharing is highest for the Federal Statistics Office (FSO). This could reflect a status quo bias where traditionally researchers and universities are more closely connected to the public sector making this a more comforting source of personal data sharing. This could also be explained by individuals believing that data held by the FSO, such as demographics or marital-status, may be less personal and therefore less privacy-violating than information held by Google or Facebook in the form of specific searches or personality traits. This higher comfort in the public sector is not mirrored in the case of commercially-purposed data sharing however. In this case, the idea of a Government authority pursuing commercial interests is clearly sub-optimal. Finally, in both research and commercial cases, concern for digital privacy seems to be higher in the case of Facebook compared to Google. This could be explained by a recently higher negative media portrayal of Facebook in the realm of data privacy, most saliently in the context of the 2016 US presidential election through the Cambridge Analytica scandal and alleged Russian interference *(Zuboff, 2019)*. The spread of the outcome variables are fairly uniform across all three organisation and both data collection purposes.

Effect of GDPR knowledge on concern

As mentioned previously, the efficacy of the GDPR depends on whether individuals are sufficiently concerned about their data privacy enough to actively make use of the rights that are anti-paternalistically at their discretion. This would most frequently manifest in the form of optimising cookie settings on websites rather than simply consenting to default settings. This leads to the question

of whether providing additional information regarding the GDPR increases concern in data-collecting organisations, in turn potentially increasing the efficacy of the policy.

Using the same experiment as discussed previously as treatment, the outcome variable will now be digital privacy concern (operationalised through comfort) rather than trust. This outcome variable takes the form of six different questions with two for each data-collecting organisation (Google, Facebook and the FSO). One of the two regards data-collection for research purposes and the other for commercial purposes, as explained previously.

A linear regression model will be used as follows:

$$y_i = \beta_0 + \beta_1 T_i + \varepsilon_i$$

Here the outcome variable y_i represents the answer to one of the questions out of the six, on a 5 point scale from "very uncomfortable" taking the value of 0, to "very comfortable" taking the value of 4. Unlike the questions for trust, these outcome variables were only asked once (post-experiment), hence removing the possibility to take differences and also reducing the potential for experimenter demand effects. The term T_i represents a dummy variable for the experiment which takes the value 1 if the individual received the passage of information regarding the GDPR and 0 if the respondent did not. Accordingly, the term β_1 represents the estimated coefficient of the effect of treatment on the outcome variable y_i, in this case the effect of providing detailed information regarding GDPR on digital privacy concern. The term β_0 represents the constant of the binary linear model and hence the predicted value of y_i if the respondent did not receive treatment ($T_i = 0$). Finally, ε_i represents the error term of the linear regression; the distance between the observed value and the predicted value for the given treatment. Controls for sex and PC ownership were added to ensure that random differences between control and treatment group were controlled for. Additionally, controls for general privacy concern and general trust as well as age were added to further increase the accuracy of the estimates.

Table 3 displays the results of six separate linear regression models regarding the effect of the experiment on an individual's concern for their personal data being collected for research purposes. Each data collecting organisation features two models; one without and one with controls. As none of the coefficients for the variable of interest are significant it can be said that on average, giving an individual in this sample extra information regarding the GDPR has no effect on their level of concern for data privacy with regard to research purposed data-collection. All point estimates are insignificant even at a level of 10%. This is indicated by the large robust standard errors that necessitate confidence intervals stretching over the threshold of zero and thus resulting in significant estimates. The effect of general privacy concern is significant at a 1% level for all three organisations, with the coefficient being negative due to the operationalisation of concern via comfort. This displays the unsurprising possibility

that general concern for privacy is positively related to and a good predictor of digital privacy concern specific to each organisation. Additionally, general trust seems to be a good predictor of digital privacy concern only in the context of the FSO, possibly due to self-reported general trust being more tied to opinions regarding government rather than the private sector.

Table 3 – RESEARCH-PURPOSED DATA COLLECTION

	Comfort Google	Comfort Google	Comfort Facebook	Comfort Facebook	Comfort FSO	Comfort FSO
GDPR Treatment	-0.031	-0.047	-0.025	-0.018	0.035	0.001
	(0.106)	(0.101)	(0.100)	(0.098)	(0.115)	(0.112)
Sex		-0.164		-0.034		-0.026
		(0.104)		(0.101)		(.104)
PC Ownership		0.033		0.456		-0.296
		(0.301)		0.250		0.339
Age		-0.005		0.001		-0.008
		(0.004)		(0.004)		0.004
General Privacy		-0.400***		-0.265***		-0.318***
Concern		(0.083)		(0.076)		0.086
General Trust		0.026		0.032		0.049**
		(0.023)		(.021)		(0.023)
Constant	2.353	3.569	2.109	2.141	2.553	3.898
	(0.079)	(0.473)	(0.066)	(0.405)	(0.088)	(0.490)
Observations	370	363	381	379	320	319

Notes: This table displays results from six linear regression models featuring the experimental treatment and the outcome variables of comfort for research-purposed data collection for specific data collecting organisations. This table features three of the six outcome variables used in this section of the empirical analysis with each respondent only being giving one, hence the sample size for each model is roughly the total sample (2117) divided by six. The models with controls feature less as some respondents left no answer to PC ownership. Each data collecting organisation has a model with and without control variables. Robust standard errors for the coefficients appear in parentheses. * represents significance of the coefficient at 10%, ** significance at 5% and *** significance at 1%.

Table 4 displays the same format as table 3, however this time referring to concern for commercially-purposed data collection. In contrast to table 3, the coefficient of interest is significant in the case of Facebook for both the model without controls and the model with controls. Without controls the effect is significant at a 10% level whilst with controls the effect is significant at a 5% level. The latter can be interpreted as a decrease of 0.2 points on a 5 point scale of comfort regarding commercially-purposed data collection on average when an individual is shown detailed information about the GDPR, all else being equal. This 0.2 decrease translates to a 7.6% increase in concern for Facebook's data-collecting practices specifically dedicated towards increasing ad revenue. This is relative to the case in which GDPR information is not given, represented by the constant. Significant effects were not found for the same question regarding Google or the FSO.

Similar results are found regarding general privacy concern as a good predictor for digital privacy concern, this time in the context of commercially-purposed data collection. All three coefficients are significant and negative indicating a positive relationship between general privacy concern and organisation-specific digital privacy concern. Additionally, general trust is found to have a significant effect in the case of Google. The general trust effect for the FSO is weakened to being significant at only 10%, possibly due to individuals associating commercially-purposed data collection less with the public sector. Finally, an interesting finding is that in the context of commercially-purposed data collection, age emerges as an important predictor of digital privacy concern. For all three organisations the coefficient is negative and significant at 1%, implying that as age increases, comfort decreases and hence concern increases. This reveals that older generations are on average more adverse to the central mechanism on which so called 'surveillance capitalism' is built. This is possibly due to younger generations having been raised closer to the information age and hence having more relative life exposure to this new market reality. Another potential explanation could be that younger generations purchase more consumer goods online and therefore enjoy the benefits from increased ad personalisation more so than their older counterparts.

Table 4 – COMMERCIALLY-PURPOSED DATA COLLECTION

	Comfort Google	Comfort Google	Comfort Facebook	Comfort Facebook	Comfort FSO	Comfort FSO
GDPR Treatment	0.068	0.075	-0.176*	-0.200**	0.002	-0.049
	(0.106)	(0.105)	(0.102)	(0.098)	(0.112)	(0.106)
Sex		-0.163		-0.120		-0.129
		(0.111)		(0.097)		(0.107)
PC Ownership		0.098		0.210		0.147
		(0.381)		(0.279)		(0.357)
Age		-0.012***		-0.015***		-0.012***
		(0.004)		(0.004)		(0.004)
General Privacy Concern		-0.192**		-0.160**		-0.402***
		(0.087)		(0.065)		(0.079)
General Trust		-0.005		0.047**		-0.040*
		(0.022)		(0.022)		(0.022)
Constant	1.939	3.015	1.915	2.641	1.879	3.624
	(0.070)	(0.526)	(0.078)	(0.396)	(0.078)	(0.467)
Observations	343	339	355	354	341	339

Notes: This table displays results from six linear regression models featuring the experimental treatment and the outcome variables of comfort for commercially-purposed data collection for specific data collecting organisations. This table features the remaining three of the six outcome variables used in this section of the empirical analysis with each respondent only being giving one, hence the sample size for each model is roughly the total sample (2,117) divided by six. The models with controls feature less as some respondents left no answer to PC ownership. Each data collecting organisation has a model with and without control variables. Robust standard errors for the coefficients appear in parentheses. * represents significance of the coefficient at 10%, ** significance at 5% and *** significance at 1%.

Account Creation and Deletion

A reoccurring concept in the growing academic field of digital privacy and ethical adaptations to the information age is that of the *privacy paradox*. This describes a general trend of inconsistency between the intentions of internet users to protect their personal data and their actual behaviour when it comes to privacy. A vast catalogue of empirical academic research has documented this phenomenon through both observational and experimental studies *(Brown 2001; Norberg et al, 2007; Young & Quan-Haase, 2013)* whilst more recent papers observe that this paradox is becoming less common as digital decisions are becoming more informed *(Dienlin & Trepte, 2015; Solove, 2021)*. Thus, in tandem with investigating the effect of GDPR awareness on trust in data-collecting organisations and digital privacy concern, an investigation into the effect of GDPR awareness on actual behaviour is justified.

This will be achieved through investigating whether there is an effect between becoming aware of the GDPR during a certain period of time and creating or deleting either Google or Facebook accounts during the same period. Social media and internet account deletion has been shown to primarily reflect privacy concerns as the services are usually free and thus keeping an account does not decrease utility through subscription fees or other monetary mechanisms *(Baumer et al, 2013; Stieger et al, 2013)*. Thus, account deletion can be operationalised as digital privacy concern reflected in actual behaviour. Account creation on the other hand does not directly reflect an absence of digital privacy concern as internet accounts can provide many benefits such as email in the case of Google's Gmail and a social media presence in the case of Facebook.

The same survey is used to investigate this question with all three waves (April, July and October of 2018) together resulting in a final total sample size of 931. A dummy variable for GDPR awareness during the time of the survey takes the value of 1 if an individual became aware of the regulation over the course of the survey. The variable takes the value of 0 if the individual was either already aware of the GDPR in wave 1 or still not aware of the GDPR by wave 3. Contradictory observations were removed for example if an individual reported being aware of the GDPR in wave 2 but not in wave 3.

Dummy variables are also constructed for those who reported to have deleted Google or Facebook accounts over the course of the survey, and those who reported to have created Google or Facebook accounts over the survey. Each wave of the survey contained a question of whether an account with a particular data collecting organisation was owned, with 1 representing that an account was owned, 2 representing that an account has never been owned and 3 representing that an account was once owned but is now no longer. For both Google and Facebook, separate account creation dummies were created where 1 represented an individual who reported not having an account in wave 1 and having an account in wave 3, and 0 represented any other case. In a similar way separate account deletion dummies were created where 1 represented an individual who reported having an account in wave 1 and not having an

account in wave 3. Similarly to the GDPR awareness variable, the observations were screened for any contradictory outcomes; such as an individual reporting to have never had an account after they had reported to have had an account in a previous wave. These observations were removed resulting in a sample size of 931 from an original merged sample of 1269.

The following linear regression model was used:

$$y_i = \beta_0 + \beta_1 T_i + \varepsilon_i$$

In the same structure as the preceding empirical analysis, y_i represents the outcome dummy variable of whether an account was created or deleted during the timeframe of the survey. T_i represents the independent variable of interest which in this case is whether an individual became aware of the GDPR during the course of the survey. β_1 represents the coefficient of the effect of this treatment on the outcome variable; in other words how much does becoming aware of the GDPR during this survey increase the likelihood of creating or deleting an account with Google or Facebook also during the survey. β_0 represents the constant of the linear regression model and hence the likelihood of creating or deleting an account over the course of the survey if an individual does not become aware of the GDPR during the survey. ε_i represents the error of the model or the distance between the observed outcome and the predicted outcome. Controls are added for sex, age, and averages for general privacy concern and general trust over all three waves of the survey.

Tables 5 and 6 show the results of the linear regression models. The insignificant coefficients for both account creation and account deletion indicate no relationship between becoming aware of the GDPR and creating or deleting accounts with either Google or Facebook. The only significant coefficient is age at a 10% level for both account creation and deletion only in the case of Google, however these values are miniscule and thus economically insignificant. Given the results of the empirical investigation into the effect of GDPR information on concern, it can be argued that these results reflect the privacy paradox in action. In that case, GDPR information was shown to effect digital privacy concern with regard to Facebook, thus GDPR awareness should effect privacy protection related behaviour such as Facebook account deletion through increased privacy concern. This final link however is missing as individuals do not always reflect their privacy concerns in their behaviour, as the paradox suggests. This argumentation has many problems however, as in the two examples the conceptualisation of GDPR awareness was significantly different. This difference being an experimental design with detailed GDPR information in the first case, and simply a measure of whether an individual became aware of the GDPR during the survey in this case. The measure for GDPR awareness in these regressions regarding account creation and deletion does not take into account the type of information the individual received regarding the GDPR or the context in which they received it. Additionally, if a result was to have been found in this section of the analysis, causality would be

difficult to prove. This is because the treatment variable is not randomly selected leaving the potential for cofounders that influence both treatment and outcome, absent of a causal link between the two.

Table 5 – ACCOUNT CREATION

	Google Account Creation		Facebook Account Creation	
Became aware of GDPR during survey	0.008 (0.019)	0.008 (0.020)	0.003 (0.008)	0.003 (0.009)
Sex		0.001 (0.020)		-0.009 (0.009)
Age		0.001* (0.001)		0.000 (0.000)
General Privacy Concern		0.005 (0.016)		-0.013 (0.008)
General Trust		0.004 (0.004)		0.001 (0.002)
Constant	0.088 (0.013)	-0.019 (0.075)	0.015 (0.006)	0.048 (0.035)
Observations	928	920	928	920

Notes: This table displays results from four linear regression models featuring the dummy treatment of whether an individual became aware of the GDPR during the survey and the outcome dummy variable of whether an individual created an account in either Google or Facebook also over the time of the survey. Two models are presented for each organisation; one with and one without controls. Robust standard errors for the coefficients appear in parentheses. * represents significance of the coefficient at 10%, ** significance at 5% and *** significance at 1%.

Table 6 – ACCOUNT DELETION

	Google Account Deletion		Facebook Account Deletion	
Became aware of GDPR during survey	-0.006 (0.008)	-0.006 (0.008)	0.005 (0.007)	0.005 (0.007)
Sex		0.005 (0.008)		-0.004 (0.008)
Age		0.000* (0.000)		0.000 (0.000)
General Privacy Concern		0.001 (0.007)		-0.009 (0.009)
General Trust		-0.001 (0.002)		0.000 (0.002)
Constant	-0.006 (0.008)	-0.010 (0.028)	0.010 (0.005)	0.037 (0.035)
Observations	928	920	928	920

Notes: This table displays results from four linear regression models featuring the dummy treatment of whether an individual became aware of the GDPR during the survey and the outcome dummy variable of whether an individual deleted an account in either Google or Facebook also over the time of the survey. Two models are presented for each organisation; one with and one without controls. Robust standard errors for the coefficients appear in parentheses. * represents significance of the coefficient at 10%, ** significance at 5% and *** significance at 1%.

Discussion

The results of the preceding empirical analyses raise many additional questions and reasons for further discussion. It is essential for governments and regulatory bodies to understand the effect that new regulations have on the populous, both in terms of their attitudes and their behaviours. In the case of anti-paternalistic policies such as the GDPR, the regulations' efficacy relies on the ability of individuals not only to understand the problem it intends to improve but to transform that understanding into action; by actively refusing cookies or requesting transparent information from data-collecting organisations.

The first section of the empirical analysis built on a recent paper which used the same survey to investigate whether the GDPR has had an impact on individuals trust with data-collecting organisations *(Bauer et al, 2021)*. This paper used an experiment which provided a random half of the survey with detailed information about the GDPR, and then used linear regression models to measure the effect on differences between trust for Google and Facebook before and after the experiment. Whilst the resulting coefficients were insignificant, further investigation of the data in my analysis revealed that a modified method controlling for the lagged dependant variable revealed weakly significant results. This method was performed in order to mitigate the chance of omitted variable bias through controlling for unbalances in the assignment of the treatment with regard to pre-experiment trust in Facebook. These weakly significant results suggested that providing individuals with information about the GDPR may increase trust in Google but not Facebook.

The mechanism through which knowledge of a new regulation should increase trust in the agents it serves to hold responsible is fairly straight forward. In this case, the regulation itself, and consequently the information about it, could serve as a potential diffuser of growing tension between surveillance capitalist technology companies and their concerned users. The regulation provides necessary reassurance that digital privacy is a concern the EU governing institutions are aware of and are taking action to protect. This reassurance could raise trust in Google as they now must abide by this new law. The reason as to why knowledge of the GDPR could increase trust in Google and not Facebook possible rises from differences in public perception of the two companies, an idea explored further in this discussion.

Even though the results were relatively week it must be restated that the repetition of the question regarding trust pre and post-experiment makes this method susceptible to experimenter demand effects. This same problem was observed in the original paper and thus can only be solved by a future restructuring of the survey in which the question regarding trust is asked only after the experiment and therefore is not influenced by the respondents awareness of the experiment or by an attempt to remain consistent throughout the survey.

Part two of the empirical analysis revealed an on average 7.6% increase in concern for Facebook collecting data for commercial purposes when an individual was shown information about the GDPR. This particular significance in the case of Facebook, rather than Google and Federal Statistics Office, raises the question of why information regarding the GDPR effected concern for only Facebook. Descriptive statistics of responses to the comfortability questions post-experiment reveal that mean concern with Facebook sharing data for both research and commercial purposes is higher than that of Google (Table 2). Despite this difference, it is well understood by academics that this mechanism of selling user data to third-parties is a major part of both companies' business models *(Zuboff, 2019)* and is only preventable through specifically "opting-out" under the GDPR *(Sanchez-Rola et al, 2019)*. This trend is backed up by multiple surveys which find similar results regarding public perception of the two tech giants. A 2017 survey of 1,520 US citizens by the *The Verge*, balanced according to US census estimates, recorded that 69% of respondents trusted Google with their information whilst only 41% trusted their information with Facebook *(Newton et al, 2017)*. A similar survey from 2020 by SEO Clarity, of a representative US sample of 1057 participants, revealed that 66% of respondents trusted Google whilst only 38% trusted Facebook *(Gandhi, 2021)*. This discrepancy in trust can in part be explained by Facebook being responsible for more salient privacy scandals and data breaches such as that of Cambridge Analytica, directly resulting from Facebook's mishandling of its users personal data. Although these survey responses stem from questions of trust, it can be assumed that trust differences are mirrored by concern differences between Google and Facebook. This assumption is supported by descriptive statistics of the survey responses to the trust questions investigated in this paper which reveal that trust in Facebook is on average a whole point lower than that of Google on an 11 point scale; a similar difference to that of concern.

Another reason for this difference in concern could be differences in the services that each company provide. Facebook provides a social media platform – a highly personal portrayal of one's digital self – therefore collecting more personal and sensitive information leading to increased concern, compared to the provider of a search engine amongst other services in Google. This difference between the companies is reflected in their revenues, with Facebook receiving 97% of revenue from personalised advertising (*Facebook, 2021)* and Google only 80% *(Alphabet, 2020)*. Thus, Google's reliance on sharing personal data with third parties is marginally diluted compared to Facebook, possibly resulting in lower public concern for their operating activities. This could explain why informing individuals about the GDPR impacts only concern for Facebook. If the GDPR further increases public knowledge regarding privacy-violating practices of technology companies, the costs may fall more heavily on Facebook as they are already in the spotlight for the same issue.

Unsurprisingly, the effect was only observed for commercially-purposed data collection, indicating that users not only care about what information is known about them, but also what said information is used for. The difference between concern for research-purposed and commercially-purposed data collection

likely stems from a traditional mistrust of corporations in which ethics are often eroded by profit motivation *(Adams et al, 2010)*. If this causal finding is genuine it is worrying news for Facebook. The company will now potentially suffer both the GDPR itself cracking down heavily on the business model of personalised advertising through data collection, as well as increased user concern from individuals becoming aware of the GDPR.

Whilst economically fairly miniscule, this causal effect of GDPR knowledge increasing concern with Facebook's commercially-purposed data collecting practices is a meaningful finding. As explained previously, the efficacy of the GDPR relies heavily on the individual agency of internet users to protect their own data by making use of the rights guaranteed by the regulation through various mechanisms of action, the most important being informed consent. In this way the regulation can be described as anti-paternalistic as it does not operate automatically[5]. Thus, for the regulation to be effective individuals must take action, a process which undoubtedly stems from concern for the issue at hand. As such, it is in the best interests of data-privacy regulators, most importantly the EU, for individuals to be concerned about their data privacy as this forms the foundation for the efficacy of the GDPR. If providing individuals with detailed information regarding the GDPR contributes towards increasing digital privacy concern, a case can be made for implementing this on an EU-wide level. Increased communication between a countries' populous and its government has always been a feature of functional democracies (*Heise, 1985)*, and in this case the communication would also result in increased efficacy of the policy. This communication could certainly be implemented through the education system however the rapid development of the industry surrounding data privacy necessitates informing the entire EU populous as a priority. As such, various communication channels such as EU social media channels and direct contact centres, as well as new infrastructures such as the European Citizens Initiative could be used to increase knowledge regarding the GDPR, in turn improving its effectiveness. However, the relatively economically insignificant effect may be negligible in the context of a cost benefit analysis of increasing communication of the regulation.

Finally, part three of the empirical analysis investigated whether becoming aware of the GDPR had any effect on either Google or Facebook account deletion. This was measured through a linear regression model with a dummy treatment of whether an individual became aware of the GDPR over the course of the survey, and outcome variables of whether an individual created or deleted their Google or Facebook accounts, also over the course of the survey. Various relevant controls were added including general privacy concern and general trust.

As mentioned previously, multiple studies have shown that deletion of social media or other internet accounts is most commonly explained by digital privacy concerns *(Baumer et al, 2013; Stieger et al,*

[5] Except with regard to organisations being obliged to notify supervising authorities within 72 hours of a data breach *(European Union, 2016)*.

2013). Thus, a significant effect between those who became aware of the GDPR and those who deleted their accounts could suggest that GDPR awareness penetrates through the privacy paradox and influences individuals to reflect their privacy concerns in their actual behaviour. Conversely, a negative relationship between GDPR awareness and account deletion could also be conceivable as individuals have more faith that their data is being protected. This could be explained by the results of part one of the empirical analysis that showed trust with Google to marginally increase. Account creation on the other hand has not been shown to reflect low privacy concerns meaning this hypothetical effect does not run contrarily to account deletion. Thus, it is unsurprising that no effect was found for account creation. In fact, no effect was found for either, indicating that GDPR awareness most likely does not have an effect on behaviour with regard to account deletion and creation.

The findings were tested for robustness using balance tests to determine whether the random assignment of the experimental treatment was successful. The results (displayed in the appendix), revealed that binary variables for both sex and PC ownership were both significantly correlated to treatment. This can only have been random variation and hence can be controlled for. These variables were added as controls along with controls for age, general privacy concern and general trust, to further increase the accuracy of the estimates. Thus, it can be argued that the results are internally valid. This is excluding the possibility that the results for the regressions regarding trust are influenced by experimenter demand effects.

Part three represented an observational, rather than an experimental methodology, as assignment of the treatment variable could not have been randomised. The same appropriate control variables were included however there is always potential for omitted variable bias as additional possible unobserved confounders may be present. As such, the internal validity is not perfect but certainly maximised under the circumstances. In any case no significance was found and as such no causal effect can be claimed.

As mentioned previously, the study focuses on a German sample and is thus not necessarily externally valid. On average, the German population is particularly sensitive to matters of information privacy *(Morey et al, 2015)*. This could possibly indicate that in countries where this concern is less present, the effect of GDPR information on trust and digital privacy concern could differ greatly. This could be in either direction as less concerned countries may be more sensitive to such an experiment as preferences are less firm, or less sensitive as the importance of the GDPR may not come across as strongly. In the same vein, account deletion or creation trends may differ greatly internationally according to the presence of each organisation in the country.

Conclusion

This paper attempted to investigate *whether the provision of detailed information regarding the GDPR influences individuals trust and data privacy concern with regard to specific data-driven organisations and specific data-collection purposes*. This is emerging as an increasingly relevant topic in the wake of a new economic framework in which large technology companies such as Google and Facebook rely on a business model of data-collection that inevitably intrudes on user privacy. The GDPR represents an EU level initiative to restore balance to the growing informational inequality between internet companies and their users. However, its composition as an anti-paternalistic policy necessitates that its efficacy relies on the agency of the society it is designed to protect, thus motivating the importance of understanding the effects of communication of the regulation.

Using experimental survey results from a German sample, I have found relatively weak evidence that providing individuals in this sample with detailed information regarding the GDPR increased trust in both Google not to share user-data with third parties. No effect was found for Facebook. This was found using linear regression models which featured responses to a question regarding trust following the experiment and using lagged dependant variables as controls. The results however were statistically weak leading to inconclusive results that mirrored a study published in May 2021 which used the difference between results to the same question asked both before and after the experiment *(Bauer et al, 2021)*. Both studies are potentially erroneous if the survey structure was subject to experimenter demand effects.

The same survey experiment featured questions regarding comfort with specific data-collecting organisations sharing data for specific purposes. Operationalising comfort as an appropriate inverse measure for data privacy concern, I found that providing information regarding the GDPR increased concern for Facebook sharing data for commercial purposes by 7.6%. No significant effects were found in the context of research-purposed data sharing or other data-collecting organisations. Having controlled for unbalanced factors, this findings can also be seen as causal as treatment was randomly assigned.

Finally, using multiple waves of the same survey, an analysis was conducted to see whether becoming aware of the GDPR had an effect on whether an individual created or deleted an account with Google or Facebook. This was motivated by a growing body of literature investigating the discrepancy between digital privacy preferences and digital privacy behaviour, also known as the privacy paradox. No significant effects were found leading to the conclusion that account creation or deletion is not influenced by awareness of the GDPR.

Suggestions for further research would include expanding the sample to other European countries to see how these effects differ within the EU. Additionally, similar research could be conducted in locations where similar regulation has been introduced such as the *California Consumer Privacy Act* (CCPA) or the *Consumer Data Protection Act* (CDPA), soon to be introduced in Virginia. As insignificant results were found for account creation or deletion, future studies could incorporate more subtle indicators of privacy concern reflected in behaviour. This could be in the form of a survey which would test whether GDPR awareness effects the likelihood of a user to accept or reject cookies when browsing on the internet.

References

Adams, J. E., Highhouse, S., & Zickar, M. J. (2010). Understanding general distrust of corporations. Corporate Reputation Review, 13(1), 38-51.

Alphabet Inc. (2020). 2020 Annual Report. Retrieved from https://abc.xyz/investor/static/pdf/2020_alphabet_annual_report.pdf?cache=8e972d2

Andreou, A., Venkatadri, G., Goga, O., Gummadi, K., Loiseau, P., & Mislove, A. (2018, February). Investigating ad transparency mechanisms in social media: A case study of Facebook's explanations. In NDSS 2018-Network and Distributed System Security Symposium (pp. 1-15).

Anandam, P., Hernandez, B., Miller, J., Voutilainen, S., & Zavesov, V. (2005, December 7). Privacy and Data. .

Bauer, P. C., Gerdon, F., Keusch, F., Kreuter, F., & Vannette, D. (2021). Did the GDPR increase trust in data collectors? Evidence from observational and experimental data. Information, Communication & Society, 1-21.

Bauer, P. C., Keusch, F., & Kreuter, F. (2019, February). Trust and Cooperation: Evidence From the Realm of Data-Sharing. SSRN Electronic Journal. doi:DOI:10.2139/ssrn.3327244

Baumer, E. P., Adams, P., Khovanskaya, V. D., Liao, T. C., Smith, M. E., Schwanda Sosik, V., & Williams, K. (2013, April). Limiting, leaving, and (re) lapsing: an exploration of facebook non-use practices and experiences. In Proceedings of the SIGCHI conference on human factors in computing systems (pp. 3257-3266).

Bashir, M., Hayes, C., Lambert, A.D. and Kesan, J.P. (2015), Online privacy and informed consent: The dilemma of information asymmetry. Proc. Assoc. Info. Sci. Tech., 52: 1-10. https://doi.org/10.1002/pra2.2015.145052010043

Bartsch, M., & Dienlinn, T. (2016, March). Control your Facebook: An analysis of online privacy literacy. Computers in Human Behaviour, 56, 147-154. doi:https://doi.org/10.1016/j.chb.2015.11.022

Brown, B. (2001). Studying the Internet experience. HP laboratories technical report HPL, 49.

Chiou, L., & Tucker, C. (2017). Search engines and data retention: Implications for privacy and antitrust (No. w23815). National Bureau of Economic Research.

Dabrowski A., Merzdovnik G., Ullrich J., Sendera G., Weippl E. (2019) Measuring Cookies and Web Privacy in a Post-GDPR World. In: Choffnes D., Barcellos M. (eds) Passive and Active Measurement. PAM 2019. Lecture Notes in Computer Science, vol 11419. Springer, Cham. https://doi.org/10.1007/978-3-030-15986-3_17

Dienlin, T., & Trepte, S. (2015). Is the privacy paradox a relic of the past? An in-depth analysis of privacy attitudes and privacy behaviors. European journal of social psychology, 45(3), 285-297.

Epstein, D., & Quinn, K. (2020). Markers of Online Privacy Marginalization: Empirical Examination of Socioeconomic Disparities in Social Media Privacy Attitudes, Literacy, and Behavior. Social Media + Society. https://doi.org/10.1177/2056305120916853

European Commission. (2019). Special Eurobarometer 487a – March 2019: The General Data Protection Regulation". European Union. 10.2838/579882

European Union (2016) Regulation 2016/679. Official Journal of the European Communities

Facebook, Inc. (2021). Facebook Reports First Quarter 2021 Results. Retrieved from https://investor.fb.com/investor-news/press-release-details/2021/Facebook-Reports-First-Quarter-2021-Results/default.aspx

Fletcher, D. (2010). How Facebook is redefining privacy.

Gandhi, M. (2021, January 11). Do Americans Trust Tech Giants?. In seoClarity.

Geissler, R. C. (2011). Private eyes watching you: Google Street View and the right to an inviolate personality. Hastings LJ, 63, 897.

Hargittai, E., Piper, A. M., & Morris, M. R. (2019). From internet access to internet skills: digital inequality among older adults. Universal Access in the Information Society, 18(4), 881-890.

Heise, J. A. (1985). Toward closing the confidence gap: An alternative approach to communication between public and government. Public Administration Quarterly, 196-217.

Isaak, J., & Hanna, M. J. (2018). User data privacy: Facebook, Cambridge Analytica, and privacy protection. Computer, 51(8), 56-59.

Kane, B. (2010). Rethinking the Internet's Privacy Dilemma: A Modest Call for Informed, Nimble Solutions. Alb. LJ Sci. & Tech., 20, 375.

Kokolakis, S. (2017, January). Privacy attitudes and privacy behaviour: A review of current research on the privacy paradox phenomenon. Computers and Security, 64, 122-134. doi:https://doi.org/10.1016/j.cose.2015.07.002

Livingstone, S., Stoilova, M. and Nandagiri, R. (2020). Data and Privacy Literacy. In The Handbook of Media Education Research (eds D. Frau-Meigs, S. Kotilainen, M. Pathak-Shelat, M. Hoechsmann and S.R. Poyntz). https://doi.org/10.1002/9781119166900.ch38

Maldoff, G. (2016, January 12). Top 10 operational impacts of the GDPR: Part 3 – consent. In IAPP. Retrieved from https://iapp.org/news/a/top-10-operational-impacts-of-the-gdpr-part-3-consent/

Morey, T., Forbath, T., & Schoop, A. (2015). Customer data: Designing for transparency and trust. Harvard Business Review, 93(5), 96-105.

Newton, C., Statt, N., & Zelenko, M. (2017, October 27). THE VERGE TECH SURVEY: How Americans really feel about Facebook, Apple, and more. In The Verge.

Norberg, P. A., Horne, D. R., & Horne, D. A. (2007). The privacy paradox: Personal information disclosure intentions versus behaviors. Journal of consumer affairs, 41(1), 100-126.

Nouwens, M., Liccardi, I., Veale, M., Karger, D., & Kagal, L. (2020, April). Dark patterns after the GDPR: Scraping consent pop-ups and demonstrating their influence. In Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems (pp. 1-13).

Peel, K., & Tretter, E. (2019). Waterfront Toronto: Privacy or Piracy?.

Pingo Z., Narayan B. (2019) Privacy Literacy and the Everyday Use of Social Technologies. In: Kurbanoğlu S. et al. (eds) Information Literacy in Everyday Life. ECIL 2018. Communications in Computer and Information Science, vol 989. Springer, Cham. https://doi.org/10.1007/978-3-030-13472-3_4

Puaschunder, J. (2020). Towards a Utility Theory of Privacy and Information Sharing and the Introduction of Hyper-Hyperbolic Discounting in the Digital Big Data Age. In Idemudia, E. C. (Eds.), Handbook of Research on Social and Organizational Dynamics in the Digital Era (pp. 157-200). IGI Global. http://doi:10.4018/978-1-5225-8933-4.ch008

Reviglio, U. (2019, September). Towards a right not to be deceived? An interdisciplinary analysis of media personalization in the light of the GDPR. In Conference on e-Business, e-Services and e-Society (pp. 47-59). Springer, Cham.

Rustad, M. L., & Koenig, T. H. (2019). Towards global data privacy standard. Florida Law Review, 71(2), 365-454.

Sanchez-Rola, I., Dell'Amico, M., Kotzias, P., Balzarotti, D., Bilge, L., Vervier, P. A., & Santos, I. (2019, July). Can i opt out yet? gdpr and the global illusion of cookie control. In Proceedings of the 2019 ACM Asia conference on computer and communications security (pp. 340-351).

Shahizam, S. (2021). Putting Consent in its Place: Proceduralism and Paternalism in Data Protection Law. LSE Law Review, 6(3).

Škrinjarić, B., Budak, J., & Rajh, E. (2019). Perceived quality of privacy protection regulations and online privacy concern. Economic research-Ekonomska istraživanja, 32(1), 982-1000.

Solove, D. J. (2021). The myth of the privacy paradox. Geo. Wash. L. Rev., 89, 1.

Stieger, S., Burger, C., Bohn, M., & Voracek, M. (2013). Who commits virtual identity suicide? Differences in privacy concerns, internet addiction, and personality between Facebook users and quitters. Cyberpsychology, Behavior, and Social Networking, 16(9), 629-634.

Susser, D., Roessler, B., & Nissenbaum, H. (2018). Online manipulation: Hidden influences in a digital world.

Utz, C., Degeling, M., Fahl, S., Schaub, F., & Holz, T. (2019, November). (Un)informed Consent: Studying GDPR Consent Notices in the Field. CCS '19: Proceedings of the 2019 ACM SIGSAC Conference on Computer and Communications Security, 973-990. doi:https://doi.org/10.1145/3319535.3354212

Van Ooijen, I., & Vrabec, H. u. (2018, December). Does the GDPR Enhance Consumers' Control over Personal Data? An Analysis from a Behavioural Perspective. Journal of Consumer Policy, 42, 91-107. doi:https://doi.org/10.1007/s10603-018-9399-7

von Briel, F., & Davidsson, P. (2019). Digital Platforms and Network Effects: Using Digital Nudges for Growth Hacking.

Wachter, S. (2018). Normative challenges of identification in the Internet of Things: Privacy, profiling, discrimination, and the GDPR. Computer law & security review, 34(3), 436-449.

Wirtz, J., Lwin, M. O., & Williams, J. D. (2007). Causes and consequences of consumer online privacy concern. International Journal of Service Industry Management.

Wu, K. W., Huang, S. Y., Yen, D. C., & Popova, I. (2012). The effect of online privacy policy on consumer privacy concern and trust. Computers in human behavior, 28(3), 889-897.

Xie, W., & Karan, K. (2019). Consumers' privacy concern and privacy protection on social network sites in the era of big data: Empirical evidence from college students. Journal of Interactive Advertising, 19(3), 187-201.

Youn, S. (2009). Determinants of online privacy concern and its influence on privacy protection behaviors among young adolescents. Journal of Consumer affairs, 43(3), 389-418.

Young, A. L., & Quan-Haase, A. (2013). Privacy protection strategies on Facebook: The Internet privacy paradox revisited. Information, Communication & Society, 16(4), 479-500.

Zizzo, D. J. (2010). Experimenter demand effects in economic experiments. Experimental Economics, 13(1), 75-98.

Zuboff, S. (2019). The Age of Surveillance Capitalism: The Fight for a Human Future at the New Frontier of Power. N.p.: Profile Books.

Appendix

Appendix A – GDPR Experiment

It is always difficult to communicate new regulations and directives of the EU in a good and comprehensive way so that they are understood by all people affected. Please read the following description and then answer a question.

Since May of this year, the General Data Protection Regulation (GDPR) has been in effect. The goal of the GDPR is to enforce better data protection rights for European citizens. The GDPR contains the following points, among others:

- Right to clear and comprehensible information on who processes your data, which data are processed and why.
- Right to information about the personal data an organization has about you.
- Right to demand a service provider to transfer your personal data to another service provider, e.g., if you switch to another social network on the Internet.
- Right to be "forgotten", that is, to have your data deleted if you no longer wish it to be processed and if there are no legitimate reasons for the organization concerned to continue storing your data.
- Organizations that want to process your data must ask for your consent und clearly indicate how your personal data is to be used.
- If your data is lost or stolen, the organization responsible must inform you and the appropriate data protection authority immediately if this data breach threatens to harm you.

Notes: This text was randomly presented to roughly half of the survey respondents. It details the most important rights guaranteed by the GDPR.

Appendix B – BALANCE TESTS FOR TRUST

	Trust in Google Pre-Experiment	Trust in Facebook Pre-Experiment
GDPR Treatment	0.032	0.214*
	(0.122)	(0.119)
Constant	3.690	2.598
	(0.087)	(0.083)
Observations	2,108	2,114

Notes: This table displays the results of balance tests for the variables listed in the top row. These balance tests are linear regression models with the GDPR experiment as the dummy treatment variable and the mentioned outcome variables of trust in Google Pre-Experiment and trust in Facebook Pre-Experiment. * represents significance of the coefficient at 10%, ** significance at 5% and *** significance at 1%.

Appendix C – BALANCE TESTS

	Age	Sex	General Privacy Concern	General Trust	GDPR Awareness
GDPR Treatment	0.386	0.046**	-0.049	0.088	0.001
	(0.598)	(0.022)	(0.031)	(0.109)	(0.019)
Constant	45.464	0.476	2.722	4.979	0.739
	(0.422)	(0.015)	(0.023)	(0.078)	(0.014)
Observations	2117	2117	2112	2112	2110

Notes: This table displays the results of balance tests for the variables listed in the top row. These balance tests are linear regression models with the GDPR experiment as the dummy treatment variable and the mentioned outcome variables of age, sex, general privacy concern, general trust and GDPR awareness. * represents significance of the coefficient at 10%, ** significance at 5% and *** significance at 1%.

Appendix D – BALANCE TESTS FOR DEVICE OWNERSHIP

	Smartphone	Non-Smartphone	PC	Tablet	Ebook
GDPR Treatment	-0.008	-0.005	-0.0156**	-0.006	0.001
	(0.016)	(0.021)	(0.008)	(0.022)	(0.019)
Constant	1.159	1.669	1.041	1.464	1.748
	(0.011)	(0.015)	(0.006)	(0.0154)	(0.014)
Observations	2115	2060	2106	2090	2076

Notes: This table displays the results of balance tests for variables relating to device ownership. These balance tests are linear regression models with the GDPR experiment as the dummy treatment variable and the mentioned outcome variables of smartphone ownership, non-smartphone ownership, PC ownership, tablet ownership and Ebook ownership. * represents significance of the coefficient at 10%, ** significance at 5% and *** significance at 1%.

Appendix E – BALANCE TESTS FOR ACCOUNT OWNERSHIP

	Google Account	Facebook Account	Twitter Account	LinkedIn Account	Xing Account
GDPR Treatment	0.006	0.050	-0.005	0.025	0.026
	(0.026)	(0.033)	(0.041)	(0.038)	(0.042)
Constant	0.857	0.924	0.508	0.385	0.519
	(0.019)	(0.023)	(0.029)	(0.026)	(0.029)
Observations	2112	2111	2105	2107	2109

Notes: This table displays the results of balance tests for variables relating to internet platform account ownership. These balance tests are linear regression models with the GDPR experiment as the dummy treatment variable and the mentioned outcome variables of Google account ownership, Facebook account ownership, Twitter account ownership, LinkedIn account ownership and Xing account ownership. Xing is a career-oriented social media platform similar to LinkedIn and popular in Germany. * represents significance of the coefficient at 10%, ** significance at 5% and *** significance at 1%.