Alma Mater Studiorum – Università di Bologna

DOTTORATO DI RICERCA IN

PhD in Management

Ciclo XXXII

Settore Concorsuale: 13/B2 – Economia e Gestione delle Imprese

Settore Scientifico Disciplinare: SECS-P/08 – Economia e Gestione delle Imprese

How do Firms ask for Consumers' Data Permission? The Value of Companies Data Practices.

Presentata da: Caterina D'Assergio

Coordinatore Dottorato

Prof. Riccardo Fini

Supervisore

Prof.ssa Sara Valentini

Co-Supervisori

Prof.ssa Elisa Montaguti
Prof. Puneet Manchanda

Esame finale anno 2021

ABSTRACT

On May 25, 2018, the EU introduced the General Data Protection Regulation (GDPR) that offers EU citizens a shelter for their personal information by requesting companies to explain how people's information is used clearly. To comply with the new law, European and non-European companies interacting with EU citizens undertook a massive data re-permission-request campaign. However, if on the one side the EU Regulator was particularly specific in defining the conditions to get customers' data access, on the other side, it did not specify how the communication between firms and consumers should be designed. This has left firms free to develop their re-permission emails as they liked, plausibly coupling the informative nature of these privacy-related communications with other persuasive techniques to maximize data disclosure. Consequently, we took advantage of this colossal wave of simultaneous requests to provide insights into two issues. Firstly, we investigate how companies across industries and countries chose to frame their requests. Secondly, we investigate which are the factors that influenced the selection of alternative re-permission formats. In order to achieve these goals, we examine the content of a sample of 1506 re-permission emails sent by 1396 firms worldwide, and we identify the dominant "*themes*" characterizing these emails. We then relate these themes to both the expected benefits firms may derive from data usage and the possible risks they may experience from not being completely compliant to the spirit of the law. Our results show that: (1) most firms enriched their re-permission messages with persuasive arguments aiming at increasing consumers' likelihood of relinquishing their data; (2) the use of persuasion is the outcome of a difficult tradeoff between costs and benefits; (3) most companies acted in their self-interest and "*gamed the system*". Our results have important implications for policymakers, managers, and customers of the online sector.

Keywords: *GDPR, Re-Permission Emails, Persuasive Communication, Online Ad Revenues, Marketing Cookies, Data-Breach, Text Analysis.*

TABLE OF CONTENTS

1. INTRODUCTION

In recent years, data have become increasingly central to firms' actions. The multiplication of data sources and the spread of analytical skills have made collecting and processing data simpler and more effective with positive and negative consequences for consumers (Bughin et al. 2018; Lohr 2012; Wedel and Kannan 2016). On the one hand, consumers have benefitted from firms' use of data as they have become the target of campaigns shaped around their needs (D'Annunzio and Russo 2020). On the other hand, consumers have observed their details turning into firms' valuable goods, increasingly generating positive revenue streams while putting their privacy at risk, as in the case of Netflix that used viewing choices to infer race and use it to target customers (Zarum 2018).

The prominent role of data in commercial practices has led several regulators worldwide (EU, Canada, Australia, and California) to guarantee strong protection for individuals regarding their personal data. The EU (GDPR, 2018) and, later, the State of California (CCPA, 2020) have introduced more stringent privacy-related regulations than ever before (Privacy Act 1974; EU Directive on Personal Data Protection 1995) to offer citizens a shelter for the protection of their personal information. The European GDPR, for instance, requires companies to be transparent about their data collection, usage, and transmission and enforces higher security standards than ever before. The idea behind this regulation was to make privacy concerns more salient, the access to consumers' data more transparent, and to hold firms accountable in order to reduce opportunities for data exploitation. The GDPR and CCPA, which might pave the way to federal law on privacy (Foote 2019), differ in several dimensions, including the scope of application, the nature and the extent of the limitations, and accountability. However, they both include requirements to better protect personal data. For example, these legislations require to inform users about which data have been gathered on them and which are the procedures to ask for data deletion in order not to face fines and litigations (Murphy 2020).

Under the European GDPR, firms' adherence to this privacy law starts with firms' asking for permission from users to allow them (firms) to use their (user) data. Therefore, to comply with the new law requirements, in 2018, European and non-European companies interacting with EU citizens engaged in a massive data re-permission-request campaign whereby firms asked their customers to grant them the right to use and trade their data. These request campaigns were mainly channeled via email and explicitly asked individuals to choose whether or not granting firms the right to use and trade their data.

The phenomenon of the GDPR re-permission emails has, consequently, attracted the attention of the business press (e.g., Mikkelsen et al. 2017; Shrimsley 2018; Weiss 2018) because of the unique and massive effort concentrated in a specific time and because of the possible negative consequences on several marketing practices that are core to current digital ad targeting (Ghosh 2018). Academic literature shows that regulations aiming at protecting individual privacy and reducing online information collection have strong collateral effects for the whole online advertising industry, which is mainly based on data and programmatic technologies. The implementation of the GDPR "*data minimization*" principle has negatively impacted the firms' ability to track users on the web and harvest data (Johnson and Shriver 2019; Libert, Graves, and Nielsen 2018; Peukert et al. 2020; Sørensen and Kosta 2019), leading to lower online advertising revenue streams (Beales and Eisenach 2014; Goldberg, Johnson, and Shriver 2019; Johnson 2013; Marotta, Abhishek, and Acquisti 2019) and concentrating the online ad market on few dominant firms (Berry, Gaynor, and Morton 2019; Brill 2011; Johnson and Shriver 2019; Libert, Graves, and Nielsen 2018; Peukert et al. 2020). Therefore, it is not surprising that the enforcement of the GDPR has seriously worried firms operating in the online sector. Yet, thanks to the GDPR re-permission emails phenomenon, companies may have found other ways to avoid the "*unintended but unavoidable*" consequences of privacy regulations.

Interestingly, the EU Regulator has been very specific in defining the conditions firms should observe to access and use customers' data. However, it did not dictate how consumers' consent should be obtained and how the privacy communication between firms and consumers should be designed in terms of text, format, or communication structure. Previous literature has shown that how data permission is asked is particularly relevant to prompt customers to disclose and achieve their opt-in (Acquisti, Brandimarte, and Loewenstein 2015; Utz et al. 2019). The "*how*" factor has been extensively investigated by literature on consumer behavior and privacy, which has identified tools and factors that managers can exploit to heighten customers' protection perception and lessen privacy concerns and feelings of vulnerability (Martin 2018). For example, companies can decide to merely obey the GDPR requirements and craft their email by stressing their *informative* content and highlighting the possibility for the user to manage the data disclosed (Martin, Borah, and Palmatier 2017; Phelps, Nowak, and Ferrell 2000; Tucker 2014). Companies can instead decide to use other strategies to maintain the data collected and sustain their data-based operations, in that trying to *persuade* their final user by using, for example, a particular framing of the message or by providing them coupons or discounts (Athey, Catalini, and Tucker 2017; Chellappa and Sin 2005; Grossklags and Acquisti 2007). Consequently, it is possible to determine two broad categories of themes characterizing privacy-related communications: *informative themes* provide information about data privacy and protection and are related to the "*GDPR principles*" (e.g., transparency and control); *persuasive themes* mainly intend to prompt customers' data disclosure behavior using marketing tools (e.g., incentives and framing).

Under the GDPR, firms were free to create their re-permission e-mails as they liked and, plausibly, with the intent of maximizing data disclosure, for example, by using persuasive arguments such as discounts in exchange for data. Notably, this is/was not in contrast with the regulator's request to make citizens explicitly decide whether or not to grant their data. Still, it begs the most effective strategy to obtain data usage consent and which factors influence firms' decision in such matters. This is what we plan to study.

First, we are interested in understanding how re-permission emails used for the European GDPR enforcement were designed and which themes were used to ask for access to customers or potential customers' personal data. Did firms emphasize more transparency and control, as suggested by the Regulator, or did they also try to persuade their customers by offering them some rewards or by highlighting the negative consequences associated with denying access to their data? Answering these questions allows us to shed some light on how firms responded to privacy regulation and how they chose to interact with their customers.

Second, we are interested in understanding whether the tradeoff between the expected benefits derived from data use and the financial and reputational costs of not adhering to the scope of the regulation influenced the design of the re-permission requests. The request of data and the use of data are intertwined decisions because the value of the data influences how they are asked. Knowing how much a firm can capitalize on its customers' details might affect how much a firm is ready to offer for them. For example, firms that have the attitude to monetize their website traffic through advertising might be more inclined to provide financial incentives into their re-permission emails as they know the value of their data. At the same time, companies that have already experienced some of the risks from not being completely compliant with the data security standards (e.g., Data-Breach Announcements) may have incentives to more strictly comply with the GDPR principles and to use re-permissions emails merely as informative tools.

Therefore, this thesis aims to study the re-permission email phenomenon and to identify the circumstances under which firms designed their privacy-related communications in a more persuasive vs. informative fashion by providing an answer to the following two main research questions:

(i) Are there any systematic patterns in how firms designed their re-permission emails to request access to users' data?

(ii) How self-interested were firms? Did the benefits (from the use of data) as well as the costs (of non-compliance) drive request content and intent?

To address these questions, we collect a sample of 1506 re-permission emails sent by 1396 firms worldwide on the occasion of the GDPR enforcement.

First, we develop a compelling modeling approach to analyze re-permission e-mails, and we empirically document the use of six key themes characterizing our sample: transparency, control, which are mainly *informative*; framing (gains vs. losses or time orientation), incentives (both monetary and non-monetary incentives) which are mainly *persuasive*. Our results also show significant heterogeneity in the way firms conceptualized their re-permission emails. Both informative and persuasive themes were used, but a considerable portion of firms used only *persuasive themes* to increase consumers' likelihood of relinquishing their data. Therefore, companies relied more on persuasion than information, which is probably not consistent with the EU Regulator's intentions.

Second, we relate the benefits firms may derive from data collection – e.g., data access and expected returns on data collection – to the risk companies may incur when collecting data – e.g., reputation or customers' reactance. Our preliminary analyses indicate that the likelihood of opting for persuasive vs. informative themes was influenced by the delicate tradeoff that firms faced in trying to balance the potential reputational damage from non-compliance with the law and the benefits they might derive from data usage. Our findings also suggest that firms designed re-permission e-mails opportunistically, providing evidence that companies, when provided with freedom of choice, behave in a self-interested way and try to "*game the system*" by designing their re-permission emails in a more *persuasive* fashion in order to achieve data access more easily. Nonetheless, we also show that this, at least, entails positive externalities for the data owners who get re-paid for the data disclosed.

Notably, this dissertation's results have important implications for marketing researchers and policymakers.

We contribute to the theoretical debate on privacy regulations in four main ways. Firstly, we demonstrate that the drivers and the mechanisms of persuasive communications (identified in previous studies done in economics, privacy, marketing, and psychology) are also applicable to the realm of privacy: companies, when deciding about the inclusion of persuasive cues in their privacy-related communications, act in a self-interested way by evaluating the benefits against the risks they might derive from the collection of users' data and opting for persuasion only if the expected benefits outweigh the expected costs. Secondly, we provide a unifying conceptual framework, which combines two streams of the privacy-related literature: studies on how to request for personal data (in order to mitigate feelings of privacy concerns and more easily achieve data permission) and studies about the impact of GDPR on firms' online performances and industry competition. Thirdly, we show how this framework translates into practice by analyzing a large set of GDPR re-permission emails and by identifying the main themes used by a large set of firms on the occasion of the massive new European GDPR enforcement in 2018. Lastly, we model and test the association between the firms' communication strategy used to obtain users' data and the expected benefits and costs which may be related to firms' data collection practices.

We also provide some empirical advancements: (i) we proposed a new measurement for the websites' economic performance; (ii) we used an unsupervised method to automatically content-analyze privacy-related communications, and we showed that it reaches consistency with more traditional text analysis approaches (e.g., manual content analysis); (iii) we developed an efficient and stable system to content-analyze texts (and that can be easily scaled up to larger datasets) which integrates a theory-based approach for textual analysis with automatic text mining tools.

Policymakers can also learn from this work. We showed that companies coupled the *informative* nature of the re-permission emails with *persuasive* arguments to entice users' disclosure behavior. We also provide additional evidence that the use of persuasion is particularly relevant for companies that are more likely to profit out of the data collection and that behave in a more opportunistic way. This may pose some concerns about the real usefulness of the GDPR. If the users' behavior is a mere consequence of the persuasion, then the whole point of getting an "*informed and explicit*" opt-in and increasing users' awareness of firms' data practices becomes meaningless.

The present dissertation proceeds as follows. After this introduction, the following chapter presents the institutional setting, illustrating the General Data Protection Regulation (GDPR), and providing a summary of the main motivations that have brought about the observed worldwide privacy spread. The third chapter illustrates the theoretical background of this thesis, describing the two main streams of literature related to the topic of privacy. Chapter 4 deals with the development of the conceptual framework and outlines the main research questions of this thesis. The dataset construction and the analysis approach are detailed in Chapter 5. The subsequent two chapters focus on each of the two research questions presented: Chapter 6 provides a detailed description of the "*themes*" that characterize re-permission emails, while Chapter 7 presents the results of the models addressing the tradeoff between the benefits and costs derived from data usage and the design of re-permission emails. The thesis ends by discussing the key findings for both firms and policymakers, by presenting the main limitations of this study, and by proposing directions for future research.

2. Institutional Background

As previously discussed, there is no doubt that data are the new oil for companies. Thanks to data, companies can make better predictions, make smarter decisions, and be more efficient and precise in detecting and exploiting pivotal opportunities that are crucial for their survival in today's hyper-competitive economic environment. Managerial press and marketing literature have shown that companies making use of data-driven strategies are more profitable and productive than their competitors, which showed a low reliance on data (Groenfeldt 2015; Lohr 2012; McAfee and Brynjolfsson 2012), suggesting that the collection and use of data is no more an optional resource for the nowadays firms.

Additionally, data have been shown to be relevant also for customers who are better-served thanks to the analysis of behavioral data about their online surfing paths. Thanks to firms' targeting and advertising strategies, customers are provided with discounts, coupons, and price cuts that companies use to tempt them to finalize the purchase.

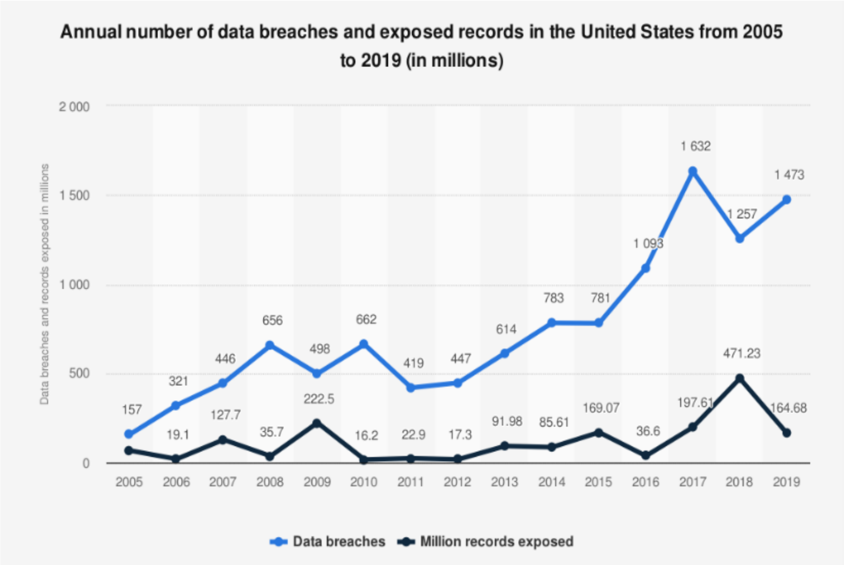
However, even if data usage presents numerous benefits for both companies and customers, it also brought about serious privacy risks, resulting in a substantial call for consumers' protection. This has resulted in a worldwide surge of new data protection policies and regulations that mainly aim at providing a set of rules and principles that data owners can use to preserve their privacy.

In this chapter, we provide a summary of the main motivations that have brought about this privacy spread, as well as some more detailed information about the General Data Protection Regulation (GDPR), which has been enforced in May 2018 and can be considered as the reference standard for the privacy regulations to come.

2.1. Privacy Regulations: The Motivations

Several elements have led to the development of privacy regulations worldwide. One of the main factors determining the emergence of the increasing need for protection by customers is the rising number of security failures experienced in the last decades (e.g., Cambridge Analytica, 2018; Equifax, 2017; Yahoo, 2013, 2014). The rise in data breach announcements (DBA) experienced in the last decades has been astonishing: from 781 data breaches in 2015 to 1.473 in 2019, reaching a peak of 1.632 in 2017 and resulting in an increase of about 88% in just four years (Figure 2.1.1). This means that people are increasingly aware of the potential negative consequences of data breaches and, consequently, are less willing to provide personal information to companies. Accordingly, academic research has shown that violation of privacy expectations, such as unauthorized selling of sensitive data to third-party vendors, is particularly important for a firm's credibility since it has a direct negative impact on trust (Martin 2018). However, trust has been found to be one of the main factors mitigating the negative effect of DBA; consequently, its absence can produce negative impacts on firms' profitability (Acquisti, Friedman, and Telang 2006; Janakiraman, Lim, and Rishika 2018; Martin, Borah, and Palmatier 2017).

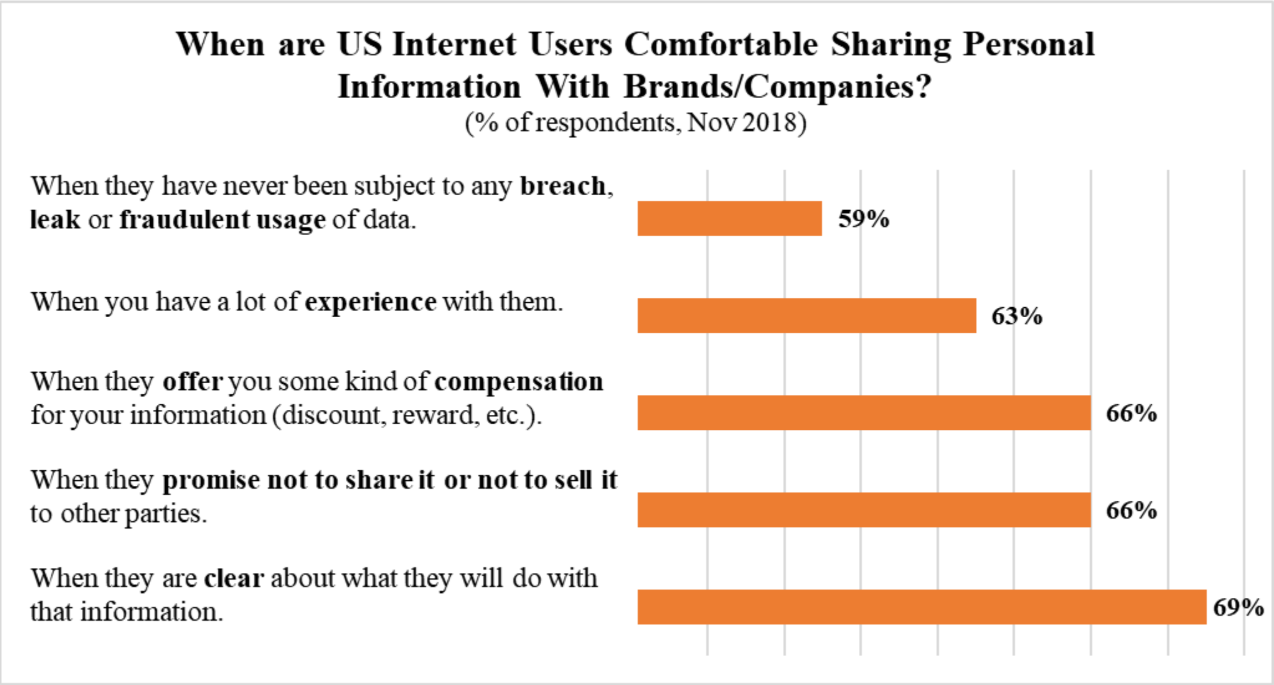
Figure 2.1.1 – Data Breaches Trend in the USA.



Source: Statista

Another serious concern that consumers are increasingly facing regards the covert collection of data; if data were collected face-to-face in the pre-Internet era, we are now constantly moving to an anonymous collection of data. In the Internet & Mobile Era, technological innovations and the birth of social media had allowed marketers to automatically collect users' personal information without or with little consciousness of the final consumer (Peppers and Rogers 2016). In a recent EMarketer report on marketing in the digital world, it is reported that 83% of internet users worldwide are concerned about their privacy (Fisher 2019). This concern becomes even stronger in the case of *covert* data collection. If users become aware of the collection and use of their data only when they receive a highly personalized advertisement or a data breach announcement on their data, then feelings of vulnerability, and the consequent privacy concerns, can have a strong effect, respectively, on the returns of the firm experiencing the privacy failure and on the effectiveness of the ad. This has been shown to lead to a phenomenon of *reactance* to the targeted advertising, meaning that the more the customers feel vulnerable, the more they react negatively to the firms' messages targeted to them, making use of highly personal information (White et al. 2008) or to the firm's personalized ads perceived as highly intrusive (Tucker 2014). Additionally, many studies have shown that if users know how their data are used, they are more willing to disclose sensitive information (Aguirre et al. 2015; Benson, Saridakis, and Tennakoon 2015). Accordingly, as it is possible to deduce from Figure 2.1.2, customers are not against data sharing *a priori*; they need some precise information about how their data is collected, used, and shared in order to feel safe in disclosing. Interestingly, customers are also increasingly conscious that their data have value. Indeed, 66% of the respondents said that they are willing to trade personal information with some sort of incentive (Figure 2.1.2).

Figure 2.1.2 – Customers' Motivations to Share Data.

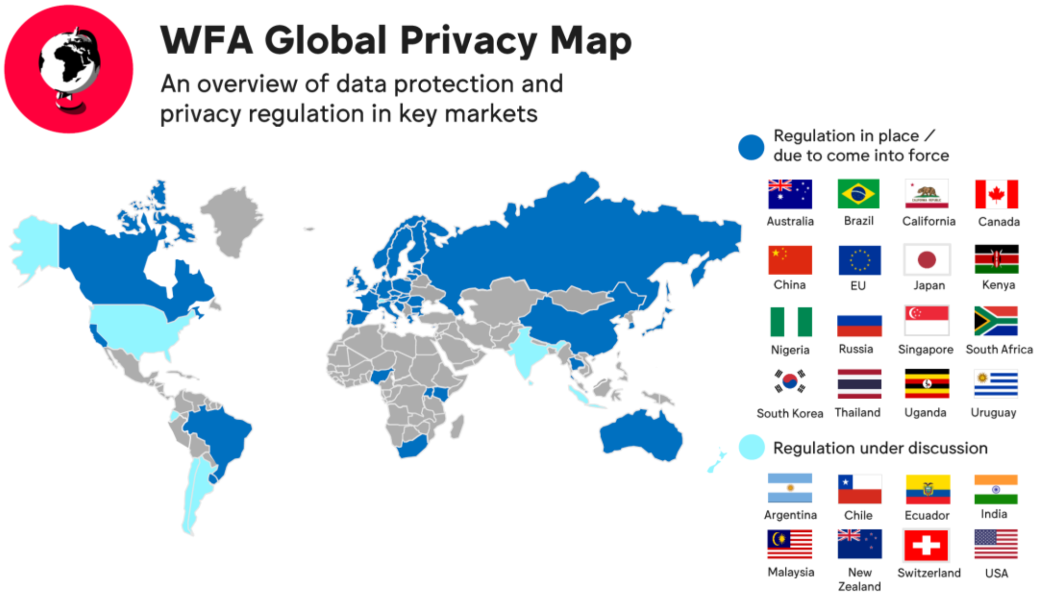


Source: eMarketer

The increasing call for protection has resulted in a worldwide spread of data protection regulations. The legislative scenario had evolved drastically in the last few years to accommodate the novelties of the digital world (Figure 2.1.3). For example, in the last two years, important regulations have been implemented in two leading continents: the General Data Protection Regulation (GDPR) in Europe and the California Consumer Privacy Act (CCPA) in North America. Both laws regard individual data privacy and protection and aim to re-empower individuals on their data propriety. Additionally, CCPA and GDPR affect businesses whether they reside in these respective states or not. The CCPA affects for-profit companies with over $25 million annual revenue or firms gathering the personal information of at least 50,000 or more California residents, whether the business is based in California, another state, or overseas. The GDPR affects all companies with more than 250 employees who do business in the EU or with EU residents, collecting and processing EU residents' data, whether the business is based in the EU or another continent.

One element of the distinctiveness between the two is that GDPR is that forces companies to ask an explicit opt-in. This implies that firms needed to outline a direct communication to users to obtain their consent. This represents a unique opportunity to observe and monitor strategies used to convey that opt-in. When asking for consent, firms should specify the purpose of using personal data (e.g., targeting, marketing, etc.), mentioning all third-party vendors who could process individual information and are recommended to ensure transparent and clear communication. GDPR requires that consent should be explicit and represent a genuine choice. For this reason, our empirical analysis is focused on GDPR, although we contend that our findings have a broader spectrum because, even if each government aims to protect personal information of their residents, data has no boundaries in a digital world (ValeoNetworks 2020). Therefore, in the next section, we are going to focus the attention on the main principles and characteristics of the GDPR.

Figure 2.1.3 – Data Protection and Privacy Regulation Map.



Source: World Federation of Advertisers (WFA)

2.2. The General Data Protection Regulation

In order to answer to the urgency for better guarantees on data collection and control by consumers, the European Union (EU) has issued the General Data Protection Regulation UE 2016/679 (GDPR) to improve customers’ data protection, providing data owners with a restrictive set of rules on sensitive data treatment and with a higher level of control on the data disclosed. The GDPR was created to substitute the previous directive (European Union 1995), which was no longer able to meet the requirements for privacy of the current digital and mobile world (Tikkinen-Piri, Rohunen, and Markkula 2018). Consequently, the European Union began enforcing the new General Data Protection Regulation starting from May 2018, two years after the EU agreed to a significant reform of its data protection GDPR framework. This reform has been heralded as the world's strongest protector of digital privacy rights (Chen 2018) with an unprecedented *wide territorial scope*. Indeed, although the reform is designed for European firms, it also affects companies operating outside Europe that need to comply with the European GDPR if they collect data from citizens who reside in the EU. Additionally, another important characteristic of the regulation regards the harsh penalties in case of non-compliance. The GDPR promises up to €20 million – or 4% of the firms' worldwide annual revenue from the preceding financial year – in case of serious infringements of the right to privacy or the right to be forgotten, which are the basis of the GDPR.

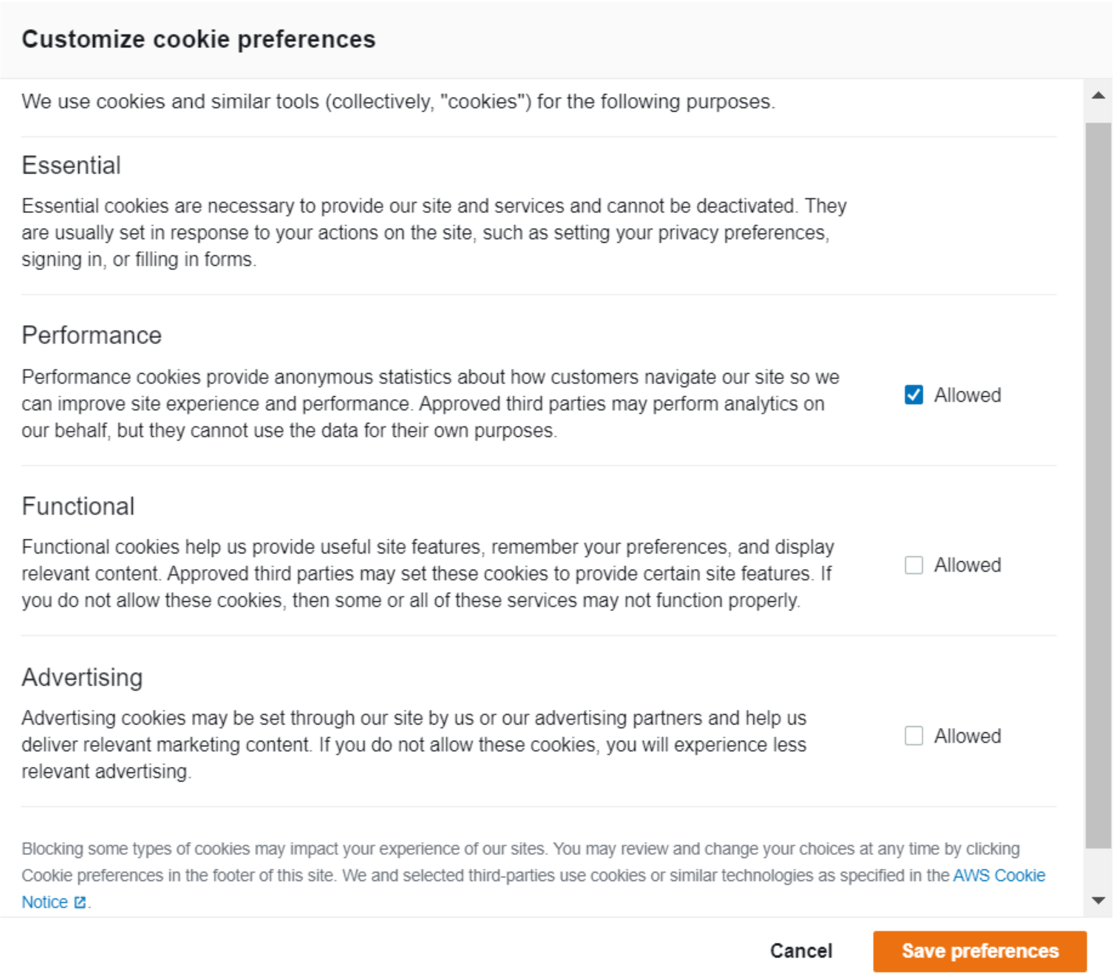
The purpose of the reform was to ensure *transparency* and *control* in the processing of personal data: communications and information provided to individuals must be clear, easily understandable, and accessible (e.g., if data are collected, users should be informed clearly about how their information will be used and who can have access to them). Additionally, the reform accords new rights and more control to individuals to manage and protect their data; for example, individuals can ask firms for an electronic copy of the data collected about them to check the truthfulness of their information. Even if this is undoubtedly a step forward to relinquish the power over data to customers, the more important shift of this new regulation regards the provision of a higher level of control to customers with the newly introduced "*right to be forgotten*" (European Union 2016, Art. 17), thanks to which individuals can request the erasure of their data or port their data elsewhere – except for data required by companies in order to be compliant with legislative obligations such as data deriving from invoices.

An additional and relevant key principle at the basis of the GDPR is the one of *consent*, extensively treated in the Art. 4. Consent has been defined as:

> *"any freely given, specific, informed and unambiguous indication of the data subject's wishes by which he or she, by a statement or by a clear affirmative action, signifies agreement to the processing of personal data relating to him or her"*

Consequently, according to it, firms need to create a clear and concise request that allows data owners to freely decide about the willingness to grant their permission to use and share the data collected. The consent should be explicit, evident, distinguishable from other requests, and specific for the particular use of the individual's personal data (e.g., profiling, targeting, marketing), and should mention all third parties who could process individual information. An example of the different consents required for cookie usage is provided in Figure 2.2.1, where it is possible to see that individuals can clearly understand which are the main purposes of the data collection and can decide whether allowing the company to use and process their data for different aims.

Figure 2.2.1 – Example of the Cookies' Consents Required by Amazon Web Services (AWS).



Customize cookie preferences

We use cookies and similar tools (collectively, "cookies") for the following purposes.

Essential

Essential cookies are necessary to provide our site and services and cannot be deactivated. They are usually set in response to your actions on the site, such as setting your privacy preferences, signing in, or filling in forms.

Performance

Performance cookies provide anonymous statistics about how customers navigate our site so we can improve site experience and performance. Approved third parties may perform analytics on our behalf, but they cannot use the data for their own purposes.

☑ Allowed

Functional

Functional cookies help us provide useful site features, remember your preferences, and display relevant content. Approved third parties may set these cookies to provide certain site features. If you do not allow these cookies, then some or all of these services may not function properly.

□ Allowed

Advertising

Advertising cookies may be set through our site by us or our advertising partners and help us deliver relevant marketing content. If you do not allow these cookies, you will experience less relevant advertising.

□ Allowed

Blocking some types of cookies may impact your experience of our sites. You may review and change your choices at any time by clicking Cookie preferences in the footer of this site. We and selected third-parties use cookies or similar technologies as specified in the AWS Cookie Notice.

Lastly, firms should be transparent with the data owners by informing them about the consequences that may derive from their choice of consents, and they should make it easy to revoke the permission at any time when the users required it. Consequently, the construction of the data consents' management system is crucial from the firm's point of view since the GDPR establishes that it is up to the companies collecting data to demonstrate that the data owners have clearly released their consents about the specific data treatments.

All in all, GDPR requires that consent should be explicit and represent a genuine choice of the data owner for the firm to be compliant and not incur the penalty described above.

However, if the GDPR provided detailed rules and principles that companies should follow to get consent, it did not explicitly define how firms should craft the communications aimed at obtaining users opt-in. This is intriguing since we observed that the different companies' communications present various elements intentionally used as leverages to get customers' permission to use data and maximize data disclosure. For example, certain firms used persuasive arguments and provided financial incentives – such as discounts – in exchange for data. Notably, this is not in contrast with the EU Regulator's request to make citizens explicitly decide whether to grant their data. Nonetheless, we contend that the use of persuasive arguments could defeat the purpose of the policy, as consumers might disclose their data because of the persuasion and not of the increased awareness.

2.3. Summary

As described in the previous sections, the legislative world is moving towards the provision of higher protection and safeguards for the consumers, limiting the power of companies (Tikkinen-Piri, Rohunen, and Markkula 2018). Customers are increasingly aware of the possible negative consequences of the firms' data collection and have requested some advancements in privacy policies worldwide. One of the most recent and stringent privacy and security law in the world is the General Data Protection Regulation, a European law that specifically aims at providing customers with more control over the data shared with firms and at posing restrictions on the usage and storage of users' information to companies dealing with EU citizen data. For example, according to this new regulation, empowered consumers can now delete all their actual and historical data thanks to the "*right to be forgotten*" (European Union 2016, Art. 17); this can deprive companies of the "*food*" for their algorithms and cause them both the loss of any competitive advantage (Peppers and Rogers 2016) and the impossibility of delivering a customer experience of superior quality.

In order not to lose data, the GDPR asked companies to obtain explicit consent (e.g., opt-in) to use and share the data collected from their customers and not incur harsh penalties. Consequently, companies that were either based in EU countries or that had EU citizens' information in their databases sent *millions* of emails, around May 2018, to both inform their customers or prospects about the key changes of the GDPR and to request their permission to use or continue using their data.

The main concrete risk firms were facing was that these emails went unread or were considered as spam. Given that the reform did not impose a specific format of the re-permission email and firms were free to develop their own campaigns, we observed that companies used different strategies to attract consumers' attention and to encourage individuals to opt-in (Davis 2018). For example, some firms focused only on providing *clear* and *understandable* information about the novelties introduced by the GDPR and on the possibility of managing the data disclosed, making a strong effort to be compliant with the principles of *transparency* and *control* on which the regulation is built. Others, instead, decided to frame the request differently, providing data owners with some sort of *incentives* – either monetary or non-monetary – and adding, to the primary informative nature of the communication, a persuasive element that increases the likelihood of the data permission.

This suggests that the content of GDPR re-permission e-mails is a key phenomenon that allows studying the role of persuasive elements in exchange for data. The content and design of these re-permission messages are crucial aspects of this work and will be treated in the central part of this thesis (Chapter 6).

3. THEORETICAL BACKGROUND

This chapter aims at describing the state of the art of literature on privacy in marketing, management, and economics. It will help better understand how the theme of privacy and data protection has been found to impact both customers' behaviors and companies' online strategies in the nowadays hyper-connected and digitalized world.

While revising this vast body of work, we focused our attention on the two main literature streams, which are more in line with this work's aim.

First, studies that have identified the main factors influencing individuals' likelihood to allow their data to be collected and used by firms (e.g., data disclosure behavior). These studies are described in the first section of this chapter and are mainly concerned with the implications of different communication stimuli on customers' disclosure behaviors.

Second, works that have studied the impact of privacy policies on both firms' actions and returns. These studies are outlined in the second part of the chapter and mainly intend to enlighten about the possible economic consequences of data protection regulations on the online ecosystem.

The chapter ends with a summary of the main results achieved in both the streams of literature and highlights the gap that this thesis aims to fill.

3.1. Introduction to the Theoretical Background

As stated in the previous chapters, the main aim of the present thesis is to study the GDPR re-permission email phenomenon by analyzing their textual content and unveiling some of the characteristics of the firms that decided to design these particular types of communications in specific ways.

With the advent of the GDPR, the EU Legislator forced companies to disclose their data protection practices and inform customers about the data collected, used, and transmitted to third-party vendors. The main aim was to reestablish the data power into the data owners' hands and make firms accountable for their data collection practices. However, if the EU Regulator was very specific in detailing which types of information companies should provide to their users, it did not regulate the content of these communications, leaving firms free to develop their data-related communication strategies. This meant that firms could potentially supplement the informative nature of these communications – required by the law – with additional communication elements proven to mitigate users' privacy concerns and ensure data disclosure behavior.

In this data-based economic scenario in which companies are increasingly relying on data to generate value, firms may be worse off when privacy protection regulations are enforced. For example, the GDPR requirement of explicitly disclose individual firm's data collection and management strategies may generate the risk, for the company, of inducing feelings of reactance in its consumers, especially if they were not aware of it (Dinev and Hart 2004; Fitzsimons and Lehmann 2004). Previous studies have demonstrated that reactance, indeed, impacts both on click-through rates and ad effectiveness (Aguirre et al. 2015; Bleier and Eisenbeiss 2015), reducing firms' profitability. Moreover, companies can experience reputational issues when privacy protection procedures are violated. Data-breach announcements have been shown to impact customer's trust in the company exposed (Acquisti, Friedman, and Telang 2006; Janakiraman, Lim, and Rishika 2018; Martin, Borah, and Palmatier 2017), which means that firms, when dealing with data, run the real risk to generate lower returns or to be fined with important sanctions by the Regulator (Bleier, Goldfarb, and Tucker

2020). Lastly, marketing research has also shown that there can be severe competition problems in the digital ecosystem, meaning that leading companies, when regulations are introduced, will have the "*bigger piece of the pie*" at the expense of small firms operating in the same sector (Johnson and Shriver 2019; Sharma, Sun, and Wagman 2019). Therefore, literature shows that the advent of privacy regulations has influenced how companies perform, interact with consumers, and compete with each other. To counteract these adverse collateral effects, firms may have turned the privacy-related communications – required by the GDPR – into opportunities and used them not only to inform but also to influence individuals' propensity to disclose.

We argue that "*how*" companies designed their privacy-related communications is, indeed, related to their data-related interests. As suggested by DellaVigna and Gentzkow (2010), the communication design is not independent of the motivations and the interests of the communication's sender:

> "*… a large share of the information on which economic and political decisions are based is provided by agents who themselves have an interest in the outcome. Information about products is delivered through advertising by the sellers, political information comes from candidates interested in winning election, and financial data is released strategically to shape the perceptions of investors. Third parties that might be more objective – certifiers, media firms, financial analysts – have complex incentives that may diverge from the interests of recipients.*" (DellaVigna and Gentzkow 2010, p. 664)

Therefore, in studying GDPR re-permission emails, it is necessary to consider that companies – when deciding the content of their re-permission emails – may act in a self-interested way depending on their specific economic objectives. For example, for a company that has based its business model on data collection and has profited out of retargeting strategies, it would be of pivotal importance to use communicative tools to prompt users to opt-in. In contrast, firms that have already experienced data breaches and need to re-build reputation and credibility should try to limit communication's distortion and opt for an accurate and transparent type of message. Nonetheless, independently from their inner motivations, companies have been shown to act in order to maximize the odds of realizing what they need to achieve and use all the communication stimuli at their disposal to reach their objectives. Persuasive cues have been found to be used in various contexts: advertising, marketing, politics, finance, economics, and CSR (DellaVigna and Gentzkow 2010; Elving et al. 2015; Habermas 1985; Leffler 1981; Mueller and Stratmann 1994; Taillard 2000; Zaharias 2010). For example, Mueller and Stratmann (1994) have described the properties of persuasive and informative US political campaigns by using economic models. In their paper, they argued:

> "*It is obvious…that persuasive campaign spending is likely to be much more attractive to candidates than informative campaign spending. The latter, unless selectively targeted, must decrease the probability of some groups supporting the candidate, while it increases the probability of others' support. Persuasive campaign spending, on the other hand, holds out the promise of increasing the votes obtained from all groups.*" (Mueller and Stratmann 1994, p. 60)

Additionally, they also stated:

> "*…if campaign expenditures do generate votes, and campaign contributions are dependent on the positions of the candidates, one can expect candidates to take positions based on the expected contributions they will generate.*" (Mueller and Stratmann 1994, p. 63)

Following their line of reasoning, it is the type of campaigns crafted by the candidates to affect their capability of attracting contributions from the voters, which, in turn, influence their chance to win. Nonetheless, political campaigns are designed by the candidate staff, which is motivated toward a prevailing goal: to ensure the candidate's victory. Therefore, the choice of a specific type of political campaign is highly dependent on the main objectives of the candidate crafting it.

A similar rationale can also be used in the privacy realm. Companies, by only informing consumers about their data practices, run the risk that to evocate users' privacy concerns about the data collected and to lose data (as candidates lose voters): users become aware of the way in which companies manage their data and get to know that oftentimes their data are shared with other vendors. However, by using other communication elements that induce users' disclosure behavior, firms expect to achieve data access from the majority of their users and to reach their final aims (as candidates expect to increase their votes and to win the election). Also from the company perspective, thus, it is more appealing to use persuasive themes than informative themes since, through persuasion, companies are more likely to reach what they want to achieve.

We claim that companies will shift their GDPR re-permission email content from informative to persuasive when they have reasons for doing so, which is when they intensively use and profit from data. In other words, in this thesis, we aim to show that also data-related communication follows the same rules that exist in other contexts (such as the political context described above) and that companies will persuasively design their re-permission emails when they see a value in doing so, by carefully evaluating the benefits and the costs that may arise from their communication choices.

In order to prove our point, we turned to privacy literature. It is an incredibly vast realm that comprehends studies from different research streams (e.g., consumer behavior, economics, marketing). Consumer behavior researchers have studied the phenomenon of privacy by analyzing the drivers and the outcomes of consumers' privacy concerns and feelings of vulnerability (Acquisti, Brandimarte, and Loewenstein 2015; Blattberg, Kim, and Neslin 2008; Bleier and Eisenbeiss 2015; Dinev and Hart 2004; Krafft, Arden, and Verhoef 2017; Martin and Murphy 2017; Prince 2018). Marketing researchers have investigated the effect of the introduction of privacy regulations on the digital ecosystem in terms of competition and firms' revenue (Johnson and Shriver 2019; Johnson, Shriver, and Du 2020; Sharma, Sun, and Wagman 2019). Information technology researchers have looked at the phenomenon from a more technical point of view by looking at the effects that privacy may have on AI technology adoption and deep learning methods and by proposing lawful algorithmic solutions to combine the users' need for privacy and security to the data-based innovations offered in the market (Chung, Wedel, and Rust 2016; Georgiadis et al. 2017; Holtrop et al. 2017; Shokri and Shmatikov 2015; Soleymanian, Weinberg, and Zhu 2019; Stahl and Wright 2018).

It is out of the aim of this thesis to provide a comprehensive literature review on the topic of privacy. However, we offer an extensive and exhaustive review of the main studies done in the following two streams, which we found of particular relevance for this thesis goal and research questions:

(i) studies on the *effects of privacy-related communications on individuals' disclosure* behavior (described in Chapter 3.2);

(ii) studies on the impact that *privacy regulations may have on companies' performance* and the entire digital ecosystem (illustrated in Chapter 3.3).

3.2. Privacy and Data Disclosure

Marketing literature has recently started to investigate the theme of customers' personal data disclosure. Thanks to the incredible development of technological tools (e.g., cookies, web beacons, fingerprinting, geo-tracking devices), companies can now track customers' actual behavior online and offline and collect vast amounts of data. This can, in principle, lead to considerable opportunities for companies (e.g., implementation of targeting actions), but it can also heighten customers' feelings of privacy intrusion, which may translate into lower willingness to disclose personal data.

Indeed, given the increasing call for privacy by customers, the EU Legislator has enforced in May 2018 the General Data Protection Regulation (GDPR) with the specific aim to give power back to customers on their data, requiring companies to be transparent about the data collection, usage, and transmission and to ensure high-security standard on the data disclosed.

The GDPR has also forced firms to obtain explicit opt-in from their customers to collect their data, which has resulted in a massive amount of re-permission emails sent by companies around May 2018 with the specific aim of gaining customers' data access. However, even if the GDPR set a lot of costly and strict requirements for the firms, it did not explicitly establish a standard template on which these communications should be crafted, meaning that firms were free to design their privacy-related communications. The Regulator mainly requested that firms develop their communication transparently and move data control into consumers' hands. However, some firms coupled these informative cues with additional elements such as incentives and framing to ease consumers' propensity to concede their data. The first part of the communication was probably in line with the Regulator's intent, whereas rewards or incentives were not. Previous literature has proved that the way in which data permission is asked is relevant in achieving customers' *opt-in*. In the paper by Utz et al. 2019, the authors found that users respond differently to *cookie consent notices* depending on how these have been proposed to them. Presenting the cookie notice in the lower (left) part of the screen increases the user's probability of interacting and providing consent. Additionally, they also proved that providing users with more choices and details about data collection leads to lower consent rates and that the use of technical language (e.g., the mention of the word "*cookie*") leads to higher interaction with the consent notices but to lower consent rates. This provides evidence that users may be nudged differently by the design of these cookie consent notices.

Consequently, it is essential that regulators not only define the specific requirements for companies (e.g., acquiring customer consent before collecting data) but also provide additional guidance on "*how*" these requirements should be operationalized (Forward Action 2018).

The "*how*" factor has been the real focal point for both academic literature and managerial press, which have increasingly proposed tools and factors that managers can exploit as communication's levers to heighten customers' protection perception and lessen privacy concerns and feelings of vulnerability (Martin 2018).

The following sections summarize the main works done about the different elements used in the privacy communication literature.

3.2.1. Control

A significant part of the recent literature has discussed the role of the provision of control on personal data to consumers, focusing on the resulting effectiveness of advertising strategies, such as *retargeting*.

As highlighted by Blattberg, Kim, and Neslin (2008), privacy concerns can be suppressed or, at least, inhibit through the concession to consumers of some kind of control on their data. Recent studies have shown that the missed provision of consumers' control over their disclosed data results in increasing levels of privacy concerns and decreasing propensity to purchase (Dinev and Hart 2004; Malhotra, Kim, and Agarwal 2004; Phelps, Nowak, and Ferrell 2000; Xu et al. 2012).

Additionally, if people know that they can control their privacy online, they are not only more willing to provide personal information but also to react positively to personalized ads and data breaches. For example, Tucker (2014) analyzed data by Facebook in the time frame in which the social media platform changed its privacy policy, implementing the possibility for the user to control the level of data disclosed publicly. The findings showed that personalized ads were twice as effective after the policy change. This effect was even larger for ads using more personal information and for target groups more likely to use opt-out privacy settings, highlighting the strong beneficial impact of the control factor in helping people perceive data collection as less intrusive and grant the firm more personal information.

Moreover, as stated in the paper by Martin, Borah, and Palmatier (2017), control can have a positive effect on returns in case of data breaches; this, again, supports the thesis that providing customers with some level of control – opt-in or opt-out – can have a crucial impact in cases of data violations since it can moderate consumers' concerns related to privacy.

However, even if control is an important tool for companies to get what they want – data – it can be a double-edged sword for consumers who often cannot thoroughly evaluate what they are granting consent to. This has been highlighted in academic literature as the phenomenon of *control paradox*, which mainly points to the fact that people, when in control of their information, feel less vulnerable and are, hence, more likely to give the possibility to access and use their sensible information also when it is highly risky (Acquisti, Adjerid, and Brandimarte 2013). Studies have proved that perception of control can lead the consumers to be more willing to answer personal questions, even when the risks of disclosure to strangers are higher (Brandimarte, Acquisti, and Loewenstein 2013; Stutzman, Gross, and Acquisti 2013). For example, individuals operating on Facebook have shown increasingly privacy-seeking behaviors regarding information shared publicly, but they also have exhibited a rise in the amount of personal information revealed on Facebook privately. This means, on one hand, that people have high concerns about the provision of personal information to strangers and, on the other hand, that the possibility to decide that some information remains private lessen these preoccupations; however, the private disclosure is not entirely confidential since it allows other entities (e.g., third-party apps, advertisers and Facebook itself) to access and use personal data often without awareness or explicit consent (Stutzman, Gross, and Acquisti 2013).

To summarize, all the studies cited above point to the main conclusion that the provision of consumers' control over their data in the firms' privacy communications can be effectively used to gain a higher level of opt-in for data collection, usage, and sharing.

3.2.2. Transparency

Another significant body of literature has examined the impact of transparency on customer's disclosure behavior.

Studies have shown that if firms covertly collect information about their users (Peppers and Rogers 2016) and implement personalized advertisements based on that, they risk to heighten customers' feelings of vulnerability and experience a phenomenon of *reactance* because customers get to know that data has been collected and used without their explicit consent (Aguirre et al. 2015; Tucker 2014). If users understand how their data are used, they are more willing to disclose sensitive information (Aguirre et al. 2015; Benson, Saridakis, and Tennakoon 2015). If, instead, users become aware of the collection and use of their data only when they receive a data breach announcement on their data or a personalized advertisement, then the feelings of vulnerability and the consequent privacy concerns can have a substantial impact, respectively, on the returns of the firm experiencing the privacy failure and on the effectiveness of the ad. On this latter effect, research has proved that people are increasingly showing signs of reactance to targeting strategies, meaning that the more they feel vulnerable, the more they react negatively either to the firms' messages targeted to them making use of highly personal information (White et al. 2008) or to the firm's personalized ads perceived as highly intrusive (Tucker 2014). Being transparent with customers about data acquisition and usage can, then, have positive externalities for firms.

Moreover, as highlighted in the paper by Acquisti, Adjerid, and Brandimarte (2013), the implementation of easy-to-read privacy notice can be a tool that companies can use to allow their consumers to make better decisions about data disclosure. Additionally, research has also shown that customers appreciate and reward the provision of privacy statements (Hui, Teo, and Lee 2007) even if they rarely read them (Farrell 2012; Milne and Culnan 2004).

However, if theoretically, the better the readability and usability of notices, the more the consumers' disclosure of information, in practice, it is not always the case (Adjerid et al. 2013). Other factors, such as the framing of the notices – high or low in protection – or the delay between the notice and the data request (Adjerid et al. 2013) or the customers' privacy concerns, trust, and comprehension of the privacy notices (Milne and Culnan 2004) might also play a role in affecting and predicting individuals' disclosure behavior.

Academic literature has tried to discuss the effects of transparency on data disclosure. Findings show, for example, that transparency can, together with control, suppress both the positive impact of vulnerability on emotional violation and its negative effect on cognitive trust (Martin, Borah, and Palmatier 2017). In a similar study, also Athey, Catalini, and Tucker (2017) proved that transparency could play a role in technology adoption; they, essentially, found that providing students with irrelevant but reassuring information on the information treatment encryption – hence, granting more transparency – led to less concern about surveillance among students meaning that transparency can impact students' behaviors.

These results suggest that providing a clearly stated privacy policy can bring to trust and lower levels of privacy concerns, which, consequently, lead to a higher probability of data disclosure.

Additionally, transparency seems to have an impact not only on users' disclosure behavior but also on customers' purchase intentions. In a recent study by (Mohan, Buell, and John 2019), it has been found that transparency – in particular cost transparency – is effective in increasing customers' purchase intention when the firm voluntarily discloses sensitive information.

However, there is also work showing that being transparent not only leads to positive effects for firms. For example, Janakiraman, Lim, and Rishika (2018), in their study on data breaches, found that being transparent with customers who were breached, sending, for example, an email to communicate that their data have been compromised, results in more relevant feelings of vulnerability, which then translate into less customer spending.

Consequently, transparency can be perceived as both a cost or a benefit for the consumers: if some studies have claimed that greater transparency translates into a higher level of perceived vulnerability, inhibiting data disclosure, other researchers have, instead, highlighted that transparency positively relates to feelings of security and trust, enhancing data provision.

Lastly, it has also been shown that transparency may affect the behavior not only of the recipient but also of the sender of the communication. A recent study by Guo, Sriram, and Manchanda (2020) found that when required by the law to disclose the payment received by pharmaceutical companies, physicians behave differently both in terms of the number of prescriptions and types of drugs prescribed (generic vs. branded). This can, in principle, also mean that the simple requirement to be more transparent by the GDPR may have led to a self-monitoring effect on firms' collecting behavior.

To summarize, it is possible to see that most of the above-cited literature agrees on a positive effect of transparency on customer behavior in terms of disclosure of personal information. The adverse effects that can result from transparency seem to be due to other contextual elements of the communication that can interplay with the positive externality of transparency (e.g., the negative effect of a data breach). Consequently, the impact of a higher degree of transparency on customers' decisions to opt-in may be dual: it may foster trust leading to higher opt-in rates; however, it can also be that the communication prompt customer to know more about firms' data practices of which he was unaware before, leading to a strong adverse reaction.

3.2.3. Incentives

Another part of studies in marketing has addressed the phenomenon of data disclosure using a cost-reward perspective showing that people tend to give access to their data if the perceived benefits obtained from the disclosure outweigh its costs (Krafft, Arden, and Verhoef 2017; Thibaut and Kelley 1959; White et al. 2008). For example, in the paper by Krafft, Arden, and Verhoef (2017), they found that there are cost- and benefit-related factors that lead people to be less or more likely to grant permission for interactive marketing. In particular, they discovered that providing messages with entertaining content lessen the negative influence of privacy concerns on the probability of granting permission, and this also holds for the intention to use mobile service and for the integration of new technological tools (Hausman and Siekpe 2009; Nysveen, Pedersen, and Thorbjørnsen 2005). Using the same rationale based on the social exchange theory, the research by White et al. (2008) analyzed the effect of message utility on both reactance and click-through rate. They found that the higher the utility, the lower the reactance and the higher the click-through rate, meaning that, once again, when the personalized ad maximizes the customer's utility – and there are perceived net benefits – he is more willing to provide personal information – which is perceived as a cost.

Consequently, the provision of benefits can be used to "*obscure*" the costs of data disclosure, and some researchers showed that incentives – both monetary (e.g., discounts) and non-monetary (e.g., lottery) - can be effectively used by companies to achieve customers' data. For example, according to the paper by Athey, Catalini, and Tucker (2017), the provision of small non-monetary incentives, such as a free pizza, to consumers can lead them to give away sensitive information easily. Similarly, Chellappa and Sin (2005) have claimed that monetary, in addition to non-monetary, incentives can push customers to give personal and preference information. Additionally, Grossklags and Acquisti (2007) have shown that customers are willing to trade data for money, even for a minimal amount of money. By contrast, in a survey examining German customers' intent to grant permission, Krafft, Arden, and Verhoef (2017) find that permission coupled with monetary incentives and lotteries do not affect consumers' likelihood to release data. These authors conclude that consumers can read efforts to "*buy*" permission as manipulative and, this might trigger negative feelings like reactance.

To summarize, this whole literature agrees on the proposition that customers are willing to provide information when benefits outweigh the costs of data disclosure. In other words, customers are susceptible to persuasive arguments that highlight gains from the *data transaction,* and companies can exploit the use of incentives to boost customers' disclosure behavior and opt-in. However, care must be taken to interpret the effectiveness of incentives only positively since they may be negatively affected by the surge of reactance.

3.2.4. Framing

Another important stream of studies in the marketing field has, instead, begun to apply "*Prospect Theory*" (Kahneman and Tversky 1979) as a conceptual framework to study the decisional process of data disclosure. It has been shown that people, when deciding on whether to disclose or keep private personal data, act differently depending on the context in which they are embedded. For example, it has been shown that the endowment effect (Gamliel and Herstein 2007; Levin, Schneider, and Gaeth 1998; Thaler 1980) plays a crucial role when privacy decisions should be taken. This hypothesis mainly states that the goods endowed are valued more than the ones not included in the endowment, meaning that if something has to be removed from the endowment, this will be perceived as a loss and, consequently, this will loom larger than the insertion of the same good, which instead is perceived as gain.

For example, the paper by Grewal, Gotlieb, and Marmorstein (1994) supports the existence of this effect, highlighting how the negative framing of the message strongly affects the relationship between the new product's price and the associated perceived performance risks. In another context, Grossklags and Acquisti (2007) have shown that when people should decide about their privacy, there are two possibilities for companies: offer them to pay for protecting their privacy (willingness-to-pay) or offer them money to get access to their information (willingness-to-accept). In the paper, they found a strong preference for money even when the amount offered is tiny: most participants decide to sell their information for 25 cents and decline to pay for protecting their personal data for the same amount of money. Similarly, another study by Acquisti, John, and Loewenstein (2013) supports the existence of the endowment effect; besides, they also showed that the order of the privacy options presented to the subjects influences the privacy decisions. In the experiment, subjects were approached by the researchers at a store of a shopping mall asking to fill in a survey in exchange for a coupon (10$ or 12$ worth) redeemable at the exit of the store; once people get back the coupon to the researchers they were presented to one of the following two situations: in the case, they were prompted with the 10$ coupon, people were then offered the possibility to exchange it with the 12$ coupon with the downside that the data collected through the survey would have become identifiable, while in the case of the 12$ coupon, people were then offered the possibility to exchange it with the 10$ coupon if they want their data to be anonymous. Findings suggest that people's care for privacy strongly depends on the context since subjects were more likely to reject cash offerings for their data in cases where they perceived that their privacy was protected by default.

All in all, the literature described till now is consistent with what "*Prospect Theory*" predicts: loss looms larger than gains for customers' disclosure decisions. In a similar vein, it is also possible that the framing effects play a role in the privacy communications' opt-in results. It can be that the request for data disclosure highlighting the negative consequences – the lack of benefits – in case of denial of data provision can have a more substantial positive impact on the customer's disclosure decision than in cases of requests of data constructed emphasizing the positive outcomes.

However, as also found in different literature streams, the reverse can be true (Hanson and Yun 2018; Ku, Yang, and Chang 2018; Rothman et al. 2006). For example, it has been found that framing the message highlighting losses instead of gains can lead to negative perceptions by customers who feel they had been treated unfairly by the company (Ku, Yang, and Chang 2018). People, when at threat of losing some service provided by a firm, can perceive a lack of appreciation from the firm itself and, hence, experience a phenomenon of reactance. Similarly, but in a different context, it has been found that the addition of positive ingredients to the nutritional elements' list of a new product announcement has a positive and significant main effect on the returns of a company. In contrast, there has been no effect removing a negative ingredient, highlighting, once again, how consumers react more positively towards gain-focused claims (Hanson and Yun 2018). Lastly, in health studies, it has been shown that the type of behavior plays a crucial role in determining the relevance of the effects of gain or loss-framed messages (Rothman et al. 2006). If the action is focused on preventing some adverse outcomes (e.g., the use of the sunscreen to avoid skin cancer), then a gain-framed message will be more effective (Detweiler et al. 1999), while the reverse is true if the behavior is focused at detecting some negative outcomes (e.g., women engaging in Breast Self-Examination) (Meyerowitz and Chaiken 1987).

In summary, this stream of literature indicates that framing may also play a role in privacy-related decisions. It can be that the request for data disclosure highlighting the negative consequences – the lack of benefits – in case of denial of data provision have a stronger positive impact on the customer's disclosure decision than in cases of requests of data constructed emphasizing the positive outcomes.

3.3. Privacy Regulations and Firms' Performance

The vast amount of data, made available by economic and automatic technological tools, is now considered a real competitive advantage for companies and marketers. As stated by McAfee and Brynjolfsson (2012), managers "*can measure, and hence know, radically more about their businesses, and directly translate that knowledge into improved decision making and performance … can make better predictions and smarter decisions … can target more-effective interventions, and can do so in areas that so far have been dominated by gut and intuition rather than by data and rigor*". Consequently, thanks to digital, mobile, and Internet-of-Things (IoT) technologies, marketers have the possibility to exploit the power of data, tracking customer behavior online and offline and gaining specific information on their prospect and actual customer (Bughin et al. 2018; Lohr 2012; Wedel and Kannan 2016).

Access to consumers' information is the real focus of the customer-centric marketing paradigm as it allows for the implementation of targeted actions to create personalized offers of goods or services. Through this type of strategy, it has been shown that firms obtain better-served customers and create a more efficient and effective delivery of customer's value across both digital and non-digital channels (Edelman and Singer 2015; Lohr 2012; Webb 2017).

Companies directly observe the journey that customers make in the digital context. Thanks to cookies, web beacons, and other tracking technologies, marketers can silently "*follow*" the customer online, recording all the websites visited, the products seen, clicked, and bought, the advertisements shown to him, and, eventually, also the action that was taken after the ad was shown. This opportunity allows companies to implement behavioral advertising practices that enable them to select more relevant advertisements for a specific customer, given his previous searches online (Boerman, Kruikemeier, and Zuiderveen Borgesius 2017). This, in turn, has been shown to lead to higher purchase probabilities and advertising revenues (Aziz and Telang 2016; Manchanda et al. 2006), higher sales, and click-through rates (Bleier and Eisenbeiss 2015; Farahat and Bailey 2013; Lewis and Reiley 2014).

Many companies are, hence, strongly dependent on data and data-driven strategies. Both managerial and academic literature has shown that companies using data-driven strategies are more profitable and productive than their competitors, which showed a low reliance on data (Groenfeldt 2015; Lohr 2012; McAfee and Brynjolfsson 2012). Additionally, as highlighted by Bughin et al. (2018), the rise of digital ecosystems, which are heavily reliant on artificial intelligence, and, hence, data collection, will account for more than $60 trillion in revenues by 2025. This consequently stresses the increasing importance of digital information in today's economic environment and suggests that data collection and use is no more an optional resource for firms.

However, if data are undoubtedly a source of revenue for firms, its collection and usage are also strictly related to privacy issues, which have been addressed multiple times in the last decades with the enforcement of privacy laws worldwide. Legislators of different countries have tried to propose regulations that assure customers data protection and higher security standards: the European "*E-Privacy Directive*" (2002), the "*AdChoice*" program (2010), the European "*General Data Protection Regulation*" (2018), and the American "*California Consumer Privacy Act*" (2020). All these regulations tried, to different degrees, to protect individual privacy, reducing the collection of online information and requiring companies to obtain consent from customers for data collection, usage, and sharing.

As it is possible to imagine, these same regulations may have strong collateral effects for the online advertising industry, which is mainly based on data and automated and programmatic technologies to deliver relevant ads to customers. Privacy policies can have "*unintended but unavoidable*" consequences for the structure of the advertising industry, leading, among others, to problems of competition (Berry, Gaynor, and Morton 2019; Brill 2011; Johnson and Shriver 2019; Libert, Graves, and Nielsen 2018; Peukert et al. 2020). As stated in different papers (Jin and Wagman 2020; Peukert et al. 2020), it is crucial to evaluate the interplay and the implications that consumers' protection and antitrust laws have on each other: data protection regulations are created to support customers in the customer-to-firm relationship and to give power back to customers on their personal data, but this usually implies considerable costs for firms that have to comply with the new regulation (Jay 2017), leads to higher barriers to entry to entrants in the market and inhibits innovation (Lambrecht and Tucker 2015; Miller and Tucker 2009, 2011). Additionally, large firms can exploit their reputation and their broader range of services to obtain consent – and, thus, data – more easily than smaller firms (Campbell, Goldfarb, and Tucker 2015). This results in markets that are more concentrated on few dominant firms that have the "*biggest piece of the pie*" (Johnson and Shriver 2019; Peukert et al. 2020; Sharma, Sun, and Wagman 2019).

This tradeoff between customers' privacy concerns and the online ad sector's profitability has been a real focus for the regulators, who want to protect customers' privacy without harming the online industry. This has led to an increasing call for additional empirical proofs to better evaluate which have been the impacts of the different privacy regulations on the online sector. Consequently, studies have tried to assess and to quantify the economic impact that privacy policies had on the web, in terms of venture capital investment (Jia, Jin, and Wagman 2019; Lambrecht 2017; Lerner 2011), technological diffusion (Miller and Tucker 2011), ad effectiveness (Goldfarb and Tucker 2011), web traffic and revenue (Aridor, Che, and Salz 2020; Goldberg, Johnson, and Shriver 2019; Marotta, Abhishek, and Acquisti 2019) and price per impression (Beales and Eisenach 2014; Marotta, Abhishek, and Acquisti 2019).

Given that GDPR has been addressed as the most comprehensive, globally leading privacy regime (Peukert et al. 2020), the main focus of the next sections will be on the impacts that this new regulation had on:

- **Companies' online business models** – mainly based on tracking technologies.
- **Players of the digital advertisement market** – publisher, advertisers, and ad network in general.

3.3.1. Impact on the Firms' Ability to Track Consumers' Behavior

Digital companies increasingly base their business models on data collection, usage, and sharing. Consequently, the GDPR can have substantial negative impacts on online firms since it can reduce the number of data they can collect; this can happen because firms decide to do so – in order not to incur the severe sanctions forecasted by the GDPR – or because customers choose not to provide consent (Goldberg, Johnson, and Shriver 2019). The GDPR is the first regulation that requires companies to obtain an explicit opt-in from customers; before it, instead, the norm was the so-called "*notice and consent*" which required customers to opt-out when not willing to be tracked.

The use of data is also vital for the online advertisement sector since it allows users to get information about what users do and like on the web and be more relevant in terms of advertisement shown. This practice is often addressed as behavioral tracking and, most of the time, happens through cookies. Cookies are pieces of code embedded in the website's HTML code and are downloaded on the user browser once the website is loaded. The cookie contains a unique identifier that companies use to identify the customer on the different websites he/she visits, allowing the recording of the customer's history of online browsing. Consequently, cookies are, as of today, the central resource used by companies to implement digital strategies and to generate online profits.

Some studies have shown that online companies made extensive use of cookies in the pre-GDPR era. In their paper, Libert and Nielsen (2018) find that, as of the first quarter of 2018, news websites use a wider variety of third-party domains and have a higher number of third-party content and cookies than popular websites. Similarly, the paper by Sørensen and Kosta (2019) has shown that private websites have more third-party URLs than public websites – this is especially true for the "*private news*" websites. In another study, Iordanou et al. (2018) find that four months before the enforcement of the GDPR, 3% of the traffic between the user and the web tracking service concerned personal information.

There is, hence, evidence that companies are consistently using these tracking technologies to get sensitive and valuable customer information, and it is not surprising that the enforcement of the GDPR has seriously worried firms in the online sector. Different studies have proved that the GDPR had a significant impact on the number of cookies that companies were able to collect. In subsequent work, Libert, Graves, and Nielsen (2018) showed a sharp 22% decrease in the number of third-party cookies recorded on the news websites, with significant losses for advertising, marketing, and social media cookies categories. These drops were recorded differently among the seven countries considered and have spillovers worldwide (Peukert et al. 2020; Sanchez-Rola et al. 2019). The same results were achieved by Sørensen and Kosta (2019), who found that the GDPR led to fewer third-party URLs on both private and public websites with a stronger decline for the private websites (entertainment, news, and travel websites categories). Peukert et al. (2020) found additional evidence that the GDPR reduced the number of third-party cookies, but they also highlight that there has been a sustained increase in the use of first-party cookies by websites, providing evidence for a "*substitution effect*" between the two types of tracking strategies. In this line, Johnson and Shriver (2019) found that the GDPR brought about a 15% drop in the number of relationships between the website and third-party vendors. However, this decrease was only short-lived and got back to the original levels by the end of 2018.

Consequently, the ability to track customers had been strongly affected by the GDPR. Cookies, which were pervasive in the pre-GDPR era, have significantly been reduced to adhere to the "*data minimization*" principle of the European regulation (European Union 2016, Art. 5(1)(c), 25(1), and Recitals 78, 156). However, the compliance of firms to GDPR principles should not be taken for granted. Research has shown that third-party domains, in the majority of the cases, are not blocked *a priori* meaning that third-party domains are still loaded without user consent (Johnson and Shriver 2019). The same result has also been reported by Sanchez-Rola et al. (2019), who found that third-party interactions happen before the user explicitly opt-in. Additionally, also Degeling et al. (2018) found that even if there has been an increase in the number of websites using "*cookie consent notices*", few websites really offer their users a real choice with regards to cookie-based tracking and are still operating on an opt-out consent mechanism.

In summary, research is quite consistent about the impact that an "*opt-in*" type of privacy regulation can have on firms' ability to track users on the web. To various degrees, companies have decided to decrease both the number of cookies and third-party vendors they are using on their websites. This, however, can be seen mainly as a measure that companies have used to avoid the severe sanctions promised by the GDPR Legislator – up to 4% of the previous-year turnover – since some studies point to a short-run effect that recovers over time. Instead, what seems to be more consistent over time is an increased concentration of the third-party web technology market, which sees the failure of small vendors and the establishment of the leading providers. Forcing companies to reduce the amount of data collected – "*data minimization*" principle – seems to have led to market failures more than customers' data protection.

3.3.2. Impact on the Online Advertising Players

Given the results highlighted in the previous section, the GDPR has impacted companies' ability to harvest customers' data, at least in the short run.

This, in principle, has harmed the behavioral tracking strategies that firms and advertising companies can implement. Data availability is not the main end for companies, but it is a means to get higher performances. Through data, advertising companies can track customers and deliver better content and ads, which results in higher effectiveness and returns (Goldfarb and Tucker 2011). If data are no longer available, not only online firms are going to lose (Goldberg, Johnson, and Shriver 2019), but also customers can experience negative externalities in terms of both ad relevance and online free services and contents (Castro 2010; D'Annunzio and Russo 2020).

Hence, privacy protection can strongly affect the online advertising sector, which is essentially based on tracking technologies and, thus, cookies.

This sector comprises different actors: the publishers – who sell advertising spaces – the advertisers – who buy advertising spaces – and the ad exchange networks – which are online platforms that allow the interaction between advertisers and publishers (e.g., DoubleClick, Right-Media). Most of the transactions happening between the publishers and the advertisers on the online platform are nowadays based on open-auctions (Marotta, Abhishek, and Acquisti 2019): advertisers engage in real-time bidding for an impression, and the ad exchange takes care of running the auction and determine the closing price at which that same impression has been sold. The real turning point is that thanks to the ad exchange's ability to track customers on the web, advertisers are ready to pay a higher price per impression, shifting from "*low-value remnant impressions to more targeted and valuable impressions*" (Johnson 2013, p. 7). Research has found that "*cookied"* impressions are more expensive than cookie-less impressions: Beales and Eisenach (2014) estimated that the use of impressions with cookies brings to a 66% increase in the CPM relative to impressions without cookies; the same results were also achieved by Marotta, Abhishek, and Acquisti (2019). Additionally, the cookies' duration may affect the evaluation of the impression since studies have found that older cookies are more valuable given their ability to store more information (Beales and Eisenach 2014; Miller and Skiera 2017).

Consequently, thanks to cookies and other tracking technologies, ad networks can retain detailed information about websites' users and classify them into specific segments of audiences. This is crucial for advertisers since, with this unique identifier placed by the cookie on the users' browser, they can extract the customer's past browsing history and correctly estimate the value of their advertisement for the selected customer (Aziz and Telang 2016). All in all, the ability to track customers seems to bring about higher revenue for all the online ad industry players since it allows for a better and more precise match between advertiser and user (Johnson 2013; Sharma, Sun, and Wagman 2019).

Thus, the implementation of privacy regulation on data collection and usage has the real potential to impact the whole sector negatively. The loss of cookies would likely affect the revenue of the online ad ecosystem as a whole: publishers are no more able to attract higher bids from advertisers, while advertisers have lower returns from advertisement since they are no more able to correctly propose the right ad at the right customer in the right moment.

To the best of my knowledge, the first research on the impact that privacy regulations have on the internet advertising industry was proposed by Johnson (2013). The author studied the effect of different types of privacy regulations on the publishers' and advertisers' revenues and found that both publisher's revenues and advertiser's surplus drop significantly under opt-in and tracking ban policies while little losses are experienced in the presence of opt-out policies. In a subsequent paper, Johnson, Shriver, and Du (2020) found additional evidence of the modest loss of an opt-out type of policies. They studied the impact of the AdChoice program and found that the inability to track opt-out users results in a loss of 8$ per opt-out user. Additional evidence has been highlighted by Marotta, Abhishek, and Acquisti (2019), who found that the presence of cookies allows publishers to increase their revenue by 4%; however, even if this value is significant from a statistical point of view, it has a low economic relevance since they estimate that the 4% gain equates to a 0.00008$ increase in revenue per advertisement.

Instead, the impact of the “opt-in” type of privacy regulation has been shown to lead to more severe losses for the ad industry. Goldfarb and Tucker (2011) found that the “*EU no cookie law*” brought to lower effectiveness of the ads; similarly, also Budak et al. (2014) spotted the same negative impact of the “*Do-Not-Track*” regulation on ad revenues. Cookies have been found to predicts user purchase intentions, leading to higher ad effectiveness (Aziz and Telang 2016); this means that targeting is effective and that the losses deriving from the inability to track and to have customer data may be quite severe. Additionally, it has also been shown that these negative externalities may not be proportional across the players in the online advertising ecosystem. Research has found that there are more severe losses for small publishers and small advertisers, especially when they interact on a smaller and weaker ad exchange network (Sharma, Sun, and Wagman 2019), bringing about, once again, concerns about the effects that this kind of regulations has on the competition in the industry.

However, research has also tried to highlight the gains that these types of regulations can bring about for firms. For example, in the paper by Aridor, Che, and Salz (2020), they found that opt-in policies allow to “*clean*” companies’ databases by the noise produced by the artificial recording of short customer’s histories. Before the implementation of opt-in privacy regulations, customers could protect their online privacy by using ad block technologies. These types of tools do not prevent the company from recording user’s behavior, but they do not let the company connecting all the information available on the same customer over time; in other words, a user who uses an ad-block tool will be recorded as if he/she is a different user each time he/she visits the website. Consequently, companies, which are now prevented from collecting any data on opt-out customers, have more reliable datasets and can make more realistic predictions on the remaining opt-in users. This, in turn, has been shown to increase the advertisers’ bids for the remaining set of opt-in customers.

In summary, the ability to track online customer’s browsing behavior has been the real revolution for the entire online advertising industry. This practice has been mainly implemented with the help of web cookies, which allow the recording of valuable information about online users that can, then, be used by companies for different purposes – e.g., to increase the user experience of a website, to implement marketing actions – which result in higher returns and advertising effectiveness (Goldberg, Johnson, and Shriver 2019; Goldfarb and Tucker 2011; Marotta, Abhishek, and Acquisti 2019). The enforcement of regulations, such as the GDPR, which aims at restricting the amount of data that companies collect in the online sector, can consequently be seen with aversion by companies operating in the online advertising ecosystem. These companies mainly base the success of their business models on the ability to use tracking technologies, and research has shown that negative externalities have been produced by different privacy policy regimes. Some studies have empirically demonstrated that the introduction of a privacy policy law has led to lower revenue for both publisher and advertisers (Johnson, Shriver, and Du 2020; Marotta, Abhishek, and Acquisti 2019); others have proved the same type of losses through the use of economic models (Johnson 2013). Additionally, the inability to use cookies has also been shown to harm the competitiveness of the firms operating in the digital advertising ecosystem (Sharma, Sun, and Wagman 2019).

3.4. Summary

Overall, while revising the vast body of works done in the privacy realm, we focused our attention on the two rich streams of literature related to the research questions presented in the introduction. Table 3.4.1 summarizes the main works in terms of the topics they investigated.

On the one hand, researchers have studied the drivers of individuals' data disclosure behavior, unveiling the most relevant factors which influence consumers' propensity to disclose their data (control, transparency, incentives, and framing). On the other hand, there is research on the impact of privacy policies on both firms' actions and returns.

Our research is unique in considering the combination of (1) communication on privacy's characteristics (informative vs. persuasive) and (2) the impact of GDPR on firms' data harvesting strategies and expected ad revenues. Our work, therefore, will try to shed some light on how firms designed their data request and prove that the use of persuasive themes is driven by firms' self-interest (see Chapter 4).

Table 3.4.1 - Summary of Prior Research on Privacy Communications and Impact of GDPR introduction on Data Harvesting and Revenues.

Papers	How to Request for Personal Data		The impact of the GDPR on:	
	Informative Themes	**Persuasive Themes**	**Harvesting of Personal Data**	**Firms Online Ad Revenues**
Acquisti, Adjerid, and Brandimarte 2013	YES Control & Transparency	NO	NO	NO
Adjerid et al., 2013	YES Transparency	NO	NO	NO
Aguirre et al., 2015	YES Transparency	NO	NO	NO
Benson, Saridakis and Tennakoon, 2015	YES Transparency	NO	NO	NO
Blattberg, Kim, and Neslin, 2008	YES Control	NO	NO	NO
Brandimarte, Acquisti, and Loewenstein, 2013	YES Control	NO	NO	NO
Dinev and Hart, 2004	YES Control	NO	NO	NO
Farrell, 2012	YES Transparency	NO	NO	NO
Guo, Sriram, and Manchanda, 2020	YES Transparency	NO	NO	NO
Hui, Teo, and Lee, 2007	YES Transparency	NO	NO	NO
Janakiraman, Lim, and Rishika, 2018	YES Transparency	NO	NO	NO
Malhotra, Kim and Agarwal, 2004	YES Control & Transparency	NO	NO	NO
Martin, Borah and Palmatier, 2017	YES Control & Transparency	NO	NO	NO
Milne and Culnan, 2004	YES Transparency	NO	NO	NO
Mohan, Buell and John, 2019	YES Transparency	NO	NO	NO
Stutzman, Gross, and Acquisti 2013	YES Control	NO	NO	NO
Tucker, 2014	YES Control & Transparency	NO	NO	NO
Peppers and Rogers, 2016	YES Transparency	NO	NO	NO
Phelps, Nowak and Ferrell, 2000	YES Control	NO	NO	NO
Xu, Teo, Tan and Agrawal, 2012	YES	NO	NO	NO

	Control			
White et al. 2008	YES Transparency	YES Incentives	NO	NO
Acquisti, John and Loewenstein, 2013	NO	YES Framing	NO	NO
Athey, Catalini and Tucker, 2017	YES Transparency	YES Incentives	NO	NO
Chellappa and Sin, 2005	NO	YES Incentives	NO	NO
Detweiler et al. 1999	NO	YES Framing	NO	NO
Grewal, Gotlieb, and Marmorstein, 1994	NO	YES Framing	NO	NO
Grossklags and Acquisti, 2007	NO	YES Incentives & Framing	NO	NO
Hanson and Yun, 2018	NO	YES Framing	NO	NO
Hausman and Siekpe 2009	NO	YES Incentives	NO	NO
John, Acquisti and Loewenstein, 2011	NO	YES Framing	NO	NO
Krafft, Arden and Verhoef, 2017	NO	YES Incentives	NO	NO
Ku, Yang and Chang, 2018	NO	YES Framing	NO	NO
Meyerowitz and Chaiken, 1987	NO	YES Framing	NO	NO
Nysveen, Pedersen, and Thorbjørnsen, 2005	NO	YES Incentives	NO	NO
Rothman et al., 2006	NO	YES Framing	NO	NO
Thibaut and Kelley, 1959	NO	YES Incentives	NO	NO
Aridor, Che and Salz, 2020	NO	NO	YES # of Unique Cookies	YES Advertisers' Bids
Degeling et al., 2018	NO	NO	YES # "Cookie Consent Notices"	NO
Johnson, Shriver and Goldberg, 2020	NO	NO	YES # Third-party Cookies Domains & Vendors	NO
Libert, Graves and Nielsen, 2018	NO	NO	YES # Third-party Cookies & Domains	NO

Libert and Nielsen, 2018	NO	NO	YES # Third-party Cookies & Domains	NO
Sorensen and Kosta, 2019	NO	NO	YES # Third-party Cookies	NO
Peukert et al., 2020	NO	NO	YES # Third-party Cookies # First-party Cookies	NO
Sanchez-Rola et al., 2019	NO	NO	YES # Third-party Cookies # First-party Cookies Cookies' Settings	NO
Aziz and Telang, 2015	NO	NO	NO	YES Advertisers' Bids & Customers' Sales
Beales and Eisenach, 2014	NO	NO	NO	YES Price of the impression
Goldberg, Johnson and Shriver, 2019	NO	NO	NO	YES # pageviews, visits, orders and website revenue
Goldfarb and Tucker, 2011	NO	NO	NO	YES Ads Effectiveness and Returns
Johnson, 2013	NO	NO	NO	YES Structural Model Estimation
Johnson, Shriver and Du, 2020	NO	NO	NO	YES Price of the impression
Marotta, Abhishek and Acquisti, 2019	NO	NO	NO	YES Publisher Revenue
Miller and Skiera, 2017	NO	NO	YES Cookies' Lifetime	YES Cookies' Lifetime Value
Sharma, sun and Wagman, 2019	NO	NO	NO	YES Structural Model Estimation

This Paper	**YES** **Control & Transparency**	**YES** **Framing, Incentives & Time Orientation**	**YES** **# of Marketing Cookies, # Persistent Marketing Cookies, and # Third-party Cookies**	**YES** **Expected Online Ad Revenue Generated by the Website of Firm *j***

4. Conceptual Framework

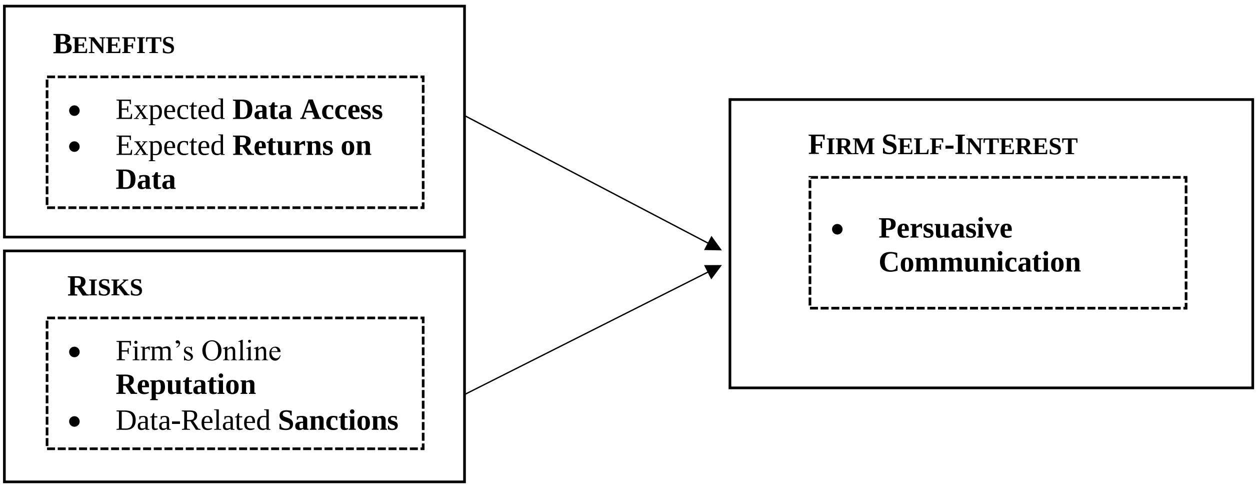
The GDPR asked companies to re-acquire explicit and informed consent to use and trade customers' data. However, the EU Regulator did not specify how consumers’ consent was to be obtained and how the communication between firms and consumers should be designed. This means that firms were left free to craft their re-permission emails to meet their data needs by using not only informative but also persuasive communication elements. We contend that firms behaved in a self-interested way and that the decision to shift towards the use of persuasive themes in the design of these privacy-related emails is strictly correlated both to the benefits from collecting and extracting value out of data and to the risks associated with not being completely compliant with the privacy law.

In this chapter, we introduce our conceptual framework to explain how re-permission emails relate to (1) a firm’s data harvesting and monetization strategy and to (2) a firm’s reputation and possible data-related sanctions.

4.1. Conceptual Framework Development

The conceptual framework depicted in Figure 4.1.1 portrays how the tradeoff between risks associated with not adhering to the law (or its spirit) and the benefits derived from collecting data might influence the surge of firms' self-interest manifested by the use of persuasive themes in GDPR re-permission emails.

Figure 4.1.1 – Conceptual Framework



As stated in the previous chapters, the EU Regulator requested firms to provide consumers information on the nature and use of the data they collected. However, information (re-permission emails' content) comes from agents (e.g., firms) who themselves are interested in what the Regulator wants to protect: data collection and usage. This poses an essential tradeoff for firms under the GDPR Regulation.

On the one side, companies need to be compliant with the new privacy regulation protecting customers' data privacy in order to not incur sanctions or reputational issues; this requires advising the data owner about the firm's data practices and management and asking them explicit consent to access and use their data.

On the other side, the increasing awareness of consumers on data privacy and the possible negative consequences of data disclosure experienced in the last decades (e.g., data breaches) pose a real risk for companies in terms of data access. This means that consumers can now deprive companies of the basic element of their retargeting algorithms. Therefore, companies had to find a way to avoid the expected data loss, which could have been derived from the GDPR enforcement.

As previously stated, the EU Regulator did not explicitly define a standard format for companies' privacy communications, meaning that companies were left free to design the re-permission messages as they prefer and according to their data need. In principle, this means that companies had the chance to exploit the only communication lever that was left under their control – the design of re-permission emails – and to use the re-permission email communications not only as a merely *informative* instrument about privacy and data security – as required by the GDPR law – but also as a *persuasive* tool aimed at maximizing data disclosure.

The distinction between informative and persuasive communication has been extensively used in studies on marketing, advertising, political speeches, and CSR (DellaVigna and Gentzkow 2010; Elving et al. 2015; Leffler 1981; Mueller and Stratmann 1994; Narayanan, Manchanda, and Chintagunta 2003). It mainly addresses the difference in the final goal of the communication itself. In the case of informative communication, the aim of the message is purely to advance the knowledge of the receiver (e.g., customer) by providing information about something (e.g., a product), while in the case of persuasive communication, the aim of the sender (e.g., firms) is to convince the receiver to have a particular point of view and to respond in a specific way to the persuasion (e.g., the purchase).

Notably, as discussed in the paper by DellaVigna and Gentzkow (2010), persuasive communication may act on consumers' behaviors in two distinct ways: (i) by directly altering the receiver prior beliefs about the communication object - by adding new valuable information for the consumer and enhancing message's elaboration and scrutiny, or (ii) by changing individuals preferences through the use of *peripheral* cues – which do not increase argument quality and are mainly used to make consumers' decisions less cognitively burdensome (Cacioppo et al. 1986; Droge 1989; Petty and Cacioppo 1984, 1986; Petty, Barden, and Wheeler 2002; SanJosé-Cabezudo, Gutiérrez-Arranz, and Gutiérrez-Cillán 2009). While the former has been shown to bring about positive externalities for the communication's recipient – who get to have more pieces of information to make a sound choice – the latter does not always lead to higher consumers' welfare – since it conveys no useful information, and it may also prompt receivers to adopt costly avoidance behaviors.

In this thesis, we define a "*persuasive message*" to be a message aimed at influencing/changing the data disclosure behavior of an agent (e.g., the customer) in line with the *preference-based* model described by DellaVigna and Gentzkow (2010). Therefore, we consider as persuasive messages those communications that include *peripheral* elements – such as incentives or other cues aimed at manipulating users' disclosure behavior – that do not directly relate with the main focus of the re-permission email communication (e.g., companies' data-related practices disclosure). Instead, we define as "*informative message*", a message merely aiming to inform and disseminate knowledge about data privacy and protection, in line with DellaVigna and Gentzkow's *belief-based* model.

We contend that there exists a conflict of interest for firms that have to communicate in a "*clear and transparent*" way about their data practices while, at the same time, trying to find a strategy to maintain the data collected and sustaining their data-based operations. Additionally, we argue that self-interest can lead firms to develop communication strategies that aim to get the consumers' opt-in and to turn their re-permission email message from merely *informative* – as suggested by the regulation – to *persuasive*.

A detailed explanation of each of the two main blocks making up our conceptual framework is provided in the following paragraphs.

4.1.1. Data Request and Email Themes

The first research question of this thesis is addressed on the right-hand side of the conceptual framework depicted above and aims at describing how re-permission emails sent by companies worldwide have been designed and at mapping the themes that companies have employed. Consequently, we aim at answering questions such as:

“*Are there any systematic patterns of themes in GDPR re-permission emails?*”

“*Did companies use only informative themes as suggested by the GDPR, or did they also add persuasive elements in the communications to prompt customers to opt-in?*”

“*Are there cases in which both informative and persuasive themes are included in the privacy-related communication?*”

As described in Chapter 3.2, there are indeed multiple ways in which firms can craft the message to be sent to their customers regarding privacy and data disclosure. They can decide to design emails that merely inform customers about their data rights and their possibility to manage the data disclosed, in that creating messages which are completely *informative* and coherent with the GDPR principles. They can also opt for a different type of message, which integrates to the informative nature of these communications, also *persuasive* elements that try to induce consumers’ opt-in behaviors. For example, companies can ask directly or indirectly, highlight the losses or the gains related to disclosure behavior, or insert default settings to encourage people to provide information. Additionally, companies can also use incentives in exchange for data. Marketing literature has investigated if giving customers rewards in exchange for personal data access is a strategy that can help firms’ efforts to collect and store data. However, studies on monetary incentives have brought to conflicting results with regards to data disclosure, meaning that it is not clear and well established how the provision of coupon, money, free premium services, or free samples can effectively push and encourage customers to grant access to personal data (Athey, Catalini, and Tucker 2017; Chellappa and Sin 2005; Krafft, Arden, and Verhoef 2017). In the best-case scenario, it can be that the provision of incentives is merely seen as a benefit and people react positively to it; however, it can also be that people, when asked to trade personal information with money, get suspicious and respond negatively, denying the disclosure of personal data.

All in all, previous research in privacy highlights that privacy concerns and customers’ disclosure behavior may be influenced by several communication factors that can be classified into two broad categories of *themes* that differ with regards to the final aim of the communication (e.g., inform vs. persuade the final user). Understanding how companies have *"played"* with *informative* and *persuasive* themes to obtain customers data access is of utmost importance to shed some lights on how firms respond to privacy regulations and choose to interact with their customers, especially in the nowadays legislative setting, which is increasingly empowering customers and potentially harming firms moving towards data-based strategies. A detailed analysis of the content of the re-permission emails collected can be found in Chapter 6.

4.1.2. Tradeoff Between Benefits and Risks of Data Collection

The left-hand side of the conceptual framework depicted above deals with the study of the circumstances under which companies decide to use persuasive communication elements and to act in a self-interested way. Therefore, we aim at answering the following research question:

"Did Benefits (from data usage) and Risks (of non-compliance) drive Data Requests Content and Intent?"

Designing a persuasive re-permission e-mail poses an essential tradeoff for the sending firm: the need for data gives it an incentive to be persuasive and "*enrich the message*", while the need to comply with the spirit of the regulation and preserve a "*clean reputation"* encourages the firm to be accurate and merely informative. We argue that the strength of persuasion increases when firms ascribe higher value to consumers' data, while, at the same time, the perceived risks of being fined by the Regulator or sanctioned by customers in terms of reputation are low.

Firms aim to exploit the data collected to get higher profits, which is why they assign a specific value to their customers' information. Past work has shown that targeting strategies and personalized marketing campaigns have a higher response rate and can increase customers' profitability (Aziz and Telang 2016; Bleier and Eisenbeiss 2015; Goldberg, Johnson, and Shriver 2019; Goldfarb and Tucker 2011; Marotta, Abhishek, and Acquisti 2019). However, both targeting and personalized marketing activities can be implemented only if consumers allow firms to access and commercially use their data. This can make obtaining permission crucial for firms intensively using data to serve their customers or leveraging on them to derive additional sources of revenues (e.g., ad revenues). For example, companies attracting numerous customers or using several tracking technologies can collect plenty of data, becoming very attractive to advertisers or third-party. Therefore, firms aiming to heavily collect data from their customers and potentially deriving higher revenues from advertisers may have stronger incentives to have used GDPR re-permission emails as persuasive tools, adding monetary incentives – such as coupons and discounts – or using other techniques (e.g., framing messages emphasizing the consequences consumers would face if they were to deny them) to prompt users to opt-in as they are more likely to know their data value.

We contend that the expected return on data feeds a firm's self-interest and leads it to use persuasive arguments.

At the same time, however, adding persuasive cues into the re-permission e-mails can generate some costs for companies. One of them could be the risk of being sanctioned by a regulator wanting to minimize the risk of consumers' manipulation. Notably, however, the EU regulator has left firms free to design their re-permission e-mails provided knowledge on the data collected and on their use was transferred. This means that firms did not face a severe risk of being fined for the content of their re-permission e-mails.

However, companies are increasingly experiencing the *negative externalities* that are related to their data collection practices. In the last decades, there has been an astonishing upward trend in data protection failures, meaning that firms are now facing the real risk that consumers take legal actions against them (Bleier, Goldfarb, and Tucker 2020; Son and Kim 2008). For example, in 2019, both Google and Facebook were accused of having exposed biometric data of millions of users and were claimed to pay millions of dollars to settle allegations (Marotti 2020; The Guardian 2021); similarly, also TikTok has recently agreed to pay 92$ Million to settle the litigation for the same facial recognition technology for which both Google and Facebook were accused of privacy violations (Walsh 2021). Additionally, research has also shown that privacy related litigation has become increasingly frequent in both the EU and US courts (Bleier, Goldfarb, and Tucker 2020; Carson 2020; Solove and Schwartz 2014).

Therefore, companies that have already experienced some of the negative effects of the data collection may behave differently from companies that have not already experienced them. In particular, it can be that past sanctions may have made firms more cautious in the way they ask for data because they have already struggled against the negative consequences that data security problems entail (e.g., in terms of company's trust and reputation). Moreover, under the GDPR legislation, bigger firms can expect greater sanctions from not being compliant with the GDPR principles (e.g., up to 4% of the firm's annual global turnover). This may have led bigger firms to adhere to the GDPR law more strictly and to consider re-permission emails more as a legal type of communications than as occasions to convince consumers to provide data access.

Therefore, providing monetary or non-monetary incentives or framing messages to influence consumers' preferences on privacy can increase a firm's risks both in terms of triggering reactance in customers and of reputation (Krafft, Arden, and Verhoef 2017).

We contend that the costs from not adhering to privacy laws (in terms of firms' reputation or possible data security failures) mitigate a firm's self-interest and lower the use of persuasive arguments in its privacy-related communications.

In summary, we claim that there may be a significant difference in the way companies decide to communicate their "*need for data*" depending on the benefits and the expected costs that may derive from the data collection. The request for data access and the collection and use of data are intertwined firms' decisions, as the value that a company assigns to data influences how they are asked for and shifts the communication towards the inclusion of persuasive cues. At the same time, companies also incur some risk when dealing with customers' data in terms of reactance and reputation. Past experience of these risks may have lead companies to privilege the informative nature of the GDPR re-permission emails, in that using them as a tool to improve their reputation and customer-to-company relationship.

We argue that firms are more likely to run the risks related to the data collection when the benefits they get from obtaining data outweigh the risks that may stem from the way in which data can be obtained (e.g., reactance), that is when firms intend to exploit their customers' data intensively and ascribe higher value to the data collected. A detailed description of this cost/benefit tradeoff and the impact it has on re-permission emails’ design can be found in Chapter 7.

5. ANALYSIS APPROACH & DATA

In the following sections, we provide an overview of the analytical approach we used in order to answer our two main research questions, and we described the creation of the dataset we have used for our analyses, which is mainly composed of two sub-datasets:

- The first one is about the collection of the re-permission emails included in the sample and contains information about the emails' content, language, sending date, and sending company.
- The second one is mainly about the characteristics of the companies which have sent the re-permission emails and contains information about the type of company and its online website.

The merge of these two databases allows us to analyze how firms' self-interest has driven the design of the privacy-related communications used by the different companies on the occasion of the GDPR enforcement.

5.1. Analysis Approach

The methodologies proposed to address the research questions described in the previous chapter are outlined in Table 5.1.1. For a more precise explanation of the methods we used, we reference Chapters 6 and 7, which address the research questions individually.

To answer the research question (1) – *which themes characterize firms GDPR re-permission emails* – we analyzed the content of the re-permission emails sent by a large sample of firms through the following main tools: Natural Language Processing (NLP) techniques, manual content analysis, and the linguistic inquiry and word count (LIWC) program and TextEvaluator online tool. The use of different methodologies allowed us to get a more exhaustive and consistent picture of how companies decided to communicate about privacy and obtain customers' opt-in. Additionally, thanks to the use of different methodologies, we were able to reach reliable results and to develop an efficient and consistent content-analysis procedure that can be used by policy makers or privacy experts to quickly analyze firms' privacy-related communications and to predict the themes in any re-permission email potentially collected.

Then, we addressed research question (2) – *How self-interested were firms?* – by analyzing if and how the *benefits* – that companies can achieve from users' data – and the *risks* – that firms may derive from not completely comply with the EU legislation – may impact the type of communication that the company implemented. Consequently, we complement the dataset by collecting additional information about (a) firms' website's marketing cookies (b) firms' website expected online ad revenues streams (c) firms' website popularity in the pre-GDPR period (d) firms experienced data breaches in the pre-GDPR period. To diagnose the correlations between risks, benefits, and re-permission emails' themes, we used fractional logit regression analysis.

Table 5.1.1 – Analyses Overview.

Research Questions	Data	Source	Analyses	Chapters
RQ1 - Are there any systematic patterns in how firms designed their re-permission emails to request access to users' data?	Collection of a sample of 1506 re-permission emails	Snowball Approach + request through a Prolific Panel	NLP - Latent Dirichlet Allocation, Manual Content Analysis, and LIWC, TextEvaluator	Chapter 6
RQ2 - How self-interested were firms? Did the benefits (from the data usage) as well as the risks (of non-compliance) drive request content and intent?	# of Marketing, Persistent Marketing, and Third-Party Cookies	Cookiebot	Fractional Logit Regression	Chapter 7
	Expected Online Ad Revenues (Google AdSense)	SEMrush Rank2Traffic		
	# Data-breaches (pre-GDPR)	Prilock HaveIBeenPwned Wikipedia		
	Website Popularity	Amazon AWIS		
	Content analyzed re-permission emails.	LDA Topics Theory-Based Themes		

5.2. The First Database: Collection of the GDPR Re-Permission Emails

The enforcement of the GDPR forced companies collecting data about EU citizens to communicate transparently and clearly about their online harvesting strategies, with the specific aim to acquire an "*explicit and informed*" opt-in by users for the collection, usage, and sharing of data by online companies. Consequently, companies worldwide sent out a massive amount of the so-called "*GDPR re-permission emails,* "which were mainly aimed at obtaining consent by users for the companies' data access.

Consequently, as previously done by Goldfarb and Tucker (2011) and Goldberg, Johnson, and Shriver (2019), we exploit the GDPR enforcement as an event study, and we collected a considerable amount of re-permission emails to study companies' privacy communication strategies. We collected these communications in three waves:

- **September 2018**

  We collected 370 communications by looking among the emails received in the authors' personal email accounts – by using search words such as "*GDPR*", "*Privacy Policy*", and "*Privacy Updates*" – and by searching online for some real-world example of these re-permission communications – by using search words such as "*GDPR email examples*", "*Privacy policy updates May 2018*", "*GDPR opt-in emails*" and "*GDPR email consent*".

- **April 2019**

  We collected additional 309 communications by asking marketing class students to look for these emails in their personal email accounts – by using the same search words mentioned above.

- **August 2019**

  We conducted a request to a Prolific panel of respondents, providing a monetary incentive to reach additional types of communications received in their email boxes. Thanks to the study, we gained additional 931 emails. For further information about the survey, see Appendix A.

This collection resulted in a total of 1610 communications. However, after an accurate inspection of all the emails, we found that 101 emails were duplicates of emails already present in the dataset; additionally, we also found that three emails were sent out by companies that do not have an existing or active website. Consequently, we proceed by removing them by the final dataset, getting to a final sample of 1506 re-permission emails sent out by various companies present in multiple countries during different periods of the year.

For each of these communications, we have recorded the information described in Table 5.2.1.

Table 5.2.1 – Information Collected for Each Re-Permission Email.

Variable	Description
ID	Unique identifier for the re-permission email
Company	Name of the company sending the re-permission email
Date	Date in which the re-permission email was sent
Text	Entire text of the re-permission email in original language and in English.
Language	Language of the re-permission email
Country	Country of the company sending the re-permission email

We describe the sample regarding each of the information collected in the following sub-chapters.

5.2.1. Re-Permission Emails’ Language and Countries

We collected a wide variety of re-permission emails from users all around the globe. Looking at the descriptive statistics about the languages and countries used in all these emails (Table 5.2.1.1), it is possible to see a wide assortment of them in our sample. It is essential to highlight that the country recorded for the email represents the country from which the email was sent rather than the country representing the company sending it. For example, there can be cases in which the email was sent by one of the subsidiaries of the main company (e.g., Facebook UK), and consequently, the country of the company (in the case of Facebook, United States) is different from the country of the email (in the example, United Kingdom).

As it is possible to see from Table 5.2.1.1, most of the email was written in English (71%), followed by Italian (17%), Poland (5%), Portuguese (3%), and Spanish (2%). Accordingly, we also found that 44% of our sample emails were sent out by companies present in Anglo-Saxon countries, such as the United Kingdom and the United States. This, once again, highlights the global scope of the GDPR and the extended territorial influence it had on companies established in countries that were not necessarily in Europe; however, as we expected, the majority of the re-permission emails (51%) were sent by European companies or subsidiaries.

To be able to analyze the content of the entirety of the re-permission emails collected, we had to translate the emails written in a language other than English. We proceeded by using Google Translator and double-checking its translation's precision by using, for a sub-sample of the emails, the translation made by a mother-tongue speaker.

Table 5.2.1.1 – Languages, Countries, and Continents of the Re-Permission Emails.

Language	% of emails
English	71.05%
Italian	16.47%
Polish	5.11%
Portuguese	2.92%
Spanish	2.19%
Hungarian	0.40%
Greek	0.33%
French	0.33%
Swedish	0.34%
Czech	0.20%
Dutch	0.20%
Latvian	0.13%
German	0.13%
Romanian	0.13%
Russian	0.07%
Total	**100.00%**

Country	% of Emails
United Kingdom	23.97%
United States	20.45%
Poland	5.05%
Italy	3.92%
Portugal	2.52%
Spain	2.19%
Ireland	1.86%
Germany	1.79%
Sweden	1.33%
Australia	1.13%
Others	35.79%
Total	**100.00%**

Continent	% of Emails
Europe	50.86%
North America	21.25%
Asia	1.33%
Australia	1.26%
South America	0.27%
Missing	25.03%
Total	**100.00%**

5.2.2. Re-Permission Emails' Date

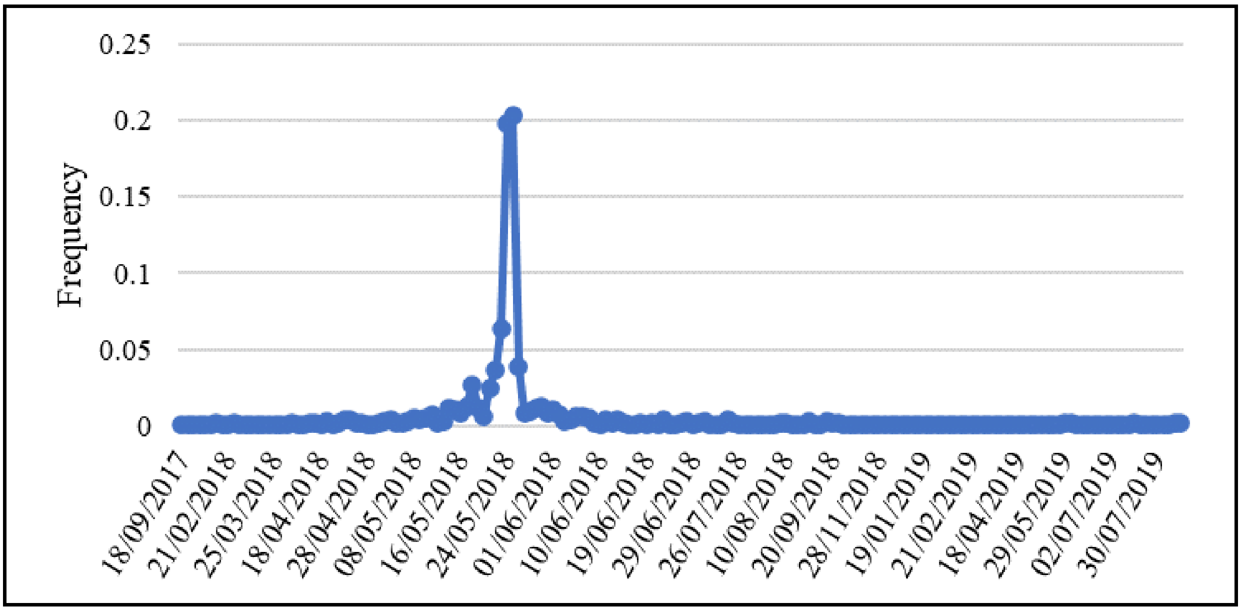
It was possible to retrieve the information about the sending date for 1120 emails of our sample since 580 emails (38%) have no data either because it was not visible in the screenshot uploaded by Prolific's users or it was not present in the examples found on the web.

Not surprisingly, and according to our expectations, most of the emails were sent out in 2018 (58%), especially in the month of May, but there is also a small percentage of emails that were sent out in 2019 (3%) – see Figure 5.2.2.1 The last date recorded for the emails collected was August 8, 2019.

The emails sent out in 2019 may still be classified as "*GDPR type of communications*" since the GDPR had effects also after its implementation. Firms were, indeed, overall unprepared and in delay with the compliance to the terms of the GDPR (Thompson 2018); consequently, some of them sent out the re-permission email after the enforcement of the GDPR.

Additionally, to be compliant with the GDPR requirement on companies' data collection, usage, and sharing, companies should ask for users' consent each year – there exist few exceptions regarding companies that were able to obtain some extension. Consequently, it can be that we recorded communications sent out by companies that were trying to get, once again, customers' data access.

Figure 5.2.2.1 – Distribution Plot of the Re-Permission Email's Sending Dates.

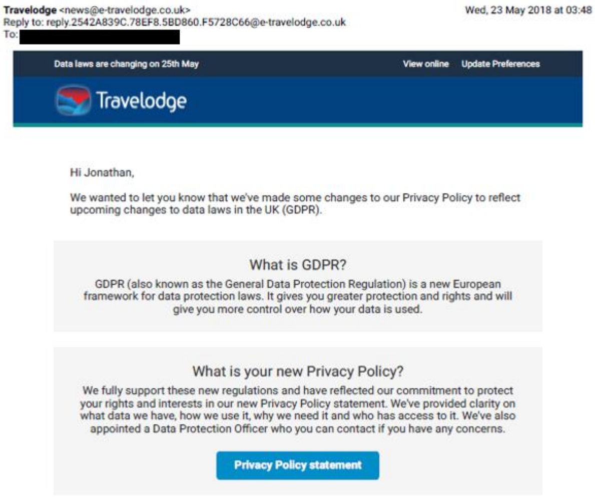
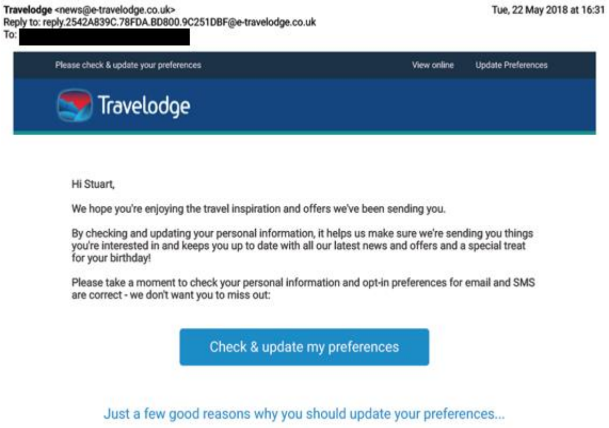


5.2.3. Re-Permission Emails' Companies

These 1506 communications were sent out by 1396 unique companies from all over the world. Some companies have sent different types of re-permission emails, meaning that they communicate the "*need for data*" in multiple ways in the observational period. We have companies with up to 5 different emails sent out for the GDPR enforcement, meaning that 6.6% of the communications present in our sample have sent out more than one re-permission communication.

For example, the accommodation company Travelodge has sent out these two communications, which exploit different communications levers to obtain users' consent (Figure 5.2.3.1). Firms may have various reasons for sending different versions of re-permission emails (e.g., different targets, areas/countries, and time periods). We do not know the reason behind this choice; therefore, we conduct robustness checks later in our analyses to test our results' sensitivity by including vs. removing the 6.6% of emails sent by the same company (Appendices F and G).

Figure 5.2.3.1 – Example of Different Re-Permission Emails Sent by the Same Company.

GDPR email 1	GDPR email 2
 **Travelodge** <news@e-travelodge.co.uk> Wed, 23 May 2018 at 03:48 Reply to: reply.2542A839C.78EF8.5BD860.F5728C66@e-travelodge.co.uk To: [redacted] Data laws are changing on 25th May Hi Jonathan, We wanted to let you know that we've made some changes to our Privacy Policy to reflect upcoming changes to data laws in the UK (GDPR). What is GDPR? GDPR (also known as the General Data Protection Regulation) is a new European framework for data protection laws. It gives you greater protection and rights and will give you more control over how your data is used. What is your new Privacy Policy? We fully support these new regulations and have reflected our commitment to protect your rights and interests in our new Privacy Policy statement. We've provided clarity on what data we have, how we use it, why we need it and who has access to it. We've also appointed a Data Protection Officer who you can contact if you have any concerns. 	 **Travelodge** <news@e-travelodge.co.uk> Tue, 22 May 2018 at 16:31 Reply to: reply.2542A839C.78FDA.BD800.9C251DBF@e-travelodge.co.uk To: [redacted] Please check & update your preferences Hi Stuart, We hope you're enjoying the travel inspiration and offers we've been sending you. By checking and updating your personal information, it helps us make sure we're sending you things you're interested in and keeps you up to date with all our latest news and offers and a special treat for your birthday! Please take a moment to check your personal information and opt-in preferences for email and SMS are correct - we don't want you to miss out: Just a few good reasons why you should update your preferences...

5.2.4. Re-Permission Emails' Content

These emails present a high heterogeneity in terms of contents and styles. We have very long emails characterized by high-quality, informative content explained using legal language or emails that use a simpler language and base their content on infographics and images to make the legislation understandable to everybody. We also have communications with few lines of text that require the users to simply re-confirm the consent or to just opt-out from the mailing list. Others, instead, use a very low register, a more colloquial type of language provides, and a very low informative content.

Consequently, from a very preliminary and simplistic analysis of the emails, it is possible to detect differences between the communication elements inserted in the emails and the various communication styles used to require data access.

Additionally, from the literature review and the conceptual framework described in Chapters 3 and 4, we outline six main themes which have been identified as factors influencing individuals' likelihood to disclose personal data: two *informational* in nature (e.g., control, transparency), and four *persuasive* (e.g., incentives, both monetary and non-monetary, and framing of the message—in terms of gains and losses and time orientation). The question is whether firms actually used them in their privacy-related communications.

This had prompted us to analyze with more rigor and more systematically all the emails collected to, then, be able to correlate them with the characteristics of the companies crafting them. We proceed with the analysis of the content of the re-permission emails collected using a three-stage approach:

1. We checked the presence of different topics in our sample of emails through NLP techniques. More specifically, we used the Latent Dirichlet Allocation (Blei, Ng, and Jordan 2003) to analyze in an unsupervised way the emails' text and to detect the main latent topics.
2. We turned to theory and retrieved more specific information about the elements that previous literature highlights as being influential in altering customer's disclosure behavior. Then, we randomly selected 20% of the total re-permission email sample and asked two independent judges to code them manually based on a theory-based coding protocol. Lastly, we collected additional text-related variables by using two automated online software (e.g., LIWC and TextEvaluator), and we predicted the likelihood that a specific theme characterizes a re-permission email by modeling the manual-coded variables on these additional text-related variables. We tested the predictive validity by using lift-charts analyses.
3. We used the models estimated in the second stage to predict the presence of the different themes in the whole sample of emails collected (N = 1506).

Notably, we tried to use two different methodologies to content-analyze the texts of the emails collected. In the first step, we chose a data-driven technique – the LDA modeling approach – that has helped us to get a first overview of the content of the emails without being biased by any theoretical background. The second and third steps, instead, are more theoretically based, reaching more granular insights than the ones obtained in the first stage. Interestingly, we found that the two methodologies are consistent and can detect the same broad categories of themes in the texts.

The description and analyses of the results obtained using this three-stage process are available in Chapter 6, which elaborates on the first research question of this thesis – "*How did firms articulate their requests for data? Which themes characterize GDPR re-permission emails?*".

5.3. The Second Database: Collection of Information About Companies

As mentioned in the previous section, our sample comprises 1396 unique companies, meaning that companies in our dataset have sent more than one email (6.6% of the total companies).

For each company, we have then collected information about its specific characteristics – e.g., the number of employees, industry, country, and age – and its online website – e.g., the number of cookies, type of cookie policy, and expected online ad revenue.

We provide additional details about the variables collected in the next sections.

5.3.1. Companies' Characteristics

We collected information about the different companies using a variety of data sources. The primary data source we relied on is Orbis, a database owned by Bureau van Dijk, which contains detailed information about companies worldwide. From this source, we collected the last data available about the number of employees, the year of foundation, the country, the NAICS code, the SIC code, and the BvD sector of each company of the dataset. If some of the information were not available on this first and primary data source, we complemented it by searching other online resources such as SimilarWeb, Crunchbase, or Owler, which are all websites containing, among others, societal data.

In the following sections, we describe and provide summary statistics for each of the variables mentioned above.

5.3.1.1. Number of Employees

We collected this information for 1300 out of the 1396 total companies in the dataset, which corresponds to roughly 93% of the full sample.

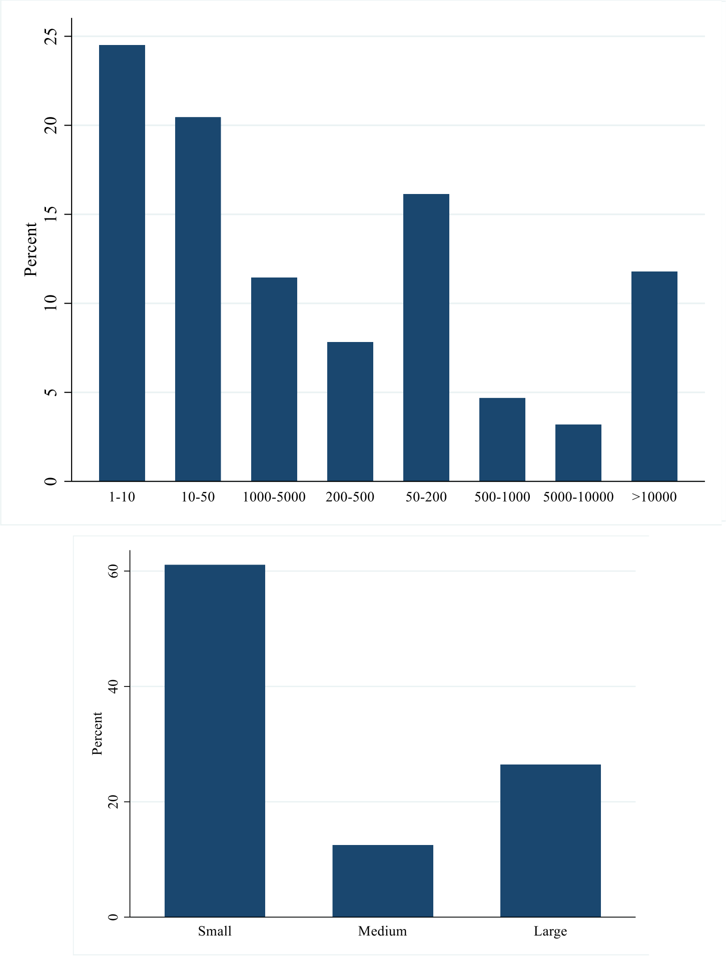
This variable has been operationalized as a categorical variable that assigns the company to the correct dimension's interval. This is one of the variables commonly used as a proxy for the company's size since it can be assumed that more prominent companies have a higher number of employees. Additionally, to corroborate this assumption, we also collected another variable commonly used to establish the company's dimension: the yearly operating turnover (information available for 94% of the companies). As it is possible to see from the table below (Table 5.3.1.1.1), companies that have a higher number of employees also have a higher value of yearly operating turnover, suggesting, once again, that the number of employees can capture the dimension of the companies in our dataset.

By looking at the histograms (Figure 5.3.1.1.1), it is possible to see a lot of variety in the companies' dimensions making up our dataset. However, if we group the categories of the number of employees in three main super categories – small, medium, and large – it is possible to see that many of the companies in our sample can be considered as small (a category that is made by grouping companies with 1 up to 200 employees: 64%). The remaining part is composed of large companies (a category that is made by grouping companies with more than 1,000 employees) that represents 24% of our sample, and medium companies (a category that is made by grouping companies with 200 up to 1000 employees) which are the remaining 12% of the sample.

Table 5.3.1.1.1 – Descriptive Statistics on the Size of the Firms Sending Re-Permission Emails (N = 1396).

Number of Employee (Category)	Mean of Last Available Turnover ($)	% of the Companies
1-10	2,659	24.07%
10-50	5,859	19.70%
50-200	46,336	15.47%
Small	**18,285**	**59.24%**
200-500	191,393	7.31%
500-1000	305,249	4.37%
Medium	**248,321**	**11.68%**
1000-5000	1,013,635	9.81%
5000-10000	2,981,338	3.15%
>10000	36,200,000	9.24%
Large	**13,398,324**	**22.20%**
Missing	**2,881**	**6.88%**

Figure 5.3.1.1.1 – Bar Charts of the Firms Sending Re-Permission Emails by their Dimension.

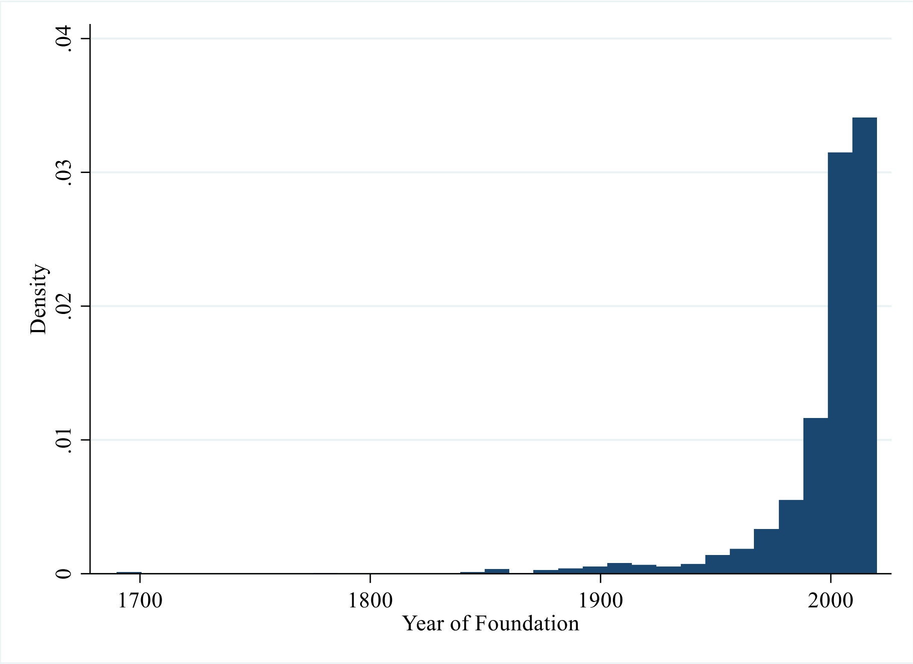


5.3.1.2. Year of Foundation

This variable aims to capture the "*experience*" of a company and its maturity in the market. In our study of companies' privacy communications, we want to control for the age that the company has since it can be that younger companies use different levers and strategies than older companies, which can exploit their reputation. We collected this information for 1306 out of the 1396 total companies in the dataset, corresponding to 94% of the full sample.

As it is possible to see from the histograms below (Figure 5.3.1.2.1), our sample is mainly composed of companies founded recently, which can be considered new in the market. They have been mainly constituted started from the years 2000's on (65%), with the majority formed in 2011 (6% of the total sample).

Figure 5.3.1.2.1 – Distribution Plot of the Firms Sending Re-Permission Emails by their Year of Foundation.



Starting from the companies' year of foundation, we also constructed another variable called "*Age*", representing the time between the initial creation of a firm and the present time (in years). This variable was, consequently, obtained in the following way:

$Age = 2020 - Year\ of\ Foundation$

This was mainly done to be able to interpret the results of our subsequent models more efficiently. Table 5.3.1.2.1 report the descriptive statistics of this variable.

Table 5.3.1.2.1 – Descriptive Statistics on the Age of the Firms Sending Re-Permission Emails (N = 1396).

	Mean	SD	Min	Max	% Missing
Firms' Age (in years)	22.49	26.10	0.00	237.00	6.45%

5.3.1.3. Country of the Firms' Headquarter

The emails collected have been sent out by companies with headquarters all around the world. Since the GDPR is characterized by a broader territorial scope than the previous privacy regulations, companies located outside the European Union are required to communicate about their data practices when collecting data about EU citizens.

It is essential to highlight a difference between the email-specific and the company-specific country variable; while the latter refers to the legal headquarter of the company sending the email (e.g., Facebook Inc.), the former is mainly signaling which of the subsidiaries, if any, is sending the communication (e.g., Facebook UK, Facebook France…).

We collected this information for 1351 out of 1396 total companies, which corresponds to 97% of our companies' entire sample.

As it is possible to see from Table 5.3.1.3.1, most of the emails in our dataset have been sent from companies with headquarters in the United Kingdom (28%), in the United States (25%), and in Italy (11%).

It is also possible to look at this information by aggregating the countries in their corresponding continents to have a more tangible sense of how much the GDPR enforcement has been felt from the other different geographical areas. Not surprisingly, as it is possible to see from Table 5.3.1.3.1, many of the emails were sent from EU companies (65%), followed by the North American ones (27%). The fact that North American companies are sending out a significant number of communications should not surprise. Europe and America were already connected in terms of data protection law by the EU-US Privacy Shield[1]. This Program was deemed in 2016 with the specific aim to protect EU citizens' personal data, which were transferred from the EU to US companies. Consequently, given the high transatlantic commerce between the EU and the US, the GDPR was strongly received by US companies that were willing to comply with it – by sending privacy communications – not to lose profits and data.

Table 5.3.1.3.1 – Descriptive Statistics on the Countries and Continents of the Firms Sending Re-Permission Emails (N = 1396).

Country of the Headquarter	% of Companies
United Kingdom	28.01%
United States	25.14%
Italy	11.10%
Poland	4.66%
Spain	3.22%
Germany	2.94%
France	2.51%
Portugal	1.93%
Ireland	1.65%
Netherlands	1.43%
Sweden	1.29%
Belgium	1.00%
Others	11.84%
Missing	3.22%
Total	**100%**

Continent of the Headquarter	% of Companies
Europe	65.33%
North America	26.50%
Asia	3.80%
Oceania	1.00%
Africa	0.07%
South America	0.07%
Missing	3.22%
Total	**100%**

Given the massive amount of emails sent out by EU companies and since we are looking at the impact of GDPR, which is a European Union Regulation, we have created an indicator variable to consider the fact that the company is located inside the EU or outside of it.

5.3.1.4. Industry

Privacy communications sent out by the different companies may also differ depending on the industry to which the company belongs. For example, there may be sectors that are more data reliant

[1] https://www.privacyshield.gov

by construction – e.g., Media, News, Banking – and others which, instead, are less affected by the GDPR since data are not central for their daily operations – e.g., Construction, Public Administration.

Consequently, a variety of information about the companies' industry has been collected using the Orbis database. The data gathered about the industry classification of each company in the sample are the following ones:

- *BvD Sector Classification*, which assigns the company to the sectors defined by the Bureau van Dijk (BvD).
- National American Industry Classification System (*NAICS*) Core Code.
- Standard Industrial Classification (*SIC*) Core Code.

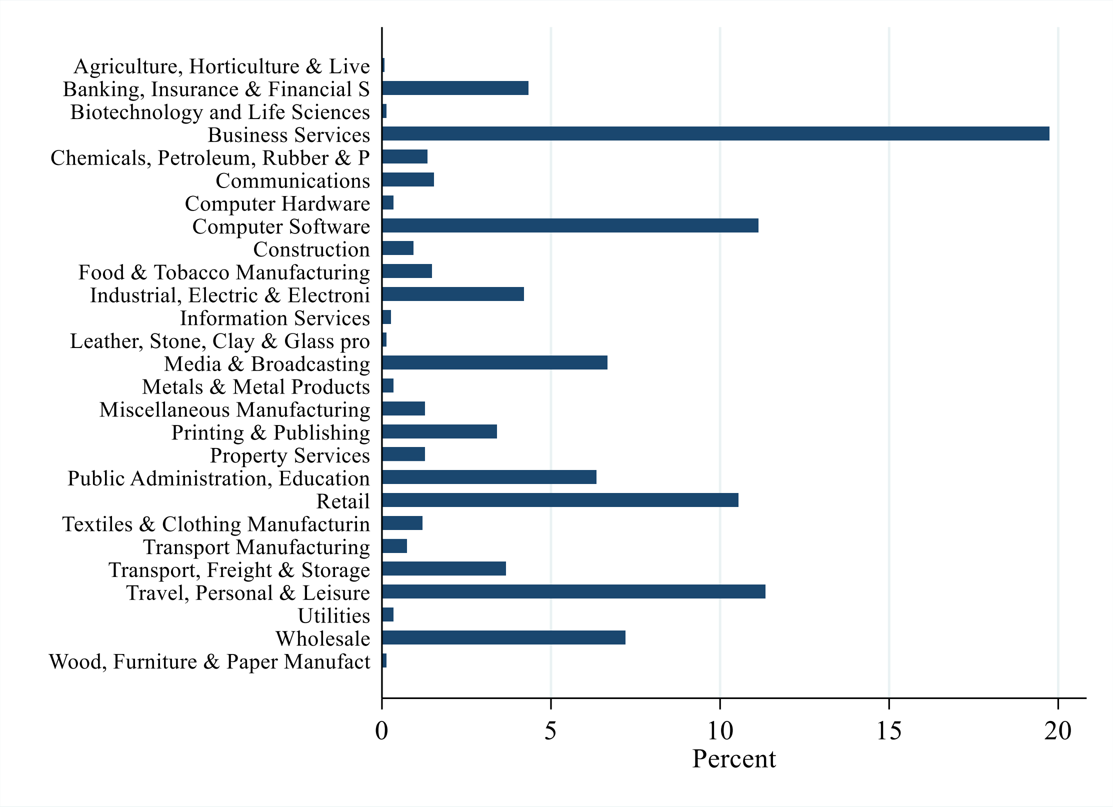
As it is possible to see from Table 5.3.1.4.1, not all the industry classification codes were recorded for all the companies. However, by using the BvD Sector classification system, we can reach up to 99.5% of the total companies collected. Consequently, given the availability of the information for a higher number of companies, the BvD sector will be adopted as our standard for the companies' classification in the corresponding industries and will be used in our subsequent models.

Table 5.3.1.4.1 – Industry Data Availability by Classification System.

	Data Available for:	
	# of Companies	**% of Companies**
NAICS Core Code	1,316	94%
SIC Core Code	1,286	92%
BvD Sectors Classification	**1,390**	**99%**

As it is possible to detect from Figure 5.3.1.4.1, most companies in our sample belong to the "*Business Services*" sector (20%), followed by the "*Retail*" and the "*Computer Softwar*e" sectors (12%) and by the "*Travel, Personal & Leisure*" sector (11%).

Figure 5.3.1.4.1 – Bar Chart of the Firms Sending Re-Permission Emails by their BvD Sectors.

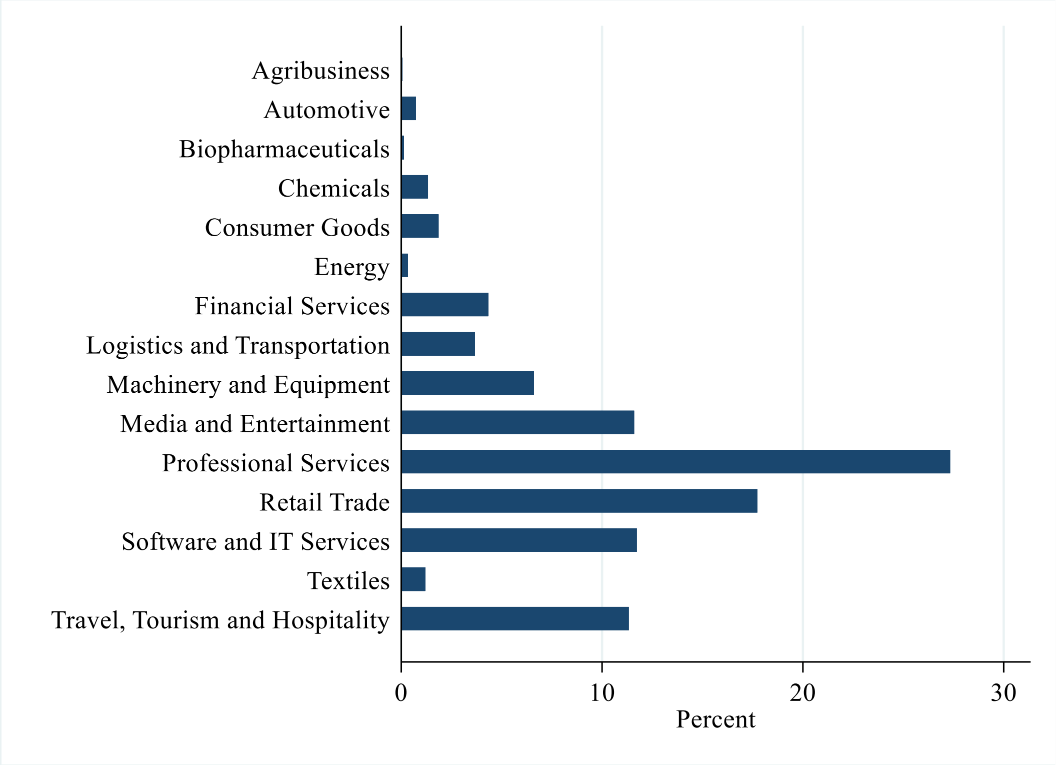


However, by looking at the possible sectors defined by the Bureau van Dijk, it is possible to notice that some of the categories can be grouped into one main super-category. For example, "*Computer Hardware*", "*Computer Software*" and "*Information Services*" can be grouped, without losing too much precision, into one category labeled "*Software and IT Services*". Consequently, we tried to reduce, consistently and systematically, the number of categories recorded in the Bureau van Dijk system. The option that could suit our aims best was to manually re-assign each of the BvD Sectors to the corresponding industries defined by the US government[2]. In the histogram below (Figure 5.3.1.4.2), it is possible to see a very similar situation to the one depicted in the histogram above: 27% of the sample is composed of "*Professional Services*" companies, 18% by "*Retail Trade*" companies, 12% by "*Software and IT Services*" companies and 11% by "*Travel, Tourism, and Hospitality*" companies. Consequently, given that we achieved the same results depicted above, we

[2] https://www.selectusa.gov/industries

can be sure that the coding was correctly done, and we can use the reduced US industry classification in our analyses, being sure we are not changing the results too dramatically.

Figure 5.3.1.4.2 – Bar Chart of the Firms Sending Re-Permission Emails by their US Industry.



By looking at Table 5.3.1.4.2, we can also proceed in further reducing the number of categories observed: with the first five categories, we can classify 80% of the companies in our sample, while the remaining ten categories contribute to providing additional information for a small part of the sample in a fragmented way. Without losing too much precision, we decided to group these last ten categories inside one unique class called “Others”. Additionally, Table 5.3.1.4.3 provides examples of firms and brands included in our sample for each of the US industries selected.

Table 5.3.1.4.2 – Descriptive Statistics on the US Industry Classification for the Firms Sending Re-Permission Emails (N = 1396).

US Industry Classification	Percent	Cumulative Percent	Grouping
Professional Services	27.36%	27.36%	**Professional Services**
Retail Trade	18.12%	45.49%	**Retail Trade**
Software and IT Services	12.11%	57.59%	**Software and IT Services**
Travel, Tourism and Hospitality	11.39%	68.98%	**Travel, Tourism and Hospitality**
Media and Entertainment	10.82%	79.80%	**Media and Entertainment**
Machinery and Equipment	6.38%	86.17%	**Others**
Financial Services	4.30%	90.47%	
Logistics and Transportation	3.51%	93.98%	
Consumer Goods	1.79%	95.77%	
Chemicals	1.29%	97.06%	
Textiles	1.22%	98.28%	
Automotive	0.72%	99.00%	
Energy	0.36%	99.36%	
Biopharmaceuticals	0.14%	99.50%	
Agribusiness	0.07%	99.57%	
Missing	0.43%	100.00%	

Table 5.3.1.4.3 – Examples of Firms and Brands for Each US Industry.

US Industry	Examples	% of Companies
Professional Services	Accenture, Aruba, PwC, LinkedIn, MailChimp	27.36%
Retail Trade	Ebay, IKEA, Selfridges, Nordstrom, Yoox	18.12%
Software and IT Services	ASUS, Coursera, DataCamp, Grammarly, Hotjar	12.11%
Travel, Tourism and Hospitality	Airbnb, Dominos, Hostelword, Lastminute.com	11.39%
Media and Entertainment	The Guardian, The Economist, Spotify, YouTube	10.82%
Others	FitBit, Pampers, Paypal, Estee Lauder, Gucci	20.20%

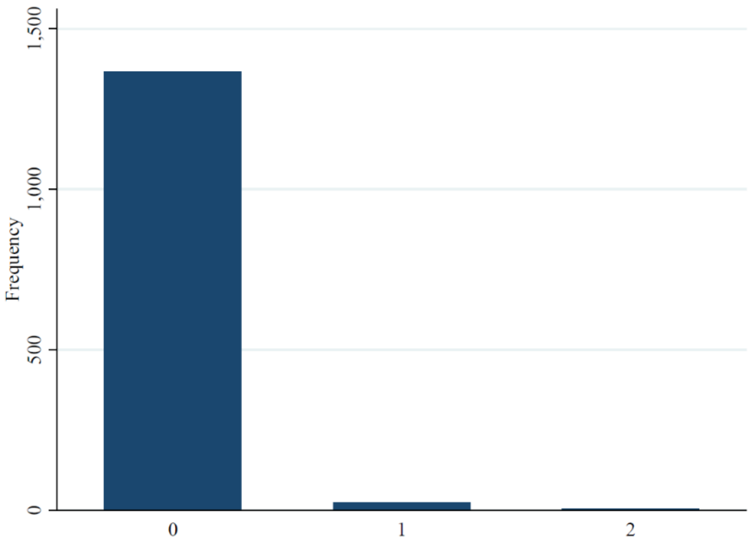
5.3.1.5. Data-Breaches Experienced

Another information that we were able to retrieve – and which is strictly related to firms' data privacy management procedures – was the number of data breaches experienced (if any) by the companies composing our sample. We collected this information by using different data sources available online: Prilock[3], Have I Been Pwned[4] , and Wikipedia list data-breached companies[5]. The collection of these additional data allowed us to control, in our subsequent models, for the possible effect that the experience of data security failures may have on the way in which companies communicate their need for data. In Chapter 4, we argued that companies act in a self-interested way by carefully evaluating not only the benefits but also the risks that may arise from not completely adhere to data protection laws. Data-breach announcements may be considered risks in which companies may incur if they do not follow the procedures and the requirements defined by data protection laws, in that becoming events that may make companies more cautious and more inclined to be compliant with the law when talking about privacy and data-related procedures, than companies which have not experienced them.

Therefore, we collected the number of data breaches experienced by the companies in our sample prior to the GDPR enforcement. Unfortunately, in our data-breach database, we only have the year in which the exposure happened; therefore, we considered, for the aim of this work, the data breaches that happened in the years before 2018 – 2018 excluded. We found that 30 companies (2%) we have collected were breached in the pre-GDPR period. Figure 5.3.1.5.1 shows the bar chart and the frequency table of this counting variable. As it is possible to see, the vast majority of the companies did not experience any breach in their data security in the pre-GDPR era; however, we have some of the companies that not only have experienced data breaches, but also have been exposed multiple times. Figure 5.3.1.5.2 show the name of the companies that have been exposed.

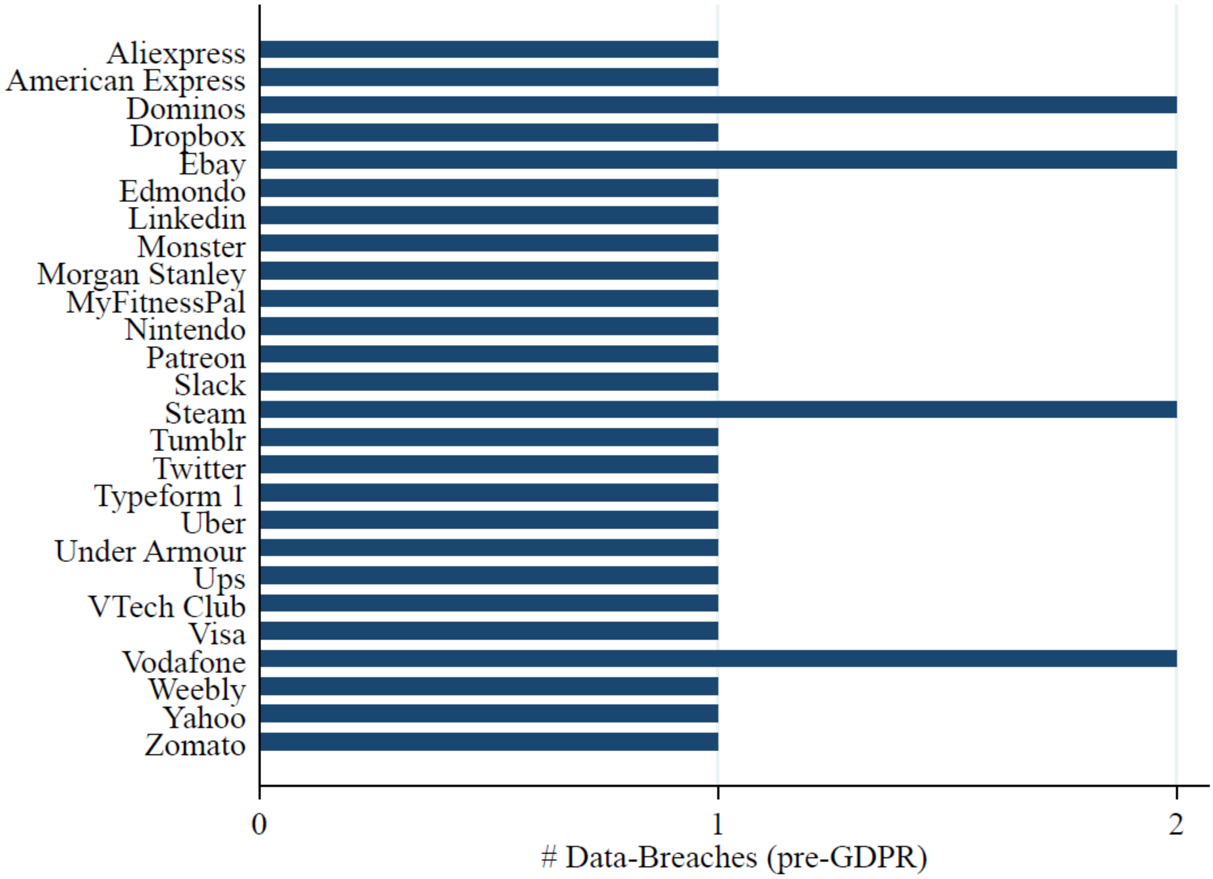
[3] https://www.prilock.com/breach_list.php
[4] https://haveibeenpwned.com/
[5] https://en.wikipedia.org/wiki/List_of_data_breaches

Figure 5.3.1.5.1 – Bar Chart and Frequency Table for the Number of Data-Breaches Experienced by the Companies (N = 1396)



# of Data-Breaches	Freq.	Perc.
0	1,366	97.50%
1	25	1.79%
2	5	0.36%
Total	1,396	100%

Figure 5.3.1.5.2 - Horizontal Bar Chart of the Number of Data-Breaches by the Companies Exposed (N = 30)



5.3.1.6. Summary of the Companies-Related Variables

For simplicity, we summarize the main company-related variables just described in Table 5.3.1.6.1 below. Small and young firms characterize our sample. These are mainly established in Europe and operate in the "*Professional Services*", "*Retail Trade*" and "*Software and IT Services*" sectors.

Table 5.3.1.6.1 – Summary of the Main Results for the Variables Related to the Firms' Characteristics.

Variable	**Operationalization**	**% of Missing Values**	**Main Descriptive Results** **Our sample is composed by:**
Number of Employees	Categorical variable with 8 levels, used as proxy of the dimension of the companies in our sample.	7%	*Small* Companies
Age of the Company	Continuous variable indicating the age of the companies in our sample.	6%	*Young* Companies
EU	Binary variable assuming value 1 if the company is located in EU and 0 otherwise.	3%	*European* Companies
US Industry	Categorical variable with 6 levels indicating the membership of the company to the US Industry Classification System.	4%	*"Professional Services", "Retail Trade" and "Software and IT Services"* types of Companies
Data-Breaches	Number of data-breaches experienced by the companies before the GDPR enfocement	-	Companies which have *not experienced data-breaches prior to the GDPR* .

5.3.2. Companies' Website Information

In order to relate the privacy communication strategy to the companies' online strategy, we collected two main types of information which can give the extent to which a company relies on online data collection:

- Information about the *web cookies* used by the companies on their online websites, which gives an idea of the data harvesting strategy which firms are implementing.
- Information about the *performances of the companies' online websites*, which provides an overview of how much the different firms may invest and profit from the online sector.

In the following sections, we present the different data sources used and the multitude of variables collected to supply measures for our main research questions adequately.

5.3.2.1. Web Cookies

In order to gain specific information about how much data a company can harvest from its website, we collected information about the cookies placed by companies on their website. We contend that the number of cookies set by firms or partner companies used for marketing and targeting purposes can be considered a proxy of a firm's data harvesting intention – e.g., firms' willingness to collect data from customers and potential customers. Previous literature recognizes that the vast majority of online data are collected via cookies, which are placed on a wide variety of websites, often with the goal of profiling consumers (Neumann, Tucker, and Whitfield 2019). Cookies are pieces of HTML code that download text files in the browser when users visit a website. They allow to uniquely identify users online, track their browsing activity, and store all the firms' needed information in the locally stored file. Consequently, it is reasonable to expect that if a company uses a multitude of marketing cookies on its online domain, it has the opportunity to gather a higher volume of data about the users surfing its website.

Our primary source of information in this regard has been Cookiebot, which is a cloud service provided by Cybot. This platform mainly aims to help companies get GDPR compliance by supporting them with the creation of specific cookie policies and consent banners and a set of tools for cookies' management – e.g., the cookie-repository tool that allows companies to know and manage the cookies placed on their domains. Its relevance for our research regards its capability to automatically detect all the cookies and tracking technologies on a company website and to provide a classification of the cookies recorded based on the goal that the specific cookie tries to achieve. By crawling the pages of the company's domain, Cookiebot can record the number and type of active cookies present in that exact moment on that particular website. Consequently, this company provided us with information about the number, duration, and variety of cookies that companies are using on their online websites.

It is essential to highlight that we asked Cookiebot to make two extractions for each website domain of the companies composing our sample: one in the period between the 25th and the 30th of October 2019 and the other on November 18, 2019.

This, in principle, allows us to control for any difference in the selection of:

- the random subsample of the 1000 webpages of every domain,
- and the specific day in which Cookiebot scraped the websites.

Cookiebot scanned up to 1,000 pages for each of the 1396 firms' domains in two snapshots, providing us with results from a total of 980,182 pages analyzed – some companies' websites have less than 1,000 subpages. According to Cookiebot reports, around 35 websites use more than 500 cookies. These websites are probably containing dynamically named cookies, which are really the same cookies but with new names for each user session. As each crawl of a website simulates many user sessions, these types of cookies are registered multiple times, leading to possible over-estimations of the number of cookies recorded for the corresponding companies' websites. Consequently, we decided to substitute these anomalies with the cookies averages to avoid distortions and false results later in our analyses.

Moreover, Cookiebot provided us with a categorization of the cookies recorded. Cookies can differ with regards to their final aim, and it is possible to classify them into the following five categories[6]:

- *Necessary Cookies*, which are cookies needed to make the website properly working and that cannot be turned off.
- *Preference Cookies*, which allow the company to remember basic information about how the website should behave or look like. These are cookies used to remember how the user sets up the webpage in terms of language, currency, geographical area.
- *Statistical Cookies*, which the company uses to get overall statistics about how its users use the website.
- *Marketing Cookies*, which are the most intrusive type of cookies and are used to keep track of the users' behaviors across websites. These cookies are used by both companies and marketing agencies to implement targeted marketing strategies and to display ads that are more relevant and engaging for the user, given his/her browsing history, and thereby more valuable for publishers and third-party advertisers.
- *Unclassified Cookies*, which are in the process of being classified by Cookiebot.

Since much research in marketing has shown that targeting strategies are actually more effective in influencing customer's behavior and bring more profitability to the firms (Aziz and Telang 2016; Bleier and Eisenbeiss 2015; Goldberg, Johnson, and Shriver 2019; Goldfarb and Tucker 2011; Marotta, Abhishek, and Acquisti 2019), the "*Marketing*" type of cookie is of utmost importance for companies which can:

- *use the information* collected by the marketing cookie *internally*; in this case, the cookie becomes an asset for the firm, allowing it to create value for its usage for marketing purposes.

[6] https://www.cookiebot.com/en/cookie-declaration/

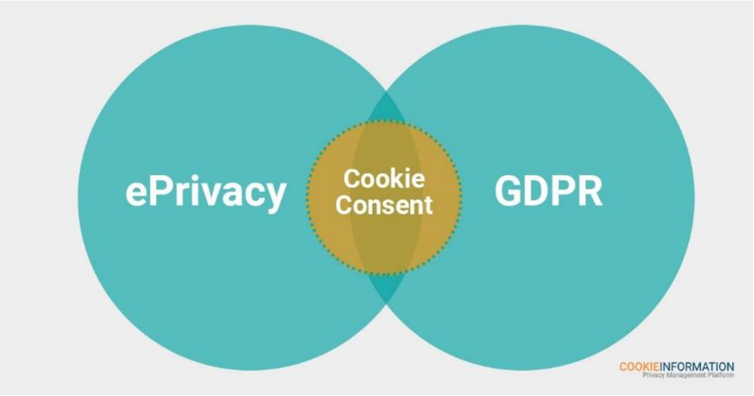
- *sell the information* collected by the marketing cookie *to third-party companies*; in this case, the cookie is a source of value per se, and the company is profiting out of it without doing anything with it. It can be seen as an extreme exploitation of the data collected.

On average, around 40% of the total cookies recorded belong to the marketing category, meaning that most cookies used by companies are implemented with the specific aim to carry out behavioral targeting activities that allow companies to achieve a higher level of profit. Consequently, it is undoubtedly relevant for the company in our sample to obtain explicit opt-in from their users to bring about effective and efficient marketing activities while being compliant with the GDPR regulation.

Additionally, cookies may differ not only in terms of their final aim but also in terms of their source and lifetime.

Regarding the source, cookies can be defined as "*first-party cookies*" – if they are placed by the firm behind the website that an individual visits – or "*third-party cookies*" – which are usually placed by partners of the 'firm's website (e.g., external domains). For example, the website may partner with advertisers to deliver ads or with an analytics company to understand how people use their site. In these cases, the external websites place their cookies on the main website to access the information they need to bring about their services.

Regarding the duration, cookies can be either *session*-specific – meaning that they are temporary and exist until the browser is open – or *persistent* – meaning that they are created to last for a more extended period of time and collect more information about users' online behavior. The central regulation trying to address issues of the length of the data storage is the European *ePrivacy Directive*, also known as "*The Cookie Law*". This is an EU directive enforced in 2002 and amended in 2009 with the specific aim of treating issues such as confidentiality of information, spam, and cookies. As opposed to the GDPR, which has the main objective of specifying how personal data should be processed and does not explicitly talk about cookies, this Directive was created to define guidelines and expectations on how to deal with online privacy, mainly regulating cookie usage. Consequently, the *ePrivacy Directive* complements and extends the GDPR's principles, and it generally takes precedence for cookies' related issues.



Source: Cookie Information Privacy Management Platform

Among the requirements set by the *ePrivacy Directive* and the GDPR, such as the need for an informed and explicit consent for cookies collection and usage except for the strictly necessary ones, there is also a reference to the *cookie duration that should not be longer than 12 months*[7].

Cookiebot provided us with information about the duration of the cookies recorded. This allows us to analyze whether companies implementing a particular type of communication aim not only to get consent for more cookies but also to store data for a longer time span to be able to implement behavioral targeting strategies. Tracking functions are, indeed, typically associated with persistent cookies (Rutz, Trusov, and Bucklin 2011).

By analyzing the data about the companies in our sample, we found evidence that most of them are using persistent marketing cookies (91.4%), and, among these, 86% are not entirely compliant with the ePrivacy Directive with regards to cookies' duration (Table 5.3.2.1.1). More importantly, about 41% of the firms are using cookies with a duration greater than two years meaning that companies are not only tracking customer behavior, but they are also storing information for

[7] https://gdpr.eu/cookies

much more time than allowed and needed. This contrasts not only the *ePrivacy Directive* but also the GDPR Art.5(e):

> " *Personal data shall be kept in a form which* ***permits identification of data subjects for no longer than is necessary for the purposes for which the personal data are processed****; personal data may be stored for longer periods insofar as the personal data will be processed solely for archiving purposes in the public interest, scientific or historical research purposes or statistical purposes in accordance with Article 89(1) subject to implementation of the appropriate technical and organizational measures required by this Regulation in order to safeguard the rights and freedoms of the data subject ('storage limitation')*"

Table 5.3.2.1.1 – Number of Companies by Marketing Cookies' Duration.

	Number of Companies	**% of Companies**
Compliant Firms		
Session-Specific Cookies	120	8.60%
Marketing Cookies' duration: between 0 and 1 year	79	5.66%
Slightly Outlaw Firms		
Marketing Cookies' duration: between 1 year and 2 years	626	44.84%
Outlaw Firms		
Marketing Cookies' duration: greater than 2 years	571	40.90%
Total	**1396**	**100.00%**

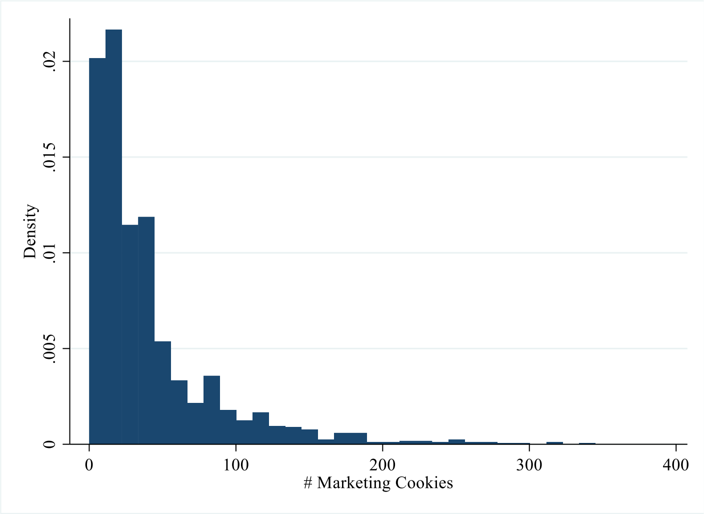
In the following sections, we present the different variables we were able to collect thanks to the collaboration with Cookiebot. We decided to focus our analyses on marketing cookies only because they are more in line with the idea of capturing companies' data harvesting strategies.

5.3.2.1.1. Number of Marketing Cookies

This variable represents the total number of marketing cookies present on the company website – regardless of their duration. As mentioned earlier, we asked Cookiebot to extract the number of cookies in two different snapshots to check for any differences in the day and pages sampled. Consequently, we have two measures of this variable. This allows us to test whether there is consistency among the data that Cookiebot provided us with. We did find a considerably high correlation between the two snapshots (r = 0.92), meaning that the numbers provided by Cookiebot are quite consistent regardless of the time of the extraction and the selection of the subsample of pages to be analyzed. Consequently, we present descriptive statistics only for one of the variables (derived from the first extraction)[8].

By looking at the histogram and table below (Figure 5.3.2.1.1.1), it is possible to notice that companies placed, on average, 34 marketing cookies on their domains with much variety around the mean. The distribution is right-skewed, meaning that most of the companies in our sample tend to have fewer marketing cookies than the average. Additionally, it is possible to notice that we have 109 companies for which no cookies were recorded.

Figure 5.3.2.1.1.1 – Distribution and Summary Statistics for the Number of Marketing Cookies present on the Firms' Websites.



Obs.	=	1396
Average	=	33.96
SD	=	41.74
Min	=	0
Median	=	20
Max	=	334

[8] Robustness checks on the two extractions' measures will be provided for each of the models in Appendices F and G.

5.3.2.1.2. Number of Marketing Cookies Domains

This variable represents the total number of external domains linked to the company website, which are placing cookies on the company webpage[9]. It can be considered a measure of the company's connection with other external companies and can also be interpreted as a measure of companies' value of data collection.

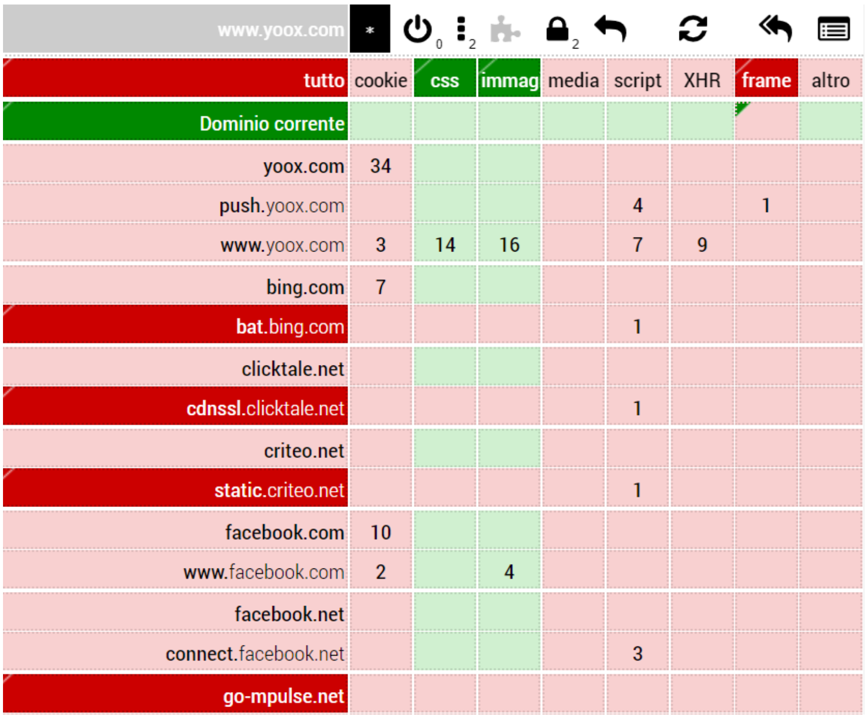
Figure 5.3.2.1.2.1 is a screenshot of some of the external connections recorded on an example website (e.g., www.yoox.com). It is possible to see that websites like Bing, Facebook, and Criteo are placing different tracking technologies on the main website. These cookies are highly relevant for these types of companies since they allow to better profile the users visiting Yoox and assign them to the right segment, out of which Facebook, Bing, and Criteo make money as advertisers.

Interestingly, this measure is highly correlated with the number of marketing cookies placed on firms' websites (r = 0.91). This is in line with our expectations since third-party owners can only read their own third-party cookies, which means that the more the external collectors, the more the marketing cookies and behavioral data the website is collecting.

Finally, we would like to clarify that Cookiebot doesn't provide any information about the type of data that websites collected through the cookies installed, but only information about the type and the total number of cookies used.

[9] https://www.paladion.net/blogs/cookie-attributes-and-their-importance

Figure 5.3.2.1.2.1 – Example of External Domains Connecting to www.yoox.com.



www.yoox.com *

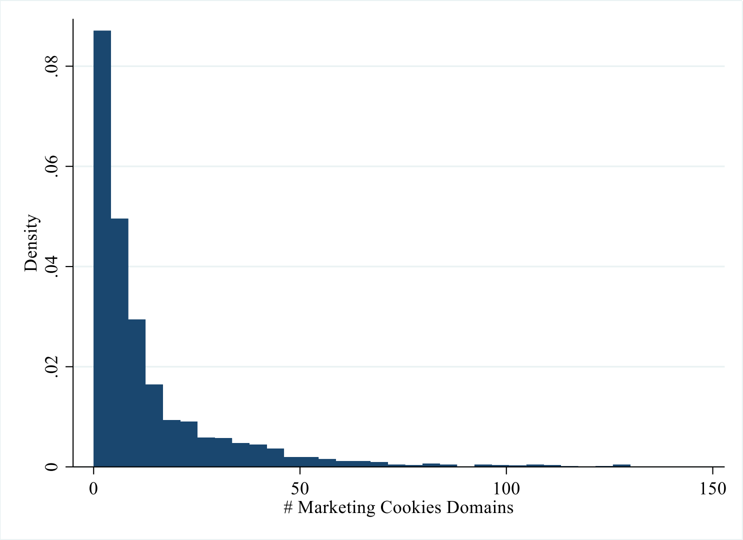
tutto	cookie	css	immag	media	script	XHR	frame	altro
Dominio corrente								
yoox.com	34							
push.yoox.com					4		1	
www.yoox.com	3	14	16		7	9		
bing.com	7							
bat.bing.com					1			
clicktale.net								
cdnssl.clicktale.net					1			
criteo.net								
static.criteo.net					1			
facebook.com	10							
www.facebook.com	2		4					
facebook.net								
connect.facebook.net					3			
go-mpulse.net								

Source: uMatrix

Also in this case, Cookiebot provided us with information about the number of external marketing collectors recorded in each of the two extractions required. The two variables are highly correlated (r = 0.92). Therefore, we decided to only present the results for the variable supplied from the first extraction.

By looking at the histogram and table below (Figure 5.3.2.1.2.2), we can see that websites have, on average, 13 connections to external websites and, also in this case, there is much variety, with most websites having less than average connections. These external collectors load, on average, 2.7 (SD = 1.39) marketing cookies on the firms' webpages.

Figure 5.3.2.1.2.2 – Distribution and Summary Statistics for the Number of Marketing Cookies Domains present on the Firms' Websites.



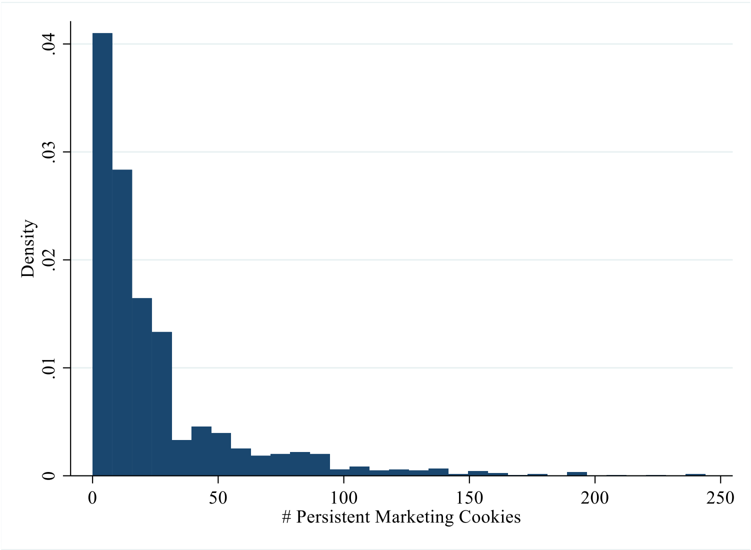
Obs.	=	1396
Average	=	13.02
SD	=	17.50
Min	=	0
Median	=	7
Max	=	130

5.3.2.1.3. Average Number of Persistent Marketing Cookies

This variable represents the total number of persistent marketing cookies – marketing cookies with a duration greater than 0 – present on the company's website. We found that only 8.6% of our entire sample uses session-specific cookies, while the remaining 91.4% is, instead, making use of persistent cookies to different degrees. This means that companies sending out re-permission emails mainly exploit this tracking technology to collect and store information about users for some time. Also in this case, we have a double measurement of this variable provided by Cookiebot. We only present the results for the variable supplied from the first extraction (r = 0.91) according to what we have done for the previously described variables.

By looking at the histogram and the table below (Figure 5.3.2.1.3.1), we can see that companies use, on average, 24 persistent marketing cookies; however, given the right-skewed distribution, we can also say that there is much variation and that many of the companies in our sample are using less than average persistent marketing cookies.

Figure 5.3.2.1.3.1 – Distribution and Summary Statistics for the Number of Persistent Marketing Cookies present on the Firms' Websites.



Obs.	=	1396
Average	=	24.30
SD	=	31.71
Min	=	0
Median	=	13
Max	=	244

5.3.2.1.4. Average Number of Marketing Cookies with Duration greater than 1 Year and 2 Years

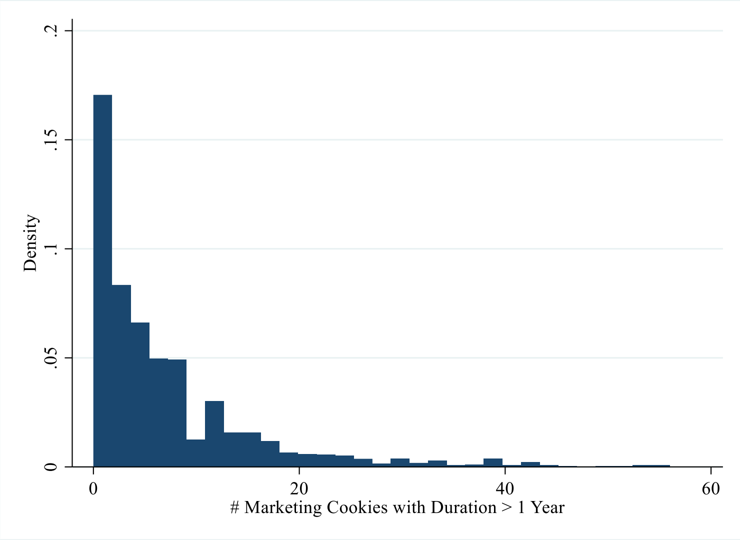
These variables represent the total number of persistent marketing cookies with a duration greater than one year and two years present on the companies' websites. This will give us a sense of how much companies can be considered as "*non-compliant*" with the e-Privacy Directive and the GDPR since they are infringing both the maximum duration established in the EU for the cookie storage and the "*data minimization*" principle, on which the new privacy regulation is based.

Also in this case, as for the previous variables, we have a double measurement of these variables provided by Cookiebot. Accordingly, to what we have done so far, we have decided to present below only the results for the variables supplied from the first extraction ($r_{\text{one\_year}}$ = 0.88 and $r_{\text{two\_years}}$ = 0.81).

By looking at the histograms and tables below (Figure 5.3.2.1.4.1 and Figure 5.3.2.1.4.2), we found that companies have, on average 7 marketing cookies with a duration greater than one year and 1 marketing cookie with a duration greater than two years. We can also see that most of the companies are storing few of these long-living cookies and that there are only some cases that are reliant on the use of long marketing cookies – the main companies using these types of cookies belong to the

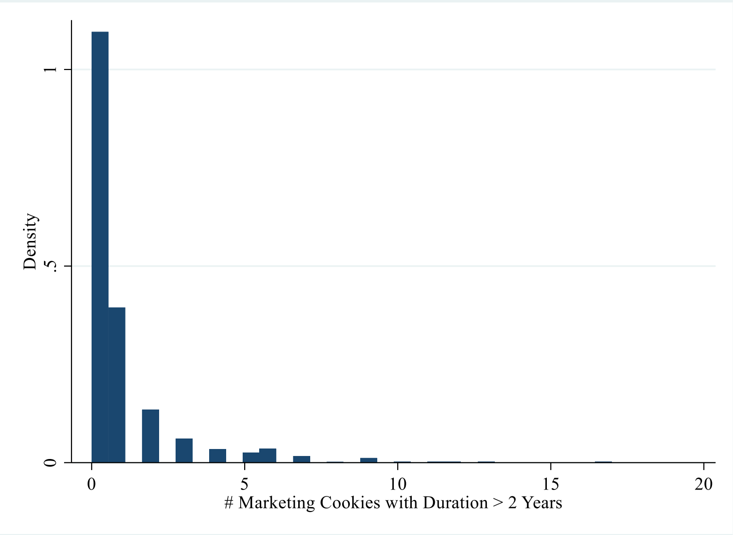
"Software & IT Service", "Professional Service" and "Media & Entertainment" sectors (e.g., Prezi, Avast, Reed).

Figure 5.3.2.1.4.1 – Distribution and Summary Statistics for the Number of Marketing Cookies with a Duration > 1 Year present on the Firms' Websites.



Obs.	=	1396
Average	=	6.75
SD	=	8.41
Min	=	0
Median	=	4
Max	=	56

Figure 5.3.2.1.4.2 – Distribution and Summary Statistics for the Number of Marketing Cookies with a Duration > 2 Years present on the Firms' Websites.



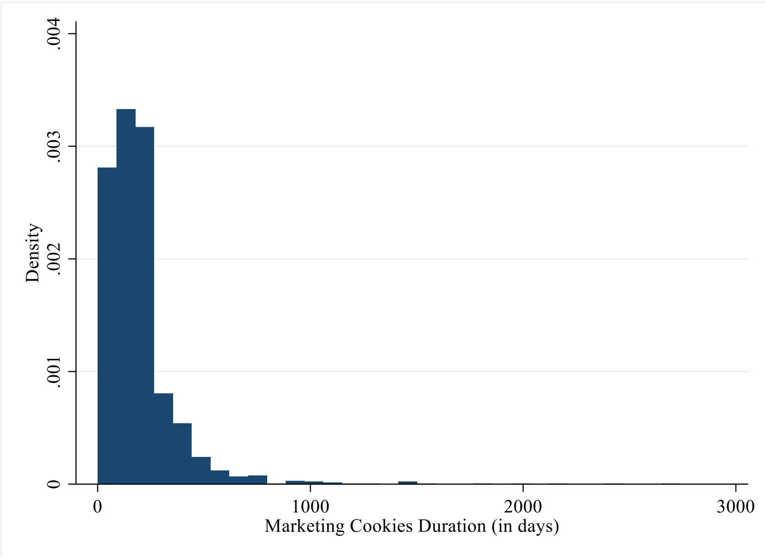
Obs.	=	1396
Average	=	0.92
SD	=	1.85
Min	=	0
Median	=	0
Max	=	17

5.3.2.1.5. Average Duration of Marketing Cookies

Another variable which we were able to calculate was the average duration of marketing cookies. This variable was calculated using the companies' Excel files provided by Cookiebot and computing the average of the duration – in days – of the cookies belonging to the "*Marketing*" category for each of the websites collected. Unfortunately, we had some problems using some files because they were not properly formatted (0.4%), and we also had to remove some of the companies since we recorded an average duration unbelievably high (4%) – greater than ten years. Consequently, we decided to substitute these anomalies with the mean of the variable not to have distortions.

By looking at the histogram and table below (Figure 5.3.2.1.5.1), we can see the same trends observed above for the average number of persistent marketing cookies and marketing cookies with long lifetime: most of the companies are storing marketing cookies for a short period of time while there are few companies which are probably strongly reliant on data collection and collects users' data for a longer than allowed period of time. The additional information we can get from this variable is that, on average, marketing cookies last 191 days (a little more than six months), and 90% of the companies set cookies with an average lifetime smaller than one year, in line with the *e-Privacy Directive*.

Figure 5.3.2.1.5.1 – Distribution and Summary Statistics for the Duration of the Marketing Cookies Domains present on the Firms' Websites (in days).



Obs.	=	1396
Average	=	190.78
SD	=	206.03
Min	=	0
Median	=	157
Max	=	2739

5.3.2.2. Other Cookie-Related Variables

The information about the number and types of cookies placed by companies on users' laptop can be obtained either by using online platforms such as Cookiebot – that can actually detect the real number of cookies set by websites emulating the user website navigation behavior – or by reading, if present, the cookie and privacy policies disclosed by the different companies on their website. Cookies policies are documents in which companies should describe the more relevant information about cookies and tracking technologies used on their websites. The GDPR establishes that companies have to obtain an explicit and informed opt-in, meaning that customers should be able to find, on the website visited, all the needed information about how data are being collected, for what purposes, who can have access to them, and for how long they are stored securely in the companies' databases. However, similarly to what happened with the re-permission emails, the Regulator has fixed all the requisites for compliance with the new data protection law without defining the modality in which these should be operationalized. Consequently, we observed a wide variety of cookie and privacy policies, and we found it interesting to map these differences. Additionally, some of the companies decided to openly disclose the names of the cookies used, allowing us to double-check the data obtained by Cookiebot and establish if "*companies say what they actually do*".

Following, we provide a description of the variables collected.

5.3.2.2.1. Type of Cookie Policy Communication

This is a manually coded variable trying to capture the variation among the Cookie Policies we observed. By observing the policies, we were able to define four different communications methodologies, which were coded in terms of completeness of the information provided:

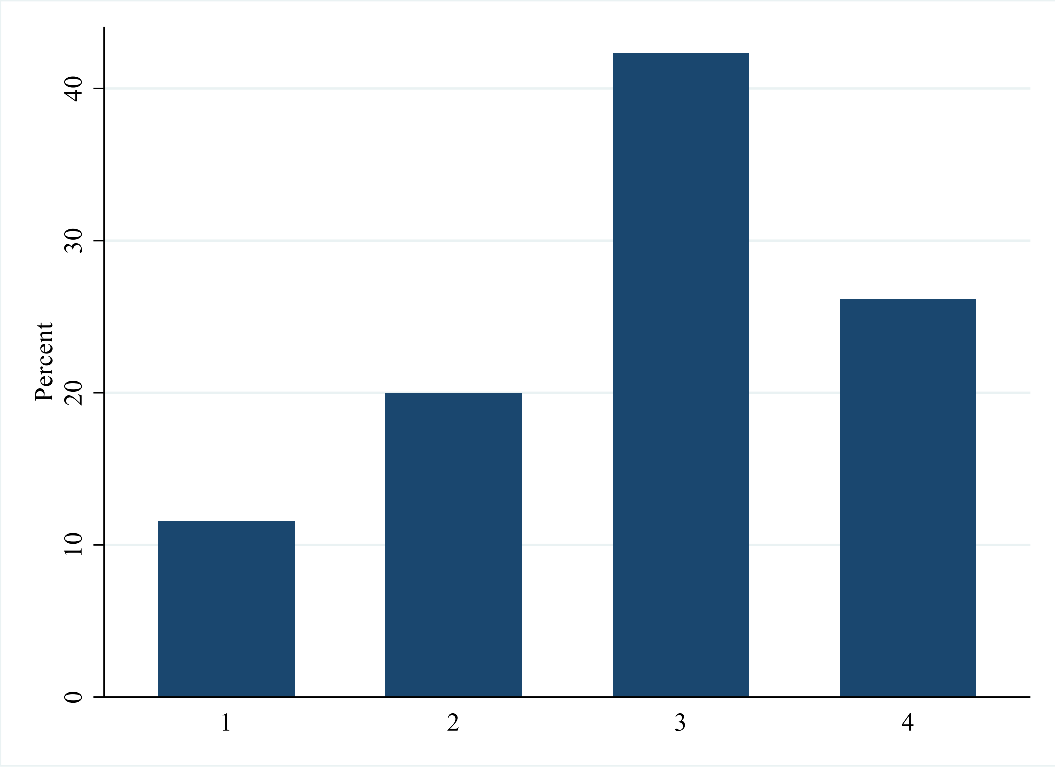
- 1: There is no cookie policy, or cookies are not mentioned or explained.
- 2: There is a referral to cookies and the general use of them on the website.
- 3: Cookies are explained in detail, and there is also some reference to the different types of cookies a website can install.

- 4: The cookie policy is rich in information about cookies and lists all the possible company's cookies used on its website.

As it is possible to understand from the histogram below (Figure 5.3.2.2.1.1), most companies are declaring and explaining to their customers what a cookie is, frequently providing specific and detailed information about their functioning, their different typologies, and the possible consequences for the user privacy. Only 12% of our sample did not provide any information about them.

However, only 26% of the companies in our sample actually list the specific cookies used on the website with a detailed description, for each cookie, of their purposes, their duration, and their provenience (e.g., *first* vs. *third*-party cookies).

Figure 5.3.2.2.1.1 – Bar Chart for the Type of Cookie Policy Used by the Firms on their Websites (N = 1396).



5.3.2.2.2. Number of Cookies Declared in the Cookie Policy

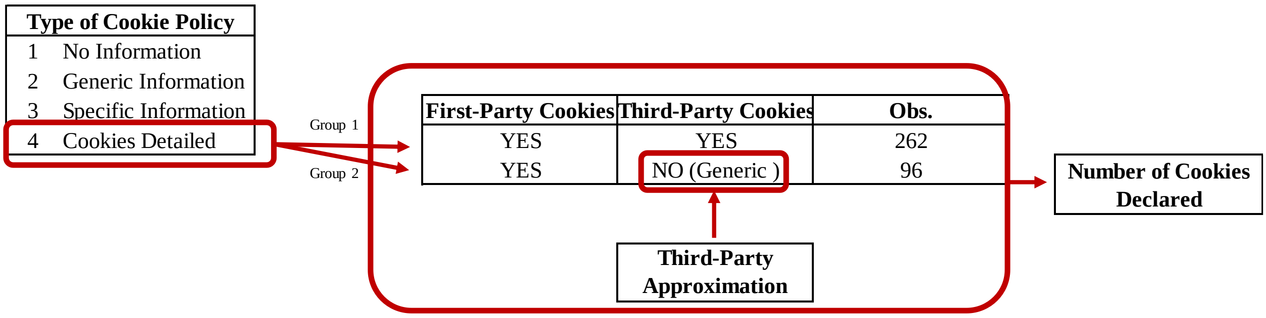
By selecting the companies that had a type of Cookie Policy coded as 4 – that provides not only a rich set of information about cookies but also specify in a detailed way all the individual cookies used on the webpage – it is also possible to manually count the cookies' declared by the company in the policy. This was done for 358 companies out of 1396 of the total sample (26%).

However, after a first inspection of the policies of the subsample identified, we found that the selected companies could be grouped into two categories:

- Those declaring, expressly, both *first*- and *third*-party cookies.
- Those declaring, specifically, the first-party cookies and approximatively the third-party cookies, often providing the link to the cookie policy of the third-party providers – such as Google, Facebook, Bing, or Twitter – to get information about their cookies.

Consequently, we found that 27% of the subsample selected belong to the second group. To cope with this issue, we better analyze the subsample of companies belonging to the first group; we then recorded the number of cookies for each specified third-party vendor declared by each company, and we averaged the total number of cookies by a third-party vendor, getting to a "*third-party approximation*". For example, we have recorded that Facebook provides, on average, 4 cookies, Google 8, Instagram 2, LinkedIn 10. The identification of the averages for each third-party provider allowed us to correct the number of cookies counted for those companies that not explicitly declared the number of third-party cookies. For example, we can have a case in which the company declares to use 10 first-party cookies, but also cookies from Google, Instagram, and Facebook. We can then estimate a total of 10 (first-party cookies) + 8 (Google) + 2 (Instagram) + 4 (Facebook) = 24 cookies for that company. Figure 5.3.2.2.2.1 provides the flow chart we follow to define the number of cookies declared by the 358 companies selected with and without third-party approximation.

Figure 5.3.2.2.2.1 – Flow Chart for the Definition of the Number of Cookies Declared in the ’Cookies’ Policies.



Type of Cookie Policy	
1	No Information
2	Generic Information
3	Specific Information
4	Cookies Detailed

First-Party Cookies	Third-Party Cookies	Obs.
YES	YES	262
YES	NO (Generic)	96

As it is possible to deduce from Table 5.3.2.2.2.1, companies are declaring, on average, 53 cookies, with much variability around it, having companies declaring not to use cookies at all and companies declaring up to 1164 total cookies. However, if we look, for the same companies, at the statistics for the total cookies recorded by Cookiebot, we can see that the average number of cookies jumps to 112, meaning that we have companies under-declaring the cookies used. This may suggest that firms do not show coherence between what they say and what they do, either voluntarily or involuntarily. Independently from the reasons behind this choice, it is important to highlight that these are not formally compliant with the GDPR requirements in case of inspection and may be severely fined.

Table 5.3.2.2.2.1 – Descriptive Statistics for the Number of Total Cookies Declared and Recorded for Firms with a Type of Cookie Policy coded as 4 (N = 358).

Variable	**Source**	**Obs.**	**Mean**	**SD**	**Min**	**Max**
Number of Cookies **Declared**	Cookie Policies	358	53.2	81.04	0	1164
Number of Cookies **Recorded**	Cookiebot	358	112.13	220.31	0	3000

5.3.2.2.3. Delta Between Cookies Declared and Cookies Recorded

By looking at the difference between cookies declared and cookies recorded – see Table 5.3.2.2.2.1 above – we decided to create another variable able to track the level of coherence with regards to what companies say (cookies declared) and do (cookies recorded). This type of analysis was performed for the 358 companies that reported, in their cookie policies, the specific cookies used on their web pages. The variable has been created using the following formula:

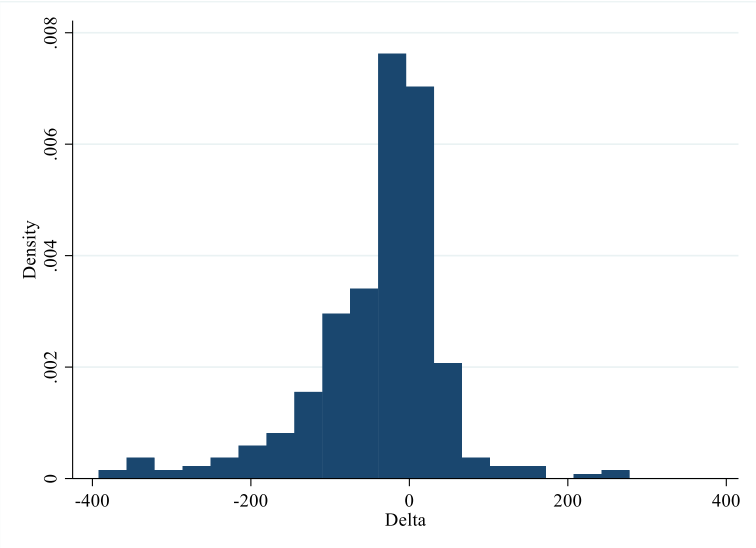
$Delta = Cookies\ Declared - Cookies\ Recorded$

and it can, consequently, present three modalities:

- *Delta greater than 0*: These are the cases in which companies are over-declaring the number of cookies placed on their websites. This can be due to the company's willingness to be compliant with the GDPR, declaring cookies that may be inactive or that the company forecast to use in the future.
- *Delta closes to 0*: These are companies that can be considered coherent in terms of communication and action regarding cookies. Small differences – either positive or negative – can be regarded as errors in the Cookiebot extraction or in the approximation done for the third-party cookies when not explicitly declared in the cookie policy.
- *Delta lower than 0*: These are, instead, companies under-declaring the cookies they are using on their websites. This may be due to a lot of factors, such as the inability of the company to keep track of the cookies used, the missed update of the cookie policy after the inclusion of additional cookies on the webpages, the under-estimation of the relevance of some of the cookies to be declared in the cookie policy – so the company decides not to report them – or, in the worst case, the low level of transparency of that companies.

By looking at the histogram and table below (Figure 5.3.2.2.3.1), we can see that the companies declaring cookies are mainly under-declaring them since the delta's mean is negative.

Figure 5.3.2.2.3.1 – Distribution and Summary Statistics for the Delta between the Cookies Declared and Recorded on the Firms' Websites (N = 358).



Obs.	=	358
Average	=	-59.14
SD	=	227.99
Min	=	-2964
Median	=	-21
Max	=	1051

Additionally, by looking at Table 5.3.2.2.3.1, we can further see that companies with a strongly negative delta are those in the "*Media & Entertainment*" and "*Professional Services*" sectors, while those with a delta close to zero are those in in the "*Software and IT Services*" sector. This may suggest that companies with higher expertise in internet-related technologies are more likely to be careful and precise in disclosing information about the cookies used because they have both the knowledge and the technical tools to do so.

Table 5.3.2.2.3.1 – Mean of the Delta by US Industries.

US Industry Classification	**Mean of Delta**
Media and Entertainment	-84.86
Professional Services	-72.28
Retail Trade	-61.28
Travel, Tourism and Hospitality	-58.48
Others	-54.93
Software and IT Services	-14.54
Total	**-59.14**

5.3.2.3. Measures of Website’ Online Performance

In order to get measures that allow us to evaluate the online performances of the companies in our sample, we decided to collect two types of information for each website of our sample:

- Its *Popularity;*
- The *Advertising Revenues* that it can generate.

These two metrics will allow us to see whether the type of communication crafted by a company regarding customers’ privacy and data protection is related to the ability of the same company to extract value from the data collection. In particular, the former measure has to do with the data availability – meaning that the higher the website popularity, the higher the traffic it attracts, and the more the variety and richness of the data that it can provide – while the latter is more about the real monetary value that company may extract since it is a direct measure of the value of the traffic that the website attracts and that can be used for marketing activities.

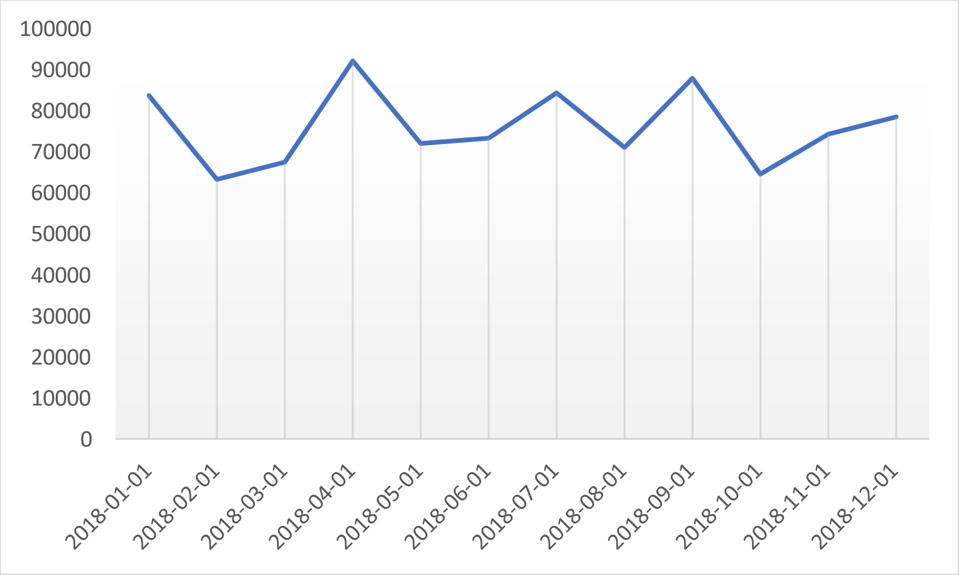
5.3.2.3.1. Website Popularity: Alexa Ranking

One of the most widespread measures used in academic papers and business press to get a sense of the website's popularity is the Alexa Traffic Ranking (Libert and Nielsen 2018; Peukert et al. 2020). This is an Amazon proprietary measure that combines the website traffic statistics with the visitor engagement data – estimated using a panel of global users – over a period of three months, ranking millions of websites in order of popularity. It returns an index that can be used to compare websites’ popularity over time and among each other, with rank n. 1 being the most popular website – a position taken by *Google.com* as of July 2020.

We collected this information by using the official Amazon AWIS API. We recorded, for each company present in our database, the Amazon Alexa Rankings from January 2018 to December 2018 on a monthly basis (every 1st day of the month). Unfortunately, not all the companies were ranked every month in the observation period, meaning that we could get the Alexa Rank score for about 61% of the companies composing our sample. Consequently, the trend plot displayed in Figure

5.3.2.3.1.1 shows the average Alexa Ranking by month for the subset of 848 companies – for which the data are available.

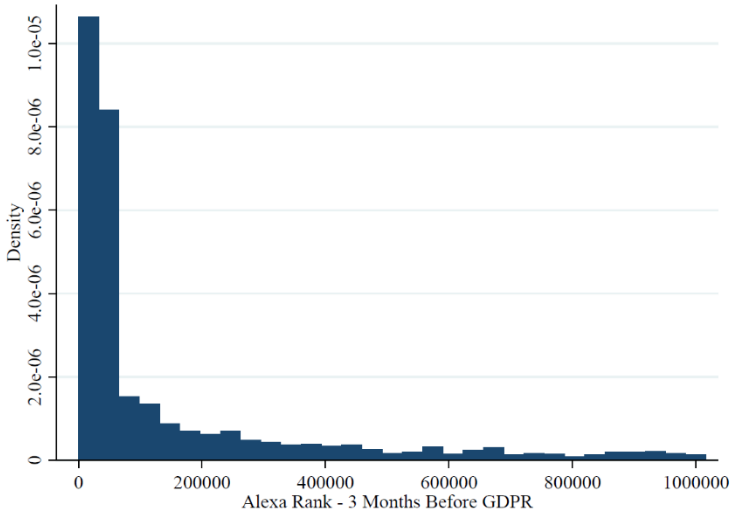
Figure 5.3.2.3.1.1 – Alexa Ranking Trend Over Time in 2018 (N = 848)



We decided to collect this information in order to be able to control, in our models, for the websites' capacity of attracting traffic, generate engagement and collect valuable users' data, just before the GDPR established new rules for the European digital scenario. This would partially account for the pre-GDPR companies' online business plan. Companies with a low rank before May 2018 (e.g., with high popularity) may have had an additional incentive to behave opportunistically because they strongly rely on the digital ecosystem for their daily operation and may perceive, to a greater degree, the risks that privacy regulations entail (e.g., lower profitability, ad effectiveness).

Therefore, we averaged the Alexa Rank scores of the three months before the GDPR enforcement (May 2018). Also in this case, by inspecting the data collected, not all the companies are ranked each of the three months, meaning that we could get the average of the Alexa Rank score for 85% of the companies composing our sample. In other words, we have 211 companies for which this information was not available for the months of February, March, and April 2018. We decided to replace the missing values using the median of the Alexa Rank scores available for the remaining part of the companies' sample in order to be able to use the entire set of companies in our subsequent analysis. Figure 5.3.2.3.1.2 shows the distribution and the summary statistics for the Averaged Alexa Rankings of the companies. Additionally, Figure 5.3.2.3.1.3 shows the "*best*" and the "*worst*" companies making up our sample – in terms of online popularity. On the one side, Google, Facebook, and YouTube occupy the top positions and can be considered as the most popular ones; on the other side, Nisa Europe, AR Hotels, and Warp Academy are companies that are far less popular ones, scoring very high values in their Alexa scores. Therefore, as highlighted, we have a lot of variability in our sample: while we do have included in our sample best in class (ranked as first), we also have collected re-permission emails belonging to less popular online firms.

Figure 5.3.2.3.1.2 - Distribution and Summary Statistics for the averaged Alexa Rank Scores for the Three Months Before the GDPR Enforcement (N = 1396)



Obs.	=	1,396
Average	=	151,271
SD	=	224,777
Min	=	1
Median	=	62,542
Max	=	1,016,361

Figure 5.3.2.3.1.3 – Examples of Most and Least Popular Firms in our Sample of Companies.

Most Popular Firms	
Firm	**Averaged Alexa Rank**
Google	1.30
YouTube	1.67
Facebook	3.00
Yahoo	6.33
Twitter	13.33
Instagram	16.00
Netflix	29.67
Linkedin	38.00
Microsoft	45.33
Aliexpress	54.00

Least Popular Firms	
Firm	**Averaged Alexa Rank**
Nisa Europe	1016361.00
AR Hoteles	1006585.00
Warp Academy	985944.70
Xcite	985335.00
Tuxedo	984307.00
Paperwave	978931.00
Natural Collection	967087.50
FSCS	961094.00
Collistar	945843.00
Android Weekly	935491.00

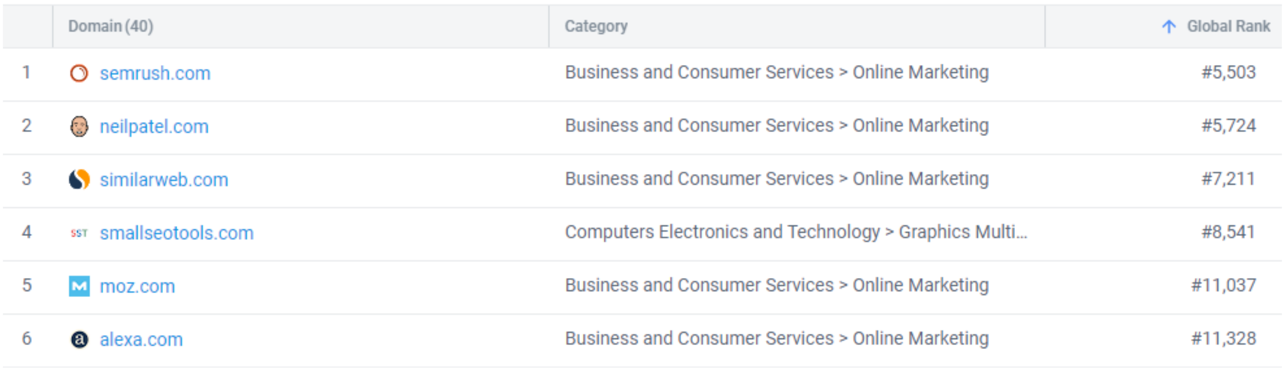
5.3.2.3.2. Measure of Website' Online Performance: Expected Online Ad Revenue

In order to evaluate the online performances of the companies in our sample, we decided to collect information about the *expected online advertising revenues* that the companies' websites can generate. This metric will allow us to see whether the type of communication crafted by a company regarding customers' privacy and data protection is related to the ability of the same company to extract value from the data collection. It can be considered as the monetary value of the traffic that the website attracts, in that becoming a measure of the expected value of the data generated in a firm's website. To the best of my knowledge, this project is one of the first to measure the expected online advertising revenue that websites are able to generate.

First, we retrieved this information using SEMrush, an online platform through which we were able to get a rough estimate of the expected online ad revenues of our sample of firms' websites. Second, we decided to use another source (Rank2Traffic) as a robustness check.

SEMrush is an online platform collecting information about 790 million domains using 5 million users worldwide, providing a wide variety of information for digital marketers. It has been used by leading companies such as eBay, Booking.com, and Quora, and it has been acknowledged with numerous digital marketing prizes in 2019. Additionally, it is one of the top websites for online marketing services according to the SimilarWeb ranking (Figure 5.3.2.3.2.1), another well-established traffic analytics tool used by experts in the digital realm. Consequently, given the reliability of the data provided (see Appendix B for additional checks) and data availability about companies' expected online advertising revenues and other traffic statistics, we decided to collaborate with this company and purchase its data access.

Figure 5.3.2.3.2.1 – SimilarWeb Ranking of Online Marketing Platforms.



	Domain (40)	Category	↑ Global Rank
1	semrush.com	Business and Consumer Services > Online Marketing	#5,503
2	neilpatel.com	Business and Consumer Services > Online Marketing	#5,724
3	similarweb.com	Business and Consumer Services > Online Marketing	#7,211
4	smallseotools.com	Computers Electronics and Technology > Graphics Multi...	#8,541
5	moz.com	Business and Consumer Services > Online Marketing	#11,037
6	alexa.com	Business and Consumer Services > Online Marketing	#11,328

Source: SimilarWeb

It is essential to highlight that the information provided by SEMrush is based only on the Google Ad Network data; consequently, we were able to extract information about the value of advertisement placed through Google Ad Sense. We could not have information about the revenue from ads placed on Facebook, Twitter, or Amazon, meaning that the information we have should be considered only a part of the total value from advertisements that each company can generate from its website's advertising. According to a recent eMarketer report (2020), Google ad revenues in the US are in a slight decline in favor of Facebook and Amazon ad revenues, which are, instead, growing. However, even if Google Ad Network seems less attractive for advertisers in 2020, it is still the global market leader for ad selling with about a 30% market share (EMarketer 2020). Consequently, we can be reassured that we are, at least, capturing a considerable part of the revenue companies in our sample are making out of advertising strategies.

Additionally, SEMrush shared the calculation they used to get the expected online advertising revenue for each company’s website. More specifically, the measure for website j is obtained as follows:

$$\begin{aligned} \mathit{Ad\ Revenue}_j = {} & \mathit{Max\ Monthly}\ \boldsymbol{Traffic}\ \mathit{of\ the\ Website\ j\ in\ Country\ c} \\ & *\ \mathit{Monthly}\ \boldsymbol{CostPerClick}\ \mathit{of\ \ Website\ j's\ Industry\ in\ Country\ c} \\ & *\ \boldsymbol{ClickThroughRate}\ \mathit{of\ Website\ j's\ Industry\ in\ \ Country\ c} \end{aligned}$$

As it is possible to infer from the investigation of the formula above, the measure of the ad revenue produced by SEMrush can be seen and interpreted as an estimation of the potential future revenue stream from advertising that each website in our dataset can generate on Google AdSense Network. Therefore, it is possible that the revenue estimated for a specific website is not the real advertising revenue that it is generating at the moment of the estimation. Consequently, it should be seen as the “*potential revenue a website could make if they monetized their site by publishing advertisements via Google AdSense*”[10].

In this regard, it may be helpful the distinction between the functions that a website can have. A website may act either as a publisher, an advertiser, or both.

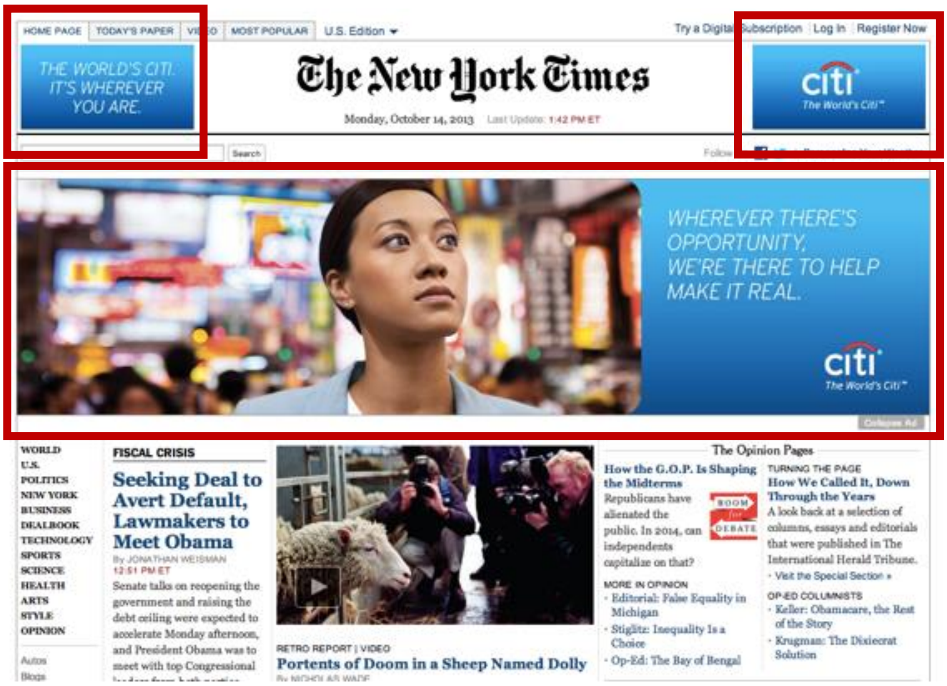
In the first case, the website is actually able to make money out of advertisements since it acts as a forum for all the advertisements produced by the external websites (e.g., advertisers who have bought ad spaces on the webpage); for example, this is the case of websites belonging to the news sector, such as “T*he New York Times*”, that mainly display advertising and sell ad spaces as a business model (Figure 5.3.2.3.2.2).

In the second case, instead, the website is mainly posting its advertisements on other websites (e.g., publishers). This is the case in which the company – that creates the advertising – buys ad spaces on external websites – which acts as publishers – to reach its audience. For example, websites such as Underarmour.com, Gucci.com, or Technogym.com are all example of advertisers type of websites,

[10] https://www.SemRush.com/kb/1008-adsense-benchmark-tool

meaning that they are very unlikely to sell ad spaces on their own web pages but are more probably using other websites (such as facebook.com or google.com) to show their advertisements (Figure 5.3.2.3.2.2).

Figure 5.3.2.3.2.2 – Example of a Publisher and an Advertiser type of websites.
In this example, "The New York Times" is acting as a publisher, whereas the "City Bank" as an advertiser.



Monday, October 14, 2013 Last Update: 1:42 PM ET

FISCAL CRISIS

Seeking Deal to Avert Default, Lawmakers to Meet Obama

By JONATHAN WEISMAN
12:51 PM ET

Senate talks on reopening the government and raising the debt ceiling were expected to accelerate Monday afternoon, and President Obama was to meet with top Congressional leaders from both parties

RETRO REPORT | VIDEO

Portents of Doom in a Sheep Named Dolly

By NICHOLAS WADE

The Opinion Pages

How the G.O.P. Is Shaping the Midterms

Republicans have alienated the public. In 2014, can independents capitalize on that?

MORE IN OPINION
- Editorial: False Equality in Michigan
- Stiglitz: Inequality Is a Choice
- Op-Ed: The Bay of Bengal

TURNING THE PAGE

How We Called It, Down Through the Years

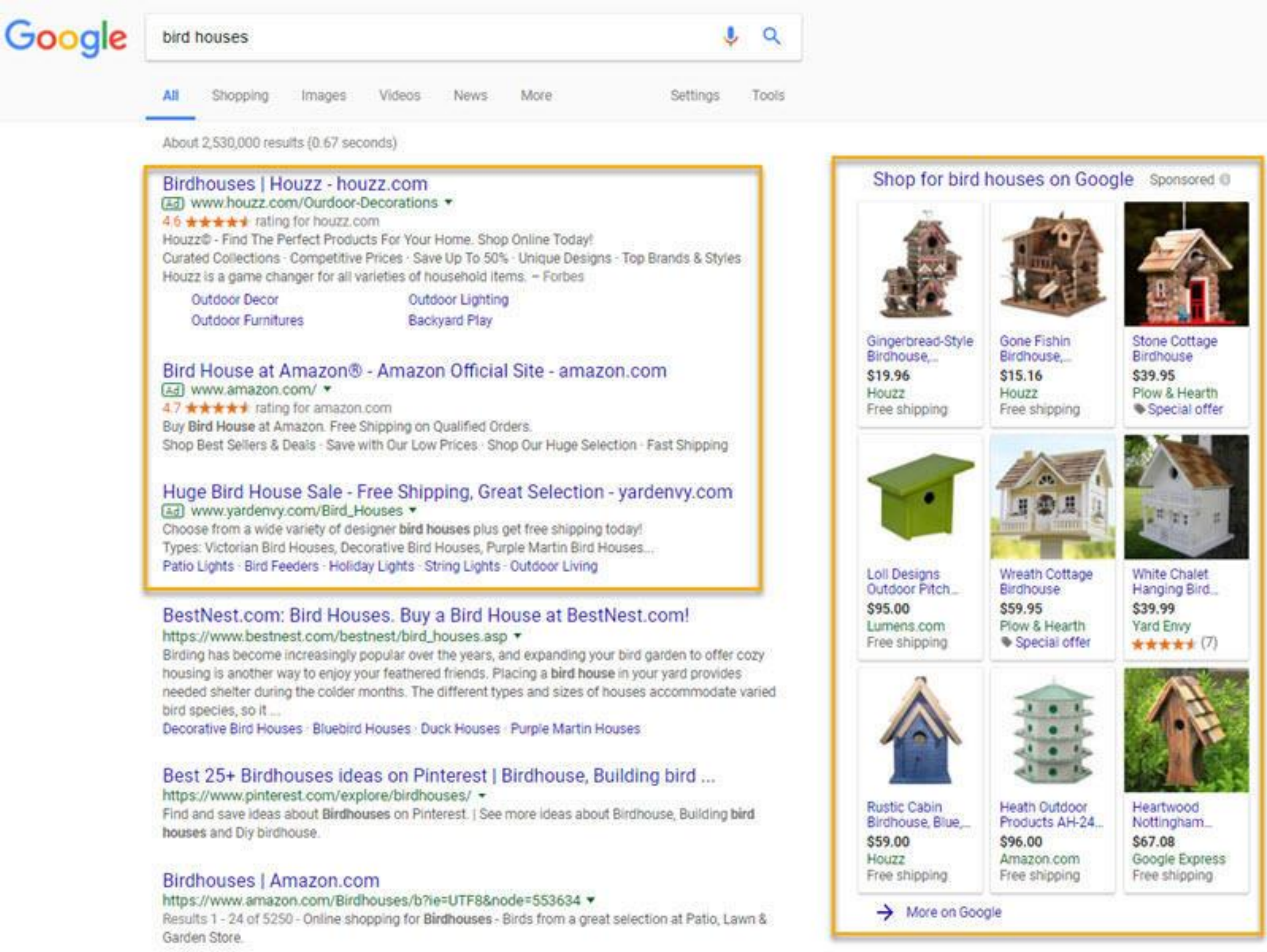
A look back at a selection of columns, essays and editorials that were published in The International Herald Tribune.
- Visit the Special Section »

OP-ED COLUMNISTS
- Keller: Obamacare, the Rest of the Story
- Krugman: The Dixiecrat Solution

Lastly, the third case represents websites that we can call as "*hybrid*" solutions that act as both publishers and advertisers. This is, for example, the case of big research engines such as Google, which both sell ad spaces (e.g., Google Announces) and publish advertising about its additional services and products (e.g., Google Fiber, Google Docs, Google AdSense, Google AdMob) on external websites (Figure 5.3.2.3.2.3 Panel A and Panel B).

Figure 5.3.2.3.2.3 – Example of Google as “Hybrid” websites.

Panel A: Google as Advertiser. Example of advertisements that Google is displaying on its main webpage.



About 2,530,000 results (0.67 seconds)

Birdhouses | Houzz - houzz.com
Ad www.houzz.com/Ourdoor-Decorations
4.6 ★★★★★ rating for houzz.com
Houzz© - Find The Perfect Products For Your Home. Shop Online Today!
Curated Collections · Competitive Prices · Save Up To 50% · Unique Designs · Top Brands & Styles
Houzz is a game changer for all varieties of household items. – Forbes
Outdoor Decor · Outdoor Lighting
Outdoor Furnitures · Backyard Play

Bird House at Amazon® - Amazon Official Site - amazon.com
Ad www.amazon.com/
4.7 ★★★★★ rating for amazon.com
Buy **Bird House** at Amazon. Free Shipping on Qualified Orders.
Shop Best Sellers & Deals · Save with Our Low Prices · Shop Our Huge Selection · Fast Shipping

Huge Bird House Sale - Free Shipping, Great Selection - yardenvy.com
Ad www.yardenvy.com/Bird_Houses
Choose from a wide variety of designer **bird houses** plus get free shipping today!
Types: Victorian Bird Houses, Decorative Bird Houses, Purple Martin Bird Houses...
Patio Lights · Bird Feeders · Holiday Lights · String Lights · Outdoor Living

BestNest.com: Bird Houses. Buy a Bird House at BestNest.com!
https://www.bestnest.com/bestnest/bird_houses.asp
Birding has become increasingly popular over the years, and expanding your bird garden to offer cozy housing is another way to enjoy your feathered friends. Placing a **bird house** in your yard provides needed shelter during the colder months. The different types and sizes of houses accommodate varied bird species, so it ...
Decorative Bird Houses · Bluebird Houses · Duck Houses · Purple Martin Houses

Best 25+ Birdhouses ideas on Pinterest | Birdhouse, Building bird ...
https://www.pinterest.com/explore/birdhouses/
Find and save ideas about **Birdhouses** on Pinterest. | See more ideas about Birdhouse, Building **bird houses** and Diy birdhouse.

Birdhouses | Amazon.com
https://www.amazon.com/Birdhouses/b?ie=UTF8&node=553634
Results 1 - 24 of 5250 - Online shopping for **Birdhouses** - Birds from a great selection at Patio, Lawn & Garden Store.

Shop for bird houses on Google Sponsored

Gingerbread-Style Birdhouse,...
$19.96
Houzz
Free shipping

Gone Fishin Birdhouse,...
$15.16
Houzz
Free shipping

Stone Cottage Birdhouse
$39.95
Plow & Hearth
Special offer

Loll Designs Outdoor Pitch...
$95.00
Lumens.com
Free shipping

Wreath Cottage Birdhouse
$59.95
Plow & Hearth
Special offer

White Chalet Hanging Bird...
$39.99
Yard Envy
★★★★★ (7)

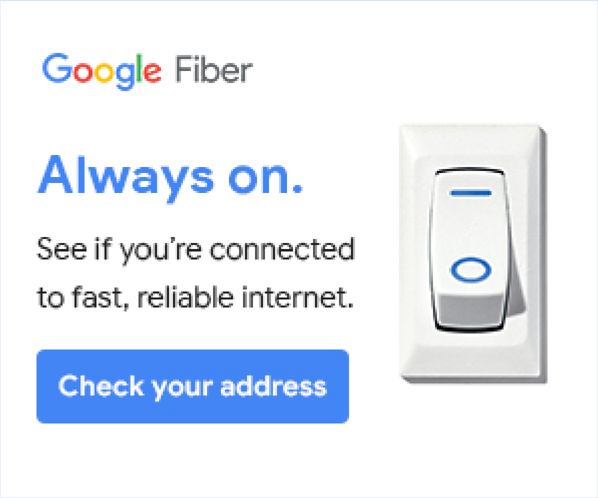
Rustic Cabin Birdhouse, Blue,...
$59.00
Houzz
Free shipping

Heath Outdoor Products AH-24...
$96.00
Amazon.com
Free shipping

Heartwood Nottingham...
$67.08
Google Express
Free shipping

→ More on Google

Panel B: Google as Publisher. Example of the advertisement that Google is displaying on other websites.



Therefore, the distinction between the function that a website may have concerning data exploitation and monetization is of fundamental importance to interpret the "*ad revenue*" figure generated by SEMrush. It may help to understand whether a website:

- Is already exploiting and monetizing data to a great degree, and, consequently, it may be severely hurt by a new privacy regulation (e.g., publishers' type of websites).
- Has an unexpressed potential to generate revenue from the data it can collect given the traffic and the industry it belongs to (e.g., advertisers' type of websites).

Consequently, we collected information about how each company in our sample uses its website – in terms of the distinction in the website usage just described. We recorded the cumulative[11] number of advertisements displayed by the company on its website and of advertisements published by the company on other websites. We also retrieved the last available date on which the advertisements have been displayed to an audience ("*last seen date*"). Unfortunately, we detect a bug in the SEMrush recordings for publishers that have shown advertisements between November 26, 2016, and December 5, 2016. Consequently, we were forced not to consider websites as publishers if their "*last seen date*" falls in that period.

Thanks to this information, we can classify the companies in our sample into three main categories: publishers, advertisers, and hybrid. By looking at Table 5.3.2.3.2.1, we can overview the composition of our sample. As it is possible to see, companies in our sample are predominantly advertisers (63%), meaning that they rather use other websites to display their ads than using their websites to show other's companies' advertisements. Additionally, it is also possible to see that there is 6% of "*hybrid*" companies acting as both advertisers and publishers. However, as highlighted above, our collection of data about the "*publishers*" suffered some problems. Therefore, given the small proportion of websites acting as publishers, it is not possible to compare the ad revenues of these different categories of websites.

[11] SEMrush provides the total number of advertisements of the company (as both publisher and advertiser) starting from the first detection they have recorded in their databases.

Table 5.3.2.3.2.1 – Frequency Table of Publishers vs. Advertisers Companies in our Sample (N = 1396).

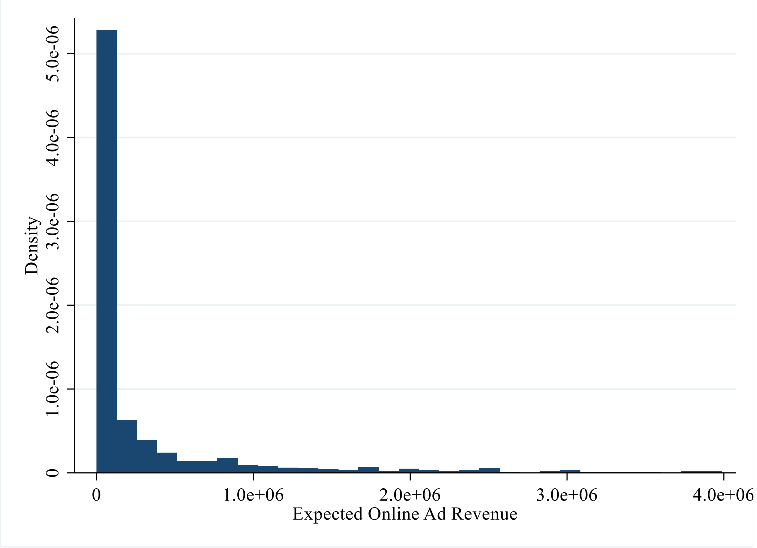
		Publisher			Total
		Yes	No	Missing	
Advertiser	Yes	89 6%	882 63%	0 0%	971 69%
	No	12 1%	0 0%	0 0%	12 1%
	Missing	0 0%	0 0%	413 30%	413 30%
Total		101 7%	882 63%	413 30%	1396 100%

All in all, given our sample's composition just described, we will interpret the advertising revenue measure more as a future potential revenue stream than as a real monetization that companies are making out of the data collection.

Figure 5.3.2.3.2.4 provides the overall distribution, and descriptive statistics of the SEMrush expected online ad revenues estimate – collected in October 2020 – which we will use later in the analysis. It is essential to highlight that since the ad revenue measure is positive in nature, we have also created the logarithmic transformation of our original variable to use in our analyses (Figure 5.3.2.3.2.5).

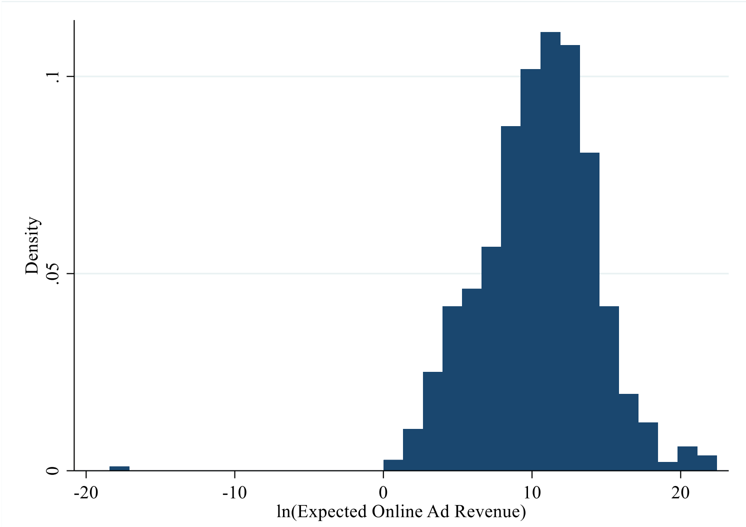
As it is possible to see, the companies composing our sample can mainly generate modest online ad revenue streams from their websites, which may be in line with the fact that most of them are advertisers rather than publishers. However, the data also show a wide variety in this variable distribution. In fact, there are companies in our dataset that are mainly operating in the online setting, which, according to the SEMrush estimation procedure, have a great potential to generate online ad revenue streams (e.g., Google, Booking.com, Etsy, Asos).

Figure 5.3.2.3.2.4 – Distribution and Summary Statistics for the Expected Online Ad Revenue Measure collected in October 2020 (N = 1255).



Obs.	=	1,255
Average	=	19,800,000
SD	=	293,000,000
Min	=	0
Median	=	31,147
Max	=	5,510,000,000

Figure 5.3.2.3.2.5 – Distribution and Summary Statistics for the log(Expected Online Ad Revenue) Measure collected in October 2020 (N = 1255).



Obs.	=	1,255
Average	=	10.08
SD	=	3.73
Min	=	-18.42
Median	=	10.35
Max	=	22.43

5.3.2.4. Summary of the Website-Related Variables

For simplicity, we summarize the main website-related variables just described in Table 5.3.2.4.1 below. Our sample is characterized by firms that are *collecting a lot of tracking data* from their websites, are *not too transparent* with regards to the *cookies* they are using, and *have the potential to generate considerable revenue streams* from the advertisements placed on their websites.

Table 5.3.2.4.1 – Summary of the Main Results for the Variables Related to the Firms' Websites Characteristics.

Variable	**Operationalization**	**% of Missing Values**	**Main Descriptive Results** **Our sample is composed by:**
Number of Marketing Cookies	Continuous variable counting the number of tracking cookies present on the companies' websites.	0%	Companies with a *lot of Marketing Cookies* placed by *different third-party domains*, which tend to *last for a long period of time* to capture users' behaviors.
Number of Marketing Cookies Domains	Continuous variable counting the number of external websites which are linking to the companies' websites through a Marketing Cookie.	0%	
Number of Persistent Marketing Cookies	Continuous variable counting the number of persistent tracking cookies present on the companies' websites.	0%	
Type of Cookie Communication	Categorical variable with 4 levels indicating the degree of disclosure of the cookie policy published on the company's website.	0%	Companies that have a high degree of disclosure about its cookies' practices, but tend to *not disclose the specific cookies* used on their webpages. Moreover, when they decide to do so, they also tend to *under-declare the cookies* used.
Delta between the Cookies Declared and the Cookies Recorded	Continuous variable indicating the difference between the number of cookies that the company declared in its cookie policy and the number of cookies actually detected on its website.	74%	
Popularity	Averaged Alexa Rank for the three months before the GDPR enforcement (February, March and April 2018).	15%	Companies with *a lot of variety in their online popularity*. We do have the most popular ones but also far less popular companies.
Ad Revenue	Continuous variable indicating the future possible online advertising revenue stream that a company can produce on its website.	10%	Companies that may generate considerable *online ad revenue* streams and are *mainly advertisers, rather than publishers*.

6. Analysis of the Content of the Re-Permission Emails

In this first part of the thesis, we are interested in the content of re-permission emails and in the specific themes that firms used in their GDPR email campaigns.

In our conceptual framework (Chapter 4.1.1), we outlined six main themes: two *informational* (control, transparency) and four *persuasive* (incentives, both monetary and non-monetary, and framing of the message – in terms of gains and losses and time orientation) that previous literature has identified as factors influencing individuals' likelihood to disclose personal data. The question is whether firms actually used them. To verify that, we conduct an in-depth content analysis of the GDPR re-permission emails collected using a three-stage approach.

First, we content analyzed our sample of emails through the use of a data-driven technique – the Latent Dirichlet Allocation (Blei, Ng, and Jordan 2003) – which allows us to detect latent topics in an unsupervised way without using pre-defined constructs and information about privacy-related communications. This procedure identified three main overarching topics in our emails: *informative*, *persuasive*, *Neither Highly Informative Nor Highly Persuasive***.**

Second, we turned to theory, and we identified the six main themes that literature found to be relevant in prompting consumers' disclosure behavior (Chapter 3.2). Then, we selected a random sub-sample of emails and developed a stable and reliable approach to content-analyze emails by integrating human (e.g., manual coding) and automated (e.g., LIWC and TextEvaluator) interventions.

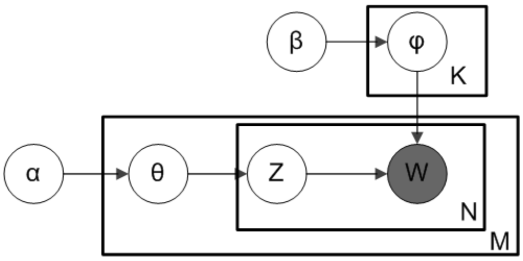
Lastly, in the third stage, we used the methodology identified in the second stage to predict the presence of the previously identified topics in all the re-permission emails collected.

Interestingly, the availability of results from two different content analysis methods allowed us to cross-validate them one against the other and to show that we were able to reach consistency in the results (see Appendix E for additional corroboration checks).

6.1. Stage 1: Latent Dirichlet Allocation Analysis

The first stage of the content analysis procedure uses Natural Language Process (NLP) techniques to process the text of the re-permission emails collected in our sample. More specifically, we used the Latent Dirichlet Allocation (Blei, Ng, and Jordan 2003) to analyze in an unsupervised fashion the data permission requests included in our database to detect several latent topics. Figure 6.1.1 outlines the model we specified.

Figure 6.1.1 – Graphic Model for LDA with Dirichlet-Distributed Topic-Word Distributions.



Given a corpus of M documents made of N words, it is possible to construct a dictionary of unique terms (W) used across the documents, which are the only observable variables among the ones depicted in Figure 6.1.1. Given this dictionary is then possible to construct the LDA model, which assumes that documents are a random mixture of K latent topics generated by:

- Picking a topic distribution for document m (θ_m) from a Dirichlet distribution with hyper-parameter α.
- Picking a word distribution for topic k (φ_k) from a Dirichlet distribution with hyper-parameter β.
- For each of the word positions (i, j) available in a document of length N_i, where $i \in \{1, \dots, M\}$ and $j \in \{1, \dots, N_i\}$:
    - Choose a topic $z_{i,j} \sim Multinomial\ (\theta_i)$
    - Choose a word $w_{i,j} \sim Multinomial\ \left(\varphi_{z_{i,j}}\right)$

This process is then re-iterated for each word in each document of the corpus.

The topic modeling analysis has been conducted using the *Gensim* package available in Python. Firstly, the usual operation of data cleaning has been brought about: the texts have been firstly split into words – tokenization – then stop-words have been removed, and words have been lemmatized. Additionally, tokens with a frequency higher than 90% or lower than 1% have been removed since too common or too rare to be useful for the analysis (Griffiths and Steyvers 2004; Lu et al. 2011), getting to a final dictionary of 1014 unique words.

In order to run an LDA model, it is necessary to specify, in an *a priori* fashion, the number of hidden topics to seek in the texts (k); consequently, to determine the optimal value for k, we have estimated different LDA models with a different number of topics, and we have compared them using the "*coherence score*" measure. This measure "*scores a single topic by measuring the degree of similarity between high scoring words in the topic*" (Kapadia 2019), helping in discriminating between the topics identified. Therefore, higher values of the coherence score measure suggest that it is possible to clearly distinguish between the topics chosen, whereas lower scores suggest overlaps and cross-contaminations between the topics. By looking at Table 6.1.1, it is possible to see that the best choice of k is obtained by estimating models with three topics since they return the highest coherence value scores (Mean = 0.48; Median = 0.50) (AlSumait et al. 2009; Mimno et al. 2011; Puranam, Narayan, and Kadiyali 2017).

Table 6.1.1 - Coherence Scores of LDA Models with Different Values of *k*.

	Number of Topics									
Model	1 Topics	2 Topics	**3** **Topics**	4 Topics	5 Topics	6 Topics	7 Topics	8 Topics	9 Topics	10 Topics
1	0.325	0.488	**0.475**	0.431	0.448	0.434	0.457	0.431	0.428	0.436
2	0.325	0.488	**0.488**	0.520	0.526	0.426	0.454	0.472	0.409	0.446
3	0.325	0.488	**0.490**	0.464	0.460	0.391	0.399	0.440	0.397	0.424
4	0.325	0.488	**0.490**	0.484	0.455	0.384	0.359	0.393	0.408	0.447
5	0.325	0.488	**0.490**	0.530	0.531	0.439	0.393	0.426	0.375	0.389
6	0.325	0.488	**0.480**	0.468	0.457	0.438	0.487	0.447	0.405	0.399
7	0.325	0.488	**0.486**	0.520	0.497	0.397	0.429	0.404	0.452	0.417
8	0.325	0.488	**0.490**	0.461	0.502	0.446	0.404	0.368	0.354	0.400
9	0.325	0.488	**0.480**	0.521	0.488	0.458	0.388	0.390	0.408	0.424
10	0.325	0.488	**0.490**	0.517	0.504	0.397	0.433	0.445	0.421	0.404
11	0.325	0.488	**0.481**	0.537	0.442	0.440	0.428	0.416	0.427	0.447
12	0.325	0.488	**0.490**	0.518	0.464	0.396	0.457	0.447	0.394	0.427
13	0.325	0.488	**0.490**	0.440	0.483	0.449	0.396	0.408	0.445	0.429
14	0.325	0.488	**0.488**	0.514	0.473	0.461	0.421	0.426	0.442	0.387
15	0.325	0.488	**0.481**	0.457	0.499	0.456	0.395	0.380	0.432	0.400
16	0.325	0.488	**0.502**	0.512	0.514	0.431	0.415	0.426	0.423	0.405
17	0.325	0.488	**0.490**	0.495	0.476	0.409	0.374	0.388	0.430	0.372
18	0.325	0.488	**0.383**	0.515	0.517	0.448	0.433	0.451	0.390	0.430
19	0.325	0.488	**0.490**	0.492	0.452	0.447	0.445	0.423	0.406	0.432
20	0.325	0.488	**0.484**	0.464	0.505	0.423	0.423	0.424	0.424	0.406
21	0.325	0.488	**0.488**	0.457	0.442	0.478	0.397	0.421	0.487	0.409
22	0.325	0.488	**0.490**	0.499	0.480	0.442	0.414	0.458	0.432	0.406
23	0.325	0.488	**0.500**	0.520	0.516	0.453	0.402	0.396	0.404	0.400
24	0.325	0.488	**0.480**	0.454	0.421	0.461	0.423	0.416	0.433	0.395
25	0.325	0.488	**0.490**	0.471	0.508	0.392	0.415	0.424	0.407	0.435
26	0.325	0.488	**0.480**	0.447	0.486	0.457	0.435	0.407	0.459	0.345
27	0.325	0.488	**0.490**	0.410	0.439	0.437	0.451	0.404	0.393	0.412
28	0.325	0.488	**0.481**	0.466	0.441	0.441	0.463	0.417	0.428	0.410
29	0.325	0.488	**0.488**	0.514	0.472	0.440	0.442	0.459	0.396	0.366
30	0.325	0.488	**0.490**	0.488	0.510	0.430	0.457	0.392	0.442	0.444
31	0.325	0.488	**0.488**	0.508	0.496	0.467	0.399	0.423	0.346	0.422
32	0.325	0.488	**0.487**	0.473	0.497	0.464	0.421	0.443	0.411	0.440
33	0.325	0.488	**0.484**	0.478	0.492	0.443	0.402	0.440	0.411	0.411
34	0.325	0.488	**0.490**	0.431	0.465	0.430	0.448	0.400	0.458	0.401
35	0.325	0.488	**0.490**	0.470	0.435	0.453	0.426	0.430	0.433	0.462
36	0.325	0.488	**0.480**	0.509	0.454	0.455	0.419	0.423	0.399	0.378
37	0.325	0.488	**0.475**	0.507	0.496	0.416	0.453	0.378	0.465	0.397
38	0.325	0.488	**0.484**	0.422	0.490	0.416	0.463	0.476	0.450	0.406
39	0.325	0.488	**0.484**	0.479	0.448	0.477	0.437	0.447	0.428	0.418
40	0.325	0.488	**0.500**	0.474	0.472	0.459	0.375	0.442	0.374	0.435
41	0.325	0.488	**0.490**	0.523	0.455	0.417	0.386	0.388	0.442	0.448
42	0.325	0.488	**0.480**	0.523	0.506	0.443	0.488	0.380	0.377	0.393
43	0.325	0.488	**0.499**	0.421	0.477	0.416	0.473	0.446	0.398	0.425
44	0.325	0.488	**0.473**	0.508	0.507	0.384	0.404	0.429	0.455	0.381
45	0.325	0.488	**0.474**	0.453	0.452	0.450	0.420	0.424	0.402	0.430
46	0.325	0.488	**0.484**	0.493	0.464	0.428	0.458	0.476	0.423	0.413
47	0.325	0.488	**0.490**	0.473	0.417	0.447	0.406	0.404	0.386	0.386
48	0.325	0.488	**0.490**	0.524	0.469	0.483	0.388	0.388	0.376	0.436
49	0.325	0.488	**0.475**	0.454	0.434	0.451	0.423	0.414	0.405	0.438
50	0.325	0.488	**0.490**	0.515	0.415	0.441	0.409	0.381	0.412	0.463

51	0.325	0.488	**0.481**	0.481	0.458	0.431	0.398	0.439	0.398	0.428
52	0.325	0.488	**0.500**	0.514	0.478	0.396	0.422	0.468	0.412	0.377
53	0.325	0.488	**0.500**	0.401	0.494	0.412	0.457	0.413	0.455	0.414
54	0.325	0.488	**0.488**	0.555	0.476	0.398	0.420	0.398	0.448	0.466
55	0.325	0.488	**0.494**	0.493	0.454	0.490	0.454	0.469	0.378	0.359
56	0.325	0.488	**0.484**	0.474	0.432	0.461	0.421	0.444	0.429	0.404
57	0.325	0.488	**0.480**	0.443	0.470	0.446	0.426	0.411	0.394	0.421
58	0.325	0.488	**0.490**	0.506	0.454	0.446	0.414	0.394	0.426	0.396
59	0.325	0.488	**0.480**	0.488	0.469	0.490	0.417	0.382	0.371	0.435
60	0.325	0.488	**0.490**	0.411	0.493	0.463	0.395	0.410	0.448	0.439
61	0.325	0.488	**0.488**	0.488	0.488	0.491	0.433	0.380	0.425	0.388
62	0.325	0.488	**0.500**	0.428	0.489	0.418	0.452	0.412	0.380	0.422
63	0.325	0.488	**0.490**	0.410	0.450	0.452	0.429	0.417	0.430	0.390
64	0.325	0.488	**0.480**	0.447	0.479	0.407	0.454	0.431	0.404	0.389
65	0.325	0.488	**0.490**	0.441	0.440	0.486	0.395	0.419	0.385	0.440
66	0.325	0.488	**0.489**	0.448	0.429	0.454	0.432	0.386	0.413	0.403
67	0.325	0.488	**0.500**	0.504	0.481	0.447	0.461	0.456	0.420	0.387
68	0.325	0.488	**0.490**	0.437	0.469	0.449	0.451	0.429	0.407	0.407
69	0.325	0.488	**0.490**	0.425	0.472	0.429	0.425	0.416	0.427	0.417
70	0.325	0.488	**0.491**	0.518	0.514	0.483	0.365	0.392	0.433	0.438
71	0.325	0.488	**0.500**	0.503	0.436	0.411	0.433	0.472	0.429	0.446
72	0.325	0.488	**0.480**	0.512	0.498	0.438	0.415	0.434	0.397	0.402
73	0.325	0.488	**0.478**	0.495	0.453	0.420	0.406	0.449	0.425	0.422
74	0.325	0.488	**0.490**	0.446	0.440	0.390	0.424	0.403	0.367	0.379
75	0.325	0.488	**0.490**	0.452	0.478	0.449	0.431	0.437	0.492	0.413
76	0.325	0.488	**0.480**	0.473	0.438	0.478	0.435	0.424	0.438	0.429
77	0.325	0.488	**0.500**	0.487	0.440	0.433	0.430	0.427	0.409	0.418
78	0.325	0.488	**0.494**	0.476	0.460	0.436	0.435	0.409	0.401	0.422
79	0.325	0.488	**0.500**	0.439	0.521	0.503	0.449	0.414	0.427	0.417
80	0.325	0.488	**0.472**	0.498	0.451	0.435	0.418	0.409	0.394	0.390
81	0.325	0.488	**0.490**	0.482	0.496	0.440	0.414	0.418	0.389	0.474
82	0.325	0.488	**0.487**	0.489	0.495	0.455	0.409	0.426	0.443	0.411
83	0.325	0.488	**0.490**	0.437	0.463	0.480	0.396	0.420	0.409	0.392
84	0.325	0.488	**0.490**	0.481	0.418	0.411	0.473	0.434	0.440	0.471
85	0.325	0.488	**0.480**	0.490	0.464	0.446	0.432	0.416	0.425	0.426
86	0.325	0.488	**0.490**	0.489	0.487	0.381	0.413	0.430	0.398	0.407
87	0.325	0.488	**0.490**	0.501	0.463	0.446	0.477	0.423	0.419	0.401
88	0.325	0.488	**0.488**	0.463	0.452	0.493	0.427	0.388	0.450	0.416
89	0.325	0.488	**0.494**	0.466	0.433	0.438	0.469	0.414	0.448	0.390
90	0.325	0.488	**0.500**	0.502	0.474	0.428	0.431	0.445	0.418	0.397
91	0.325	0.488	**0.484**	0.482	0.515	0.480	0.413	0.417	0.428	0.428
92	0.325	0.488	**0.480**	0.422	0.514	0.450	0.398	0.417	0.431	0.427
93	0.325	0.488	**0.480**	0.444	0.514	0.431	0.446	0.389	0.409	0.408
94	0.325	0.488	**0.500**	0.494	0.492	0.468	0.437	0.437	0.367	0.420
95	0.325	0.488	**0.488**	0.423	0.455	0.527	0.416	0.405	0.420	0.435
96	0.325	0.488	**0.506**	0.458	0.509	0.400	0.367	0.464	0.366	0.427
97	0.325	0.488	**0.490**	0.438	0.487	0.452	0.413	0.406	0.407	0.433
98	0.325	0.488	**0.490**	0.456	0.422	0.459	0.377	0.418	0.440	0.411
99	0.325	0.488	**0.484**	0.500	0.500	0.441	0.402	0.443	0.466	0.452
100	0.325	0.488	**0.488**	0.438	0.504	0.470	0.429	0.423	0.429	0.448
Average	0.325	0.488	**0.487**	0.477	0.473	0.442	0.424	0.421	0.417	0.416
Median	0.325	0.488	**0.490**	0.480	0.473	0.443	0.423	0.421	0.420	0.417

We graphically present the final three topics in Figure 6.1.2. The plot shows the inter-topic differences calculated using the Jensen-Shannon divergence: the less the overlap between three topics, the more the distance between them. As it is possible to see, the three topics appear well distinguished among them.

Figure 6.1.2 – LDA Results: Three Latent Topics Plotted to Evaluate their Differences.

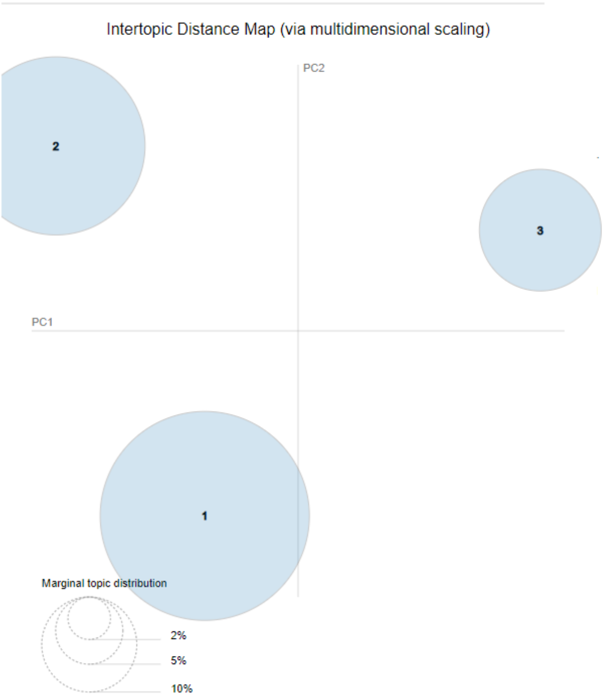


Table 6.1.2 shows each of the three topics identified with the list of its 30 most probable words listed in descending order of relevance, which has been calculated accordingly to Sievert and Shirley (2014):

$$r(w,k|\lambda) = \lambda \log(\phi_{kw}) + (1-\lambda)\log\left(\frac{\phi_{kw}}{p_w}\right) \qquad (1)$$

where w indicates the word, k indicates the topic, ϕ_{kw} denote the probability of term w for topic k, p_w indicates the marginal probability of term w in the corpus, and λ is a balance factor that we have set to 0.5 to give equal weight to the probability of term w for topic k and its lift (Sievert and Shirley 2014). This allows to decrease the rankings of frequent words in the corpus and increase the relevance of rare words.

Table 6.1.2 – Topics Revealed from NLP Analysis.

Topic	Bag of Words (Lemmatized)	Label	% of Documents for Each Topic
1	**update**, **term**, **datum**, **change**, **service**, **user**, make, **collect**, **read**, take, condition, **control**, account, thank, provide, share, team, **understand**, use, effect, question, **right**, include, **transparency**, **easy**, review, cookie, part, full, protect.	**Informative (Transparency)** GDPR novelties: update of terms of service, collection of data, users' control, increase transparency, need to understand and read.	50%
2	**email**, **receive**, **click**, **want**, **keep**, **offer**, preference, send, **continue**, would, news, time, unsubscribe, communication, link, come, **consent**, **stay**, need, **newsletter**, hear, know, **touch**, **event**, like, **marketing**, whish, still, list, late.	**Persuasive** Explicit consent to data usage, users' clicks to access offers, events, newsletter and marketing actions.	41%
3	**datum**, **processing**, process, purpose, **right**, provision, application, request, **transfer**, **administrator**, entity, period, **contract**, particular, claim, necessary, address, **legal**, **complaint**, basis, **obligation**, provide, implementation, object, case, **authority**, conclude, carry, accordance, payment.	**Neither Highly Informative nor Highly Persuasive** GDPR legal material: Privacy by design and default, new consumers' rights, data administrator.	17%

The results above outline the dominant topics characterizing GDPR permission requests. We find that these requests center on three main topics. Interestingly, some of these themes indicate that firms are treating these communications mainly as informative and legal material, whereas we also have a topic that highlights the firms' willingness to persuade consumers to make actions (e.g., click, consent).

Therefore, this unsupervised technique suggests that it is possible to detect two main broad categories of words that companies are using to get users opt-in: *informative* and *persuasive*. Notably, most of the re-permission emails analyzed tried to mainly address the GDPR requirements to communicate in a “*clear and transparent*” way about data collection and data processing procedures and to return *control* of information disclosed to the final users (61%). However, there is also a considerable percentage of communications that were predominantly trying to convince consumers to perform an action (39%). This provided a first evidence that some of the companies took advantage of the re-permission email phenomenon and strategically crafted these communications.

6.2. Stage 2: Analytic Strategy for Uncovering Themes

The second stage of the content analysis procedure involves the creation of an analytical approach which, by integrating theory and automated tools, can be applied to larger datasets and can, for example, be used by policymakers or firms to detect the degree of persuasion and of information in privacy-related communications.

In order to achieve this goal, we used a 20% randomly drawn sub-sample of emails (n=308), and we content-analyzed it by following these steps:

Step 1: We *manually coded* the emails for the six themes identified in the literature: control, transparency, gain/loss framing, time framing, monetary incentives, and non-monetary incentives.

Step 2: We *collected additional text-related variables* on the content of the re-permission emails by using two online software: LIWC and TextEvaluator.

Step 3: We *estimated six regression models*, one for each theme (*Y*), with the LIWC and TextEvaluator variables as covariates (*Xs*), and we *checked the predictive accuracy* of parameters obtained through in- and out-of-sample lift charts.

6.2.1 Step 1: Manual Coding

The first step of this approach is based on the manual content analysis of the re-permission emails collected by using the constructs that literature found relevant in altering consumers’ propensity to grant data access (Chapter 3).

Therefore, we randomly choose 20% of our sample, and we asked two independent judges to code them manually in terms of the construct that literature (Chapter 3.2) suggests being influential in affecting customers' disclosure behavior: Control, Transparency, Framing (Gain/Loss), Monetary and Non-Monetary Incentives and Time Orientation.

We provided the judges with a protocol to content analyze the emails and operationalize the six themes identified into variables. Table 6.2.1.1 summarizes the protocol for the operationalization of the variables we provided to the coders. Examples of how different re-permission emails were coded are available in Appendix C.

Table 6.2.1.1 - Coding Protocol Used to Identify the Presence of Informative and Persuasive Themes in GDPR Re-Permission Emails.

	Communication Themes	Levels	Definition
Informative Themes	**Control**	0 - 1	Coded as 1 if the e-mail highlights and stresses how the user can control personal data. It is coded as 0 if the e-mail provides only basic information about the possibility to control.
	Transparency: Five-level variable assessing the level of transparency of companies in describing their data related activities and data security standards.	1	Minimum level of transparency. The e-mail provides only minimal information to inform the user about how personal data will be processed and used.
		2	Low level of transparency. Information provided to the user is few and generic.
		3	Average level of transparency. Information is provided with references and links to sources to better understand the conditions.
		4	High level of transparency. Information provided is accessible, clear, and easily understandable.
		5	Very High level of transparency. Information provided appears complete, clear, and accessible, with a specific focus on every aspect of the data protection domain.
Persuasive Themes	**Framing: Gain / Losses**	0 - 1	Coded as 1 if the e-mail indicates the presence of gain frame, loss frame, or both in the e-mail.
	Monetary Incentives	0 - 1	Coded as 1 if the e-mail provides a monetary incentive such as discounts or offers in the communication. (Chandon, Wansink, and Laurent 2000)
	Non-Monetary Incentives	0 - 1	Coded as 1 if the e-mail provides a non-monetary incentive such as invitation to events or free trials in the communication. (Chandon, Wansink, and Laurent 2000)
	Framing: Time Orientation	0 - 1	Coded as 1 if the e-mail has some form of time orientation - past, present, or future, and 0 otherwise.

Some of the variables identified above are more objective and some more subjective in terms of the judges' evaluation. For example, the presence of monetary and non-monetary incentives, as well as the type of frame of the emails, can be seen as a more objective type of variable to be coded. In contrast, the presence of control and the degree of transparency can be seen as more subjective variables to be coded. Consequently, we proceed to calculate the inter-judge reliability for the two subjective variables using both the Krippendorff's Alpha and the Cohen's Kappa (Table 6.2.1.2). As it is possible to see, all the values are above the common threshold of 0.8 and fall within the domain of accepted reliability for content analysis. This means that the judges agree on most of the codes assigned at the variables and that the coding is the result of rational and non-casual reasoning based on efficient predefined criteria.

Table 6.2.1.2 – Inter-Judge Reliability Metrics.

Variable	Krippendorff's Alpha	Cohen's Kappa
Control	0.90	0.83
Transparency	0.94	0.82
Average	**0.92**	**0.82**

The main results are presented in the table below (Table 6.2.1.3). In the following sections, we provide a brief analysis of each of the constructs manually coded.

Table 6.2.1.3 – Results of the Manual Coding Procedure (N=308).

Variable		% of Emails
Control		
	Yes	50.65%
	No	49.35%
Transparency		
	1	18.51%
	2	33.77%
	3	20.45%
	4	21.43%
	5	5.84%
Framing: Gain/Loss		
	Yes	37.66%
	No	62.34%
Monetary Incentives		
	Yes	33.12%
	No	66.88%
Non-Monetary Incentives		
	Yes	29.55%
	No	70.45%
Framing: Time Orientation		
	Focus on the Past	21.10%
	Focus on the Future	13.96%

6.2.1.1. Control

The provision of control to users on the data disclosed to the company is one of the main novelties that GDPR has brought about (European Union 2016, Art. 14). According to the regulation, customers can define which data will be collected, how they will be processed, and who can access them. Consequently, companies are, to different degrees, incorporating this opportunity in their email communications, highlighting how users can decide and play around with data consent.

By looking at our sample (Table 6.2.1.3), we have a very striking result: half of the emails did not talk about data control (49%). This is likely not to be the results that the GDPR Regulator wanted to achieve by forcing companies to actively communicate about customers' data privacy and protection.

By inspecting the remaining 51% of the re-permission emails, they seem to be more coherent with the GDPR principles, providing either general information about data management and control or more specific and complete details about how consumers can practically control the data disclosed.

6.2.1.2. Transparency

Another relevant variable is the one that regards the level of transparency provided by the company on its data practices. The GDPR establishes that companies should inform clearly and understandably about how data are processed (European Union 2016, Art. 12), and customers should provide informed consent to the companies' privacy terms. By a first visual inspection of the emails, it is possible to notice that companies present information about privacy in many different ways. It is possible to have cases that range from emails in which a clear and extensive explanation about data protection and customers' rights is provided until emails in which companies describe in a very simplistic and concise way the rights and principles of the new regulation, giving references to external websites for additional insights.

By looking at our sample (Table 6.2.1.3), we have a second striking result: 53% of the emails have a very low level of transparency: 19% of the emails do not talk about transparency, while 34% contain very few and general information about privacy. Of the remaining half, roughly 41% of the emails are specific and accurate in talking about data protection – supplementing information with links, infographics, and images – and 6% provide high-quality and complete information about data protection.

6.2.1.3. Framing: Gain/Loss

As highlighted by the literature on privacy communication (Chapter 3.2), companies can try to alter customers' privacy concerns by using subtle marketing tools and strategies (Acquisti, Brandimarte, and Loewenstein 2015). One of the tools that companies may exploit is by framing the phrases of their communications to highlight the gains that customers can derive from data disclosure or the losses that they can experience from data denial. This has been shown to alter customer disclosure behavior.

We consequently have kept track of the different framing present in our emails, and we found (Table 6.2.1.3) that we have more emails framed as gains (31.7%) than framed as losses (17.2%). This is consistent with studies in advertising that mainly adopt a gain-type of framing to evocate positive feelings and memories in the customers. Additionally, this is also consistent with research in the health sector, according to which, for behaviors focused on preventing some negative outcomes, a gain-type of framing is more effective in prompting users to act accordingly to the behavior.

6.2.1.4. Incentives: Monetary & Non-Monetary

Similar to the framing stimulus, also incentives may alter the valuation of customers for their data. As shown in literature, people are afraid of giving away information, but they actually do disclose personal information when in the presence of incentives (Athey, Catalini, and Tucker 2017) because the benefits outweigh the costs of data disclosure. Therefore, people's privacy preferences are malleable and can be modified by using some monetary or non-monetary incentives in exchange for data.

If we look at the number presented in Table 6.2.1.3, we can see that companies are using incentives to get customers' consent. It seems to be a pretty common practice among the companies in our sample since 41.1% of them are using at least one of the two incentives, and 21.4% are using both of them in the same communication. Moreover, it is also possible to see that monetary incentives (33.1%) are more frequent than non-monetary incentives (29.6%).

6.2.1.5. Framing: Time Orientation

Literature in marketing has discussed the role of time and time frames on customers' behavioral decisions for a long time. Companies may use time orientations to alter customers' perceptions by playing on the feelings evoked by events and relations described. For example, the use of past orientation has been shown to be linked to feelings of nostalgia, which has been extensively studied in marketing, given its effects on the decision to buy products and services (Havlena and Holak 1991). Additionally, companies may also decide to use future framing to attract and tempt customers in prompting a specific behavior. For example, in the case of the firm's decision to craft a privacy communication, it may be both crafted around past feelings and events – to prompt customers to remember to "*good old times*" and to confirm, once again, the trust in the company – or future possibilities – to entice consumers to opt-in in order to "*wait and see*".

Consequently, we decided to record the presence of referrals to past and future events in the re-permission emails to see whether firms are actually using this type of lever to achieve customers' data access. By inspecting Table 6.2.1.3, we found that firms use, to a small extent, time framings in their privacy communications (28.9%); in particular, they tend to focus more on the history (21.1%) with the customers than on the future (13.9%), leveraging feelings of nostalgia.

6.1.2. Step 2: Collection of Additional Text-Related Variables

In order to be able to find an approach that can be easily used to screen large datasets of emails, we needed to collect text-related variables that can be automatically retrieved. Therefore, we used two online software to get additional metrics about the texts' complexity and the psychological constructs used in the emails: the former set of metrics was obtained through TextEvaluator (ETS), while the latter through LIWC. We present a description of the main variables obtained by these two online programs in the following sub-chapters.

6.1.2.1 Linguistic Inquiry and Word Count (LIWC)

The Linguistic Inquiry and Word Count (LIWC) is a text analysis program able to extract information about how thoughts, feelings, personality, and motivations are present in a text. It is based on pre-loaded dictionaries that identify the percentage of different constructs in a text. As described on the website, "*the dictionary identifies which words are associated with which psychologically-*

relevant categories. After the processing module has read and accounted for all words in a given text, it calculates the percentage of total words that match each of the dictionary categories”[12].

LIWC has been widely used in academic marketing research (Tang and Guo 2015; Tausczik and Pennebaker 2010), and it calculates, for a wide-range list of meaningful categories, the percentage of the counts of the words falling into a specific category. The possible metrics are approximately 90 and can be grouped in the following categories:

- Summary language variables
- General descriptor categories
- Standard linguistic dimensions
- Word categories tapping psychological constructs
- Personal concern categories
- Informal language markers
- Punctuation categories

As highlighted above, all the metrics produced are measured as a percentage of text. Consequently, if, for example, one of the communications scores “*25*” on the category “Risk”, it means that 25% of the total words used in that particular text may be categorized as risk-related words.

Table 6.1.2.1.1 reports the main LIWC variables that we used later in our analyses, as well as some descriptive statistics. By looking at it, we can say that the types of privacy communications companies are bringing about due to GDPR requirements are 200 words long and contain a text organized logically and formally (Analytic = 70.1%). This is probably because of the legal type of text they should contain to comply with the GDPR. Additionally, these communications prompt mainly positive emotions and are more focused on the present rather than on the past or the future,

[12] https://liwc.wpengine.com/how-it-works/

which is in line with what we found through the manual coding analysis. Looking at the statistics for rewards, money, and risk, these are present only in a tiny proportion in the texts.

Table 6.1.2.1.1 - Definition and Descriptive Statistics of the LIWC Metrics (N = 1506).

Variable	Definition	Range	Mean	Std. Dev.	Min	Max
WC	Words count.	-	199.65	156.56	12	2622
Analytic	Percentage of text that contains words that suggest formal, logical, and hierarchical thinking patterns.	0-100	70.61	19.84	1	99
Clout	Percentage of text that relates to the relative social status, confidence, or leadership that people display through their writing or talking.	0-100	97.48	4.13	44.48	99
Authentic	Percentage of text that contains words that suggest a more personal, humble, and vulnerable way of writing.	0-100	23.2	17.37	1	99
Tone	This is a single summary variable that puts together negative emotion and positive emotion. Values below 50 indicate a negative emotional tone.	0-100	79.56	19.77	1.37	99
WPS	Words per Sentence.	-	20.51	6.79	4	102
Function	Percentage of text that contains function words.	0-100	48.18	4.84	24	66.67
Posemo	Percentage of text that can be classified as pertaining to positive emotions.	0-100	3.9	1.95	0	25
Negemo	Percentage of text that can be classified as pertaining to negative emotions.	0-100	0.29	0.54	0	5.13
Social	Percentage of text that can be classified as pertaining to social processes.	0-100	17.42	4.31	2.44	36.62
Cogproc	Percentage of text that can be classified as pertaining cognitive processes.	0-100	12.1	3.54	0	24.49

Percept	Percentage of text that can be classified as pertaining to perceptual processes.	0-100	1.21	1.28	0	10.53
Affiliation	Percentage of text that can be classified as pertaining to the need for affiliation.	0-100	7.48	2.91	0	22
Achieve	Percentage of text that can be classified as pertaining to the need for achievement.	0-100	1.29	1.27	0	25
Power	Percentage of text that can be classified as pertaining to the need for power.	0-100	4.28	1.84	0	14.29
Reward	Percentage of text that can be classified as reward-related.	0-100	1.08	1.13	0	25
Risk	Percentage of text that can be classified as risk-related.	0-100	1.24	0.96	0	6.45
Relativ	Percentage of text that can be classified as pertaining to relativity.	0-100	12.16	3.54	3.28	33.33
Work	Percentage of text that can be classified as work-related.	0-100	5.31	2.47	0	20.51
Leisure	Percentage of text that can be classified as leisure related.	0-100	0.86	0.98	0	8.33
Home	Percentage of text that can be classified as home related.	0-100	0.15	0.41	0	6.45
Money	Percentage of text that can be classified as money-related.	0-100	1.05	1.23	0	10.71
Death	Percentage of text that can be classified as death related.	0-100	0.01	0.15	0	3.28
Informal	Percentage of text that can be classified as informal.	0-100	0.5	0.98	0	23.08
Focus on the Past	Percentage of text that can be seen as having a focus on the past.	0-100	1.52	1.29	0	8.86
Focus on the Present	Percentage of text that can be seen as having a focus on the present.	0-100	9.69	3.14	1.33	30.77

Focus on the Future	Percentage of text that can be seen as having a focus on the future.	0-100	1.92	1.34	0	8.33

Additionally, we also inspected whether communications sent by firms in different continents are different regarding the type of communication used by the company. In order to statistically test for the existence of differences in the means of the levels of the LIWC variables listed in Table 6.1.2.1.1, we conducted a series of multiple comparison tests – with the Tukey's correction – by using the MultiComparison command available in Python. The main result obtained regards the difference between Europe and North America and is presented in Table 6.1.2.1.2. As it is possible to see from Table 6.1.2.1.2, European communications, compared to North Americans' ones, are using more negative emotional words, are focused more on the future than on the past, and contain more risk-related words. This is likely due to the fact that European companies perceived more heavily the enforcement of GDPR and its principles.

Table 6.1.2.1.2 - Multiple Comparison of Means for LIWC Variables (Tukey's Correction, α = 0.05).

Negative Emotion						
Group 1	**Group 2**	**Mean Diff.**	**P-Adj**	**Lower**	**Upper**	**Reject**
Africa	Asia	-0.21	0.90	-1.74	1.32	False
Africa	Europe	-0.14	0.90	-1.67	1.38	False
Africa	N. America	-0.28	0.90	-1.80	1.25	False
Africa	Oceania	0.08	0.90	-1.49	1.65	False
Africa	S. America	-0.47	0.90	-2.62	1.68	False
Asia	Europe	0.07	0.90	-0.13	0.27	False
Asia	N. America	-0.06	0.90	-0.27	0.14	False
Asia	Oceania	0.30	0.39	-0.14	0.73	False
Asia	S. America	-0.26	0.90	-1.79	1.27	False
Europe	**N. America**	**-0.13**	**0.00**	**-0.22**	**-0.04**	**True**
Europe	Oceania	0.23	0.56	-0.17	0.62	False
Europe	S. America	-0.33	0.90	-1.85	1.20	False
N. America	Oceania	0.36	0.11	-0.04	0.76	False
N. America	S. America	-0.19	0.90	-1.72	1.33	False
Oceania	S. America	-0.55	0.90	-2.12	1.02	False

Risk

Group 1	Group 2	Mean Diff.	P-Adj	Lower	Upper	Reject
Africa	Asia	1.02	0.90	-1.75	3.79	False
Africa	Europe	0.81	0.90	-1.94	3.56	False
Africa	N. America	0.62	0.90	-2.13	3.37	False
Africa	Oceania	0.94	0.90	-1.90	3.78	False
Africa	S. America	0.91	0.90	-2.97	4.79	False
Asia	Europe	-0.21	0.56	-0.57	0.15	False
Asia	**N. America**	**-0.40**	**0.03**	**-0.77**	**-0.03**	**True**
Asia	Oceania	-0.08	0.90	-0.87	0.71	False
Asia	S. America	-0.11	0.90	-2.88	2.66	False
Europe	**N. America**	**-0.19**	**0.01**	**-0.35**	**-0.03**	**True**
Europe	Oceania	0.13	0.90	-0.59	0.84	False
Europe	S. America	0.10	0.90	-2.65	2.85	False
N. America	Oceania	0.32	0.78	-0.40	1.04	False
N. America	S. America	0.29	0.90	-2.46	3.04	False
Oceania	S. America	-0.03	0.90	-2.87	2.81	False

Focus on Past

Group 1	Group 2	Mean Diff.	P-Adj	Lower	Upper	Reject
Africa	Asia	0.41	0.90	-3.23	4.06	False
Africa	Europe	0.46	0.90	-3.16	4.08	False
Africa	N. America	0.83	0.90	-2.79	4.45	False
Africa	Oceania	1.10	0.90	-2.64	4.83	False
Africa	S. America	0.43	0.90	-4.68	5.54	False
Asia	Europe	0.05	0.90	-0.42	0.52	False
Asia	N. America	0.41	0.16	-0.08	0.91	False
Asia	Oceania	0.69	0.42	-0.36	1.73	False
Asia	S. America	0.02	0.90	-3.63	3.66	False
Europe	**N. America**	**0.37**	**0.00**	**0.15**	**0.58**	**True**
Europe	Oceania	0.64	0.39	-0.30	1.58	False
Europe	S. America	-0.03	0.90	-3.65	3.59	False
N. America	Oceania	0.27	0.90	-0.68	1.22	False
N. America	S. America	-0.40	0.90	-4.02	3.22	False
Oceania	S. America	-0.67	0.90	-4.40	3.07	False

Focus on Future

Group 1	**Group 2**	**Mean Diff.**	**P-Adj**	**Lower**	**Upper**	**Reject**
Africa	Asia	-1.68	0.78	-5.45	2.09	False
Africa	Europe	-1.79	0.72	-5.54	1.95	False
Africa	N. America	-2.14	0.57	-5.89	1.60	False
Africa	Oceania	-2.01	0.65	-5.88	1.85	False
Africa	S. America	-3.10	0.54	-8.39	2.19	False
Asia	Europe	-0.11	0.90	-0.60	0.38	False
Asia	N. America	-0.46	0.10	-0.97	0.04	False
Asia	Oceania	-0.33	0.90	-1.41	0.74	False
Asia	S. America	-1.42	0.89	-5.19	2.35	False
Europe	**N. America**	**-0.35**	**0.00**	**-0.57**	**-0.13**	**True**
Europe	Oceania	-0.22	0.90	-1.19	0.76	False
Europe	S. America	-1.31	0.90	-5.05	2.44	False
N. America	Oceania	0.13	0.90	-0.85	1.12	False
N. America	S. America	-0.96	0.90	-4.70	2.79	False
Oceania	S. America	-1.09	0.90	-4.95	2.78	False

6.1.2.2 TextEvaluator

TextEvaluator is an online tool by ETS that returns different variables that give an indication of the level of complexity of a text. ETS is an organization that aims at "*advancing quality and equity in education by providing fair and valid assessments, research and related services*"[13]; its most known tests are the GRE, the TOEFL, and the TOEIC, which all aim at evaluating different sets of students' skills and knowledge. To help teachers, instructors, and researchers, the same organization has also developed a tool that can be used to establish the complexity of a text to be used in instruction and assessment.

As described in the TextEvaluator manual (TextEvaluator 2017), the output variables can be grouped into four macro-categories, indicating different cognitive processes. Additionally, a final overall variable is generated, signaling the global level of complexity of the text. In Table 6.1.2.2.1,

[13] https://www.ets.org/mission

we reported the descriptions and the summary statistics of all the variables and constructs made available by TextEvaluator for the re-permission email collected in our sample.

Table 6.1.2.2.1 - Definition and Descriptive Statistics of the TextEvaluator Variables (N = 1506).

Type	**Variable**	**Definition**	**Range**	**Mean**	**Std. Dev.**	**Min**	**Max**
Understanding Words	Academic Vocabulary	Extent to which the language of a text is more characteristic of academic texts than of nonacademic texts.	0-100	73.24	11.7	2	100
	Word Unfamiliarity	Extent to which the language comprises unusual words.	0-100	76.17	9.75	4	100
	Concreteness	Extent to which the text contains more concrete words and is more likely to evoke meaningful mental images.	0-100	22.71	9.24	1	80
Understanding Sentences	Syntactic Complexity	It is composed of phrase-level elements such as the paragraph's length or the number of dependent clauses per sentence.	0-100	51.82	12.87	1	100
Inferring Connections Across Sentences	Lexical Cohesion	Extent to which a text can be considered as a "coherent message" rather than a collection of unrelated clauses and sentences.	0-100	47.96	7.1	28	100
	Level of Argumentation	Extent to which it is easy or difficult to infer connections across sentences.	0-100	24.13	18.25	7	93

Using knowledge of discourse organization to generate additional inferences	Degree of Narrativity	Extent to which a text contains past tense verbs and third-person singular pronouns, which are characteristic of narrative texts.	0-100	57.39	8.11	4	100
	Interactive/ Conversational Style	Extent to which a text exhibits an interactive/conversational style.	0-100	39.72	22.67	5	100
Overall Text Complexity Scores		It provides a single overall measure of text complexity for a text.	100-2000	1007.7	194.34	50	2000

We can see that privacy communications tend to use a more complex type of vocabulary, using both rare worlds and a more academic type of language. They are also a type of communication that is more abstract, less argumentative, and less conversational in nature. Overall, our re-permission emails have a complexity score of 1000, which has been shown to corresponds to a Common Core Grade Level equal or greater than 9 (Table 6.1.2.2.2). This means that, on average, these communications are difficult to read and can be understood by people with a higher level of education, highlighting the inherent complexity of privacy-related communications and partially explaining the reasons why people tend not to read them (Milne and Culnan 2004).

Table 6.1.2.2.2 – TextEvaluator to Common Core Concordance Conversion Table (Sheehan et al. 2014).

Common Core Grade Level	**TextEvaluator Score Range (100-200 Scale)**
2	100-525
3	310-590
4	405-655
5	480-720
6	550-790
7	615-860
8	685-940
9	750-1025
10	820-1125
11	890-1245
12	970-1360

Additionally, as previously done with LIWC, we statistically tested for the existence of differences in the means of the levels of the TextEvaluator variables listed in Table 6.1.2.2.1 We conducted a series of multiple comparison tests – with the Tukey's correction – by using the MultiComparison command available in Python. The main result obtained regards the difference between Europe and North America and is presented in Table 6.1.2.2.3. As it is possible to see from Table 6.1.2.2.3, North American communications, compared to Europeans' ones, are more narrative and cohesive in the language used.

Table 6.1.2.2.3 - Multiple Comparison of Means for TextEvaluator Variables (Tukey's Correction, α = 0.05).

Degree of Narrativity						
Group 1	**Group 2**	**Mean Diff.**	**P-Adj**	**Lower**	**Upper**	**Reject**
Africa	Asia	1.32	0.90	-23.03	25.68	False
Africa	Europe	-0.10	0.90	-24.28	24.07	False
Africa	N. America	1.56	0.90	-22.63	25.75	False
Africa	Oceania	0.87	0.90	-24.09	25.82	False
Africa	S. America	3.00	0.90	-31.17	37.17	False
Asia	Europe	-1.42	0.77	-4.59	1.74	False
Asia	N. America	0.24	0.90	-3.05	3.53	False
Asia	Oceania	-0.46	0.90	-7.41	6.50	False
Asia	S. America	1.68	0.90	-22.68	26.03	False
Europe	**N. America**	**1.66**	**0.01**	**0.25**	**3.08**	**True**
Europe	Oceania	0.97	0.90	-5.32	7.25	False
Europe	S. America	3.10	0.90	-21.07	27.28	False
N. America	Oceania	-0.70	0.90	-7.05	5.65	False
N. America	S. America	1.44	0.90	-22.75	25.63	False
Oceania	S. America	2.13	0.90	-22.82	27.09	False

Lexical Cohesion						
Group 1	**Group 2**	**Mean Diff.**	**P-Adj**	**Lower**	**Upper**	**Reject**
Africa	Asia	11.77	0.62	-9.98	33.53	False
Africa	Europe	9.40	0.79	-12.19	30.99	False
Africa	N. America	11.14	0.66	-10.47	32.75	False
Africa	Oceania	9.73	0.79	-12.55	32.02	False
Africa	S. America	17.00	0.59	-13.52	47.52	False
Asia	Europe	-2.38	0.16	-5.20	0.45	False
Asia	N. America	-0.63	0.90	-3.57	2.30	False
Asia	Oceania	-2.04	0.90	-8.25	4.17	False
Asia	S. America	5.23	0.90	-16.53	26.98	False
Europe	**N. America**	**1.74**	**0.00**	**0.48**	**3.01**	**True**
Europe	Oceania	0.34	0.90	-5.28	5.95	False
Europe	S. America	7.60	0.90	-13.99	29.19	False
N. America	Oceania	-1.41	0.90	-7.08	4.27	False
N. America	S. America	5.86	0.90	-15.75	27.47	False
Oceania	S. America	7.27	0.90	-15.02	29.55	False

6.1.3. Step 3: Model Estimation and Validation

As previously described, the two online text analysis tools described in Section 6.1.2 return, for each communication, the degree to which various categories of words are used in a text. We contend that the LIWC words' categories and the text complexity indicators can detect the six themes identified in previous literature and predict the likelihood that a specific theme characterizes a re-permission email. Therefore, in the third step, we tried to relate the variables manually coded by the judges with the variables that the two online software provided us with. This was done by using regression models that included, as covariates, the words' categories identified through LIWC and TextEvaluator and as dependent variables the manually coded variables. We estimated the models on 75% of the manually coded observations (N = 232), and we tested them on the remaining 25% (N = 76).

We proceeded as described below. For clarity, we present the approach used to estimate the likelihood of the degree of *transparency* of the emails, but this is generalizable to all the other five themes manually coded by the judges.

Firstly, we looked at the operationalization of the "*transparency*" variable obtained through the manual coding at stage two and used it as a dependent variable in our model. Secondly, we included, as covariates, the words' categories identified through LIWC and TextEvaluator to estimate a model predicting the degree of *transparency* of a re-permission email *j*. Thirdly, we tested the predictive accuracy of the models through a lift chart analysis.

Table 6.1.3.1 and Figure 6.1.3.1 present the results of the estimation procedure for two of the six manually coded variables (see Appendix D for complete results). By looking at the more relevant variables in each model estimated (Table 6.1.3.1), we find evidence that some of the constructs identified by LIWC and TextEvaluator are relevant in predicting the manually coded theme. For example, the presence of "*monetary incentives*" is predominantly predicted by higher values of the LIWC "*money*", while the level of "*transparency*" by the length of the text, according to our expectations. Additionally, Figure 6.1.3.1 shows that the models have good predictive power both in- and out-of-sample, signaling that we can use the estimated models to predict the presence of the theory-based themes into larger datasets of privacy-related texts.

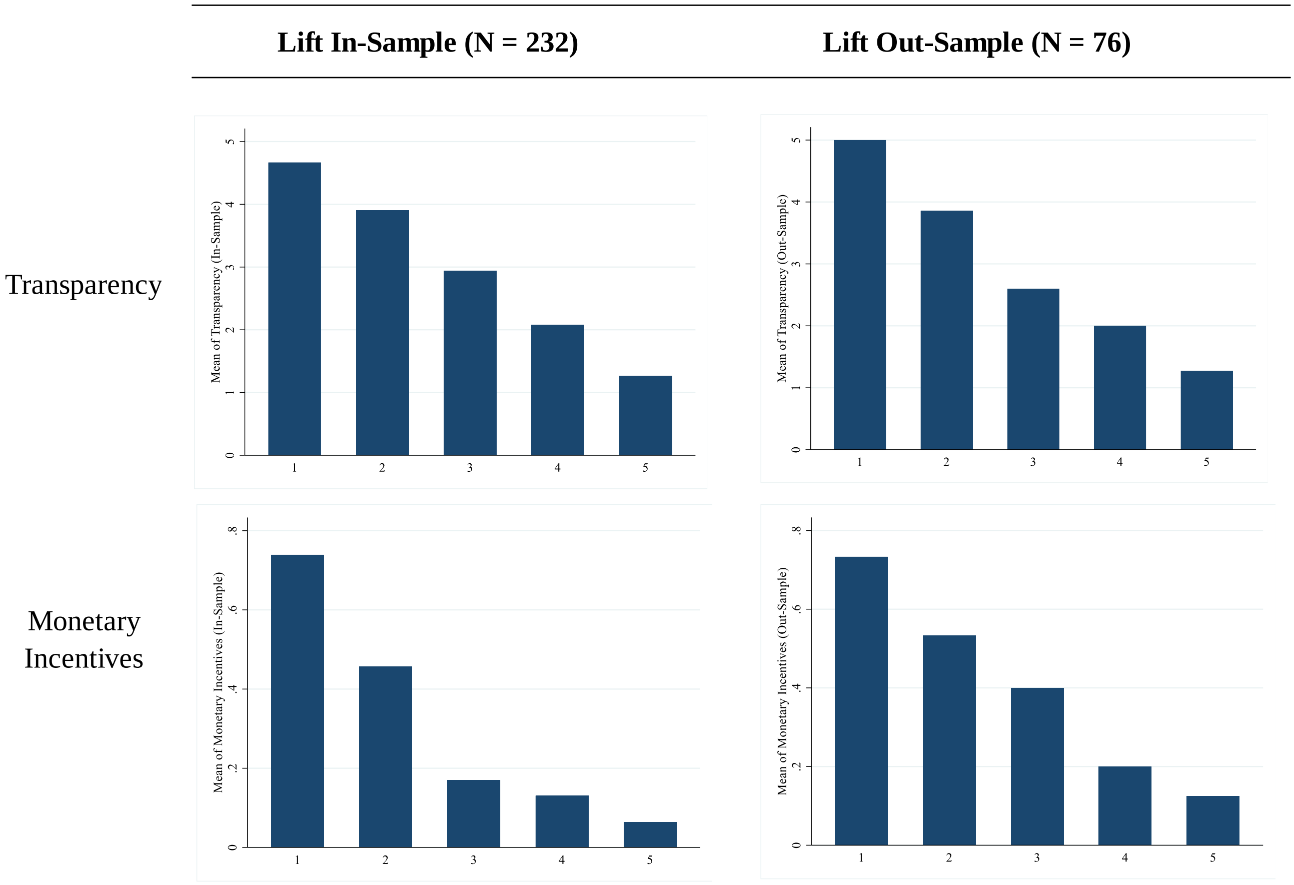
Table 6.1.3.1 – Results of the Predictive Models – DVs = Transparency; Monetary Incentives (N = 308)

		Trasparency		**Monetary Incentives**	
LIWC	wc	0.037 ***	(0.003)	0.000	(0.001)
	analytic	-0.017 *	(0.01)	-0.018	(0.012)
	clout	0.118 **	(0.048)	0.060	(0.056)
	authentic	0.014	(0.014)	0.012	(0.015)
	tone	-0.023 *	(0.014)	-0.037 **	(0.016)
	wps	0.024	(0.018)	0.044 **	(0.022)
	function	-0.076 *	(0.046)	0.022	(0.052)
	posemo	0.029	(0.138)	0.263	(0.165)
	negemo	-0.214	(0.295)	-0.749 **	(0.336)
	social	-0.096 **	(0.048)	0.097 *	(0.053)
	cogproc	0.117 **	(0.051)	-0.066	(0.058)
	percept	-0.080	(0.128)	0.165	(0.148)
	affiliation	0.169 ***	(0.063)	-0.092	(0.073)
	achieve	-0.048	(0.146)	0.154	(0.167)
	power	-0.020	(0.085)	0.104	(0.099)
	reward	0.391 **	(0.166)	0.330	(0.205)
	risk	-0.099	(0.169)	0.518 **	(0.206)
	relativ	-0.100	(0.078)	0.107	(0.09)
	work	0.019	(0.07)	-0.127	(0.091)
	leisure	0.055	(0.152)	0.226	(0.174)
	home	0.496	(0.321)	0.547	(0.36)
	money	-0.070	(0.113)	0.460 ***	(0.14)
	death	-0.062	(2.859)	0.034	(3.968)
	informal	-0.110	(0.175)	0.027	(0.202)
Text Evaluator	Academic Vocabulary	0.073	(0.055)	-0.131	(0.094)
	Concreteness	-0.014	(0.064)	0.145	(0.107)
	Degree of Narrativity	-0.017	(0.02)	-0.058 **	(0.023)
	Interactive/Conversational	0.000	(0.021)	0.020	(0.034)
	Level of Argumentation	-0.011	(0.018)	-0.045	(0.03)
	Lexical Cohesion	-0.010	(0.025)	0.017	(0.034)
	Syntactic Complexity	-0.064	(0.079)	-0.127	(0.134)
	Word Uunfamiliarity	-0.049	(0.061)	-0.048	(0.018)
	Complexity Score	0.004	(0.011)	0.018	(0.102)

Model	Ordered Logit	Logit
# Obs.	308	308
Log-Likelihood	-234	-142
Pseudo R^2	0.49	0.27

* p < 0.1; ** p < 0.05; *** p < 0.01

Figure 6.1.3.1 – Lift Chart Analysis (In- and Out-of-Sample)



Lift In-Sample (N = 232)

Lift Out-Sample (N = 76)

Transparency

Monetary Incentives

6.3. Stage 3: Prediction of the Themes for the Whole Sample of Emails

Thanks to the procedure identified in Stage 2 (Section 6.2), we were able to develop an efficient and theoretically sound approach for content-analyze privacy-related communications that integrates both human and automated interventions. This system can be used to evaluate larger datasets of privacy-related communications.

Therefore, we used the parameter estimates to predict the presence of the manually coded variables in the whole sample of emails (N = 1506). This estimation procedure allowed us to assess the degree of presence of the main constructs found in the literature for all the re-permission emails collected. The estimated variables are the ones we are going to use for our analyses later, given the satisfactory results obtained in the second step.

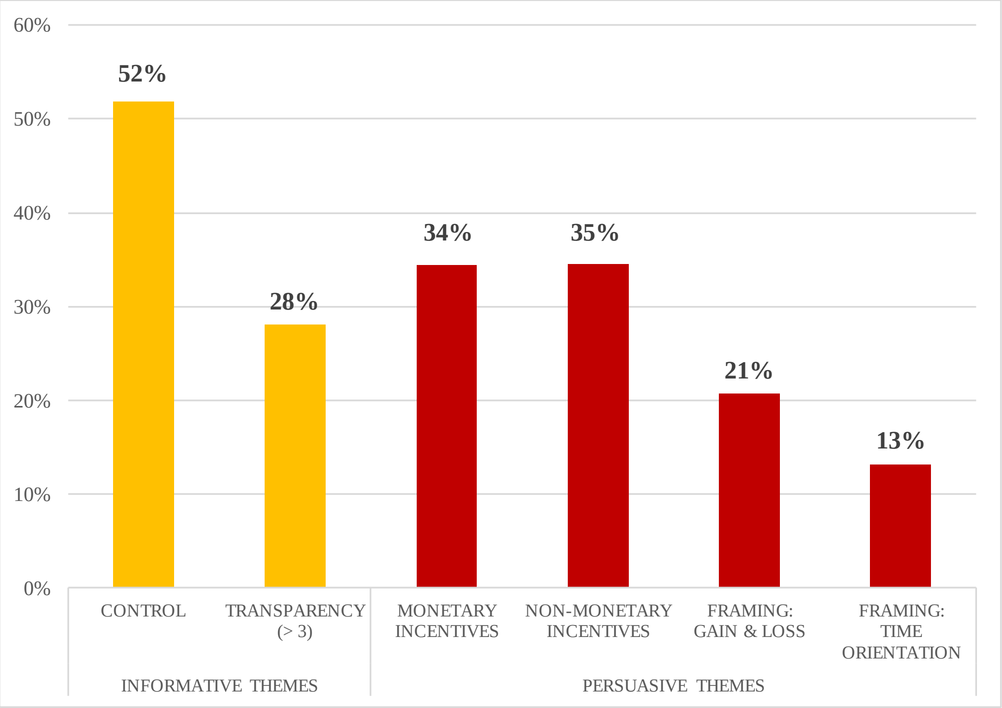
In order to be able to interpret them directly and in an easier way in our subsequent models and to assess the differences in the frequency of emails showing the different themes, we decided to perform some transformations on the estimated variables. In particular, for variables such as "*monetary incentives*", "*non-monetary incentives*", "*gain and loss framing*" and "*control*", which were estimated using logistic models, we dichotomized the probability associated with each variable by using the maximum value obtained for the Jouden-Index of a specific variable and assigning the value to 1 if the estimated probability was greater than the index found, and 0 otherwise (Lehmann, Gupta, and Steckel 1998, p. 663). Figure D.3 (Appendix D) shows the Youden-Index values chosen as cutoffs to dichotomize the variables estimated. For the variable "*transparency*", which was estimated using an ordered logit model, we assigned the new categorical variable to the category with the highest estimated probability.

Figure 6.3.3.1 and Figure 6.3.3.2 outline the presence of the main themes associated with our sample of emails.

As it is possible to see in Figure 6.3.3.1, results show that firms extensively highlight the possibility of controlling personal information in their messages. *Control* is the topic most frequently stressed (52%) in re-permission emails. Additionally, 28% of emails are conceived to be perceived as highly *transparent*. This is not only in line with the spirit of the reform that explicitly calls for communications "*concise, transparent, intelligible and easily accessible form, using clear and plain language*" (European Union 2016, Art. 12), but also stress the extra effort that companies have made to be perceived as trustable and "*clean*" by their users. Interestingly, however, also persuasive themes have been extensively used. Particularly, *incentives* (both monetary and non-monetary) that were used in 34% of re-permission emails, as well as *framing* in terms of gain and/or losses (21%) and in terms of future or past time orientation (13%).

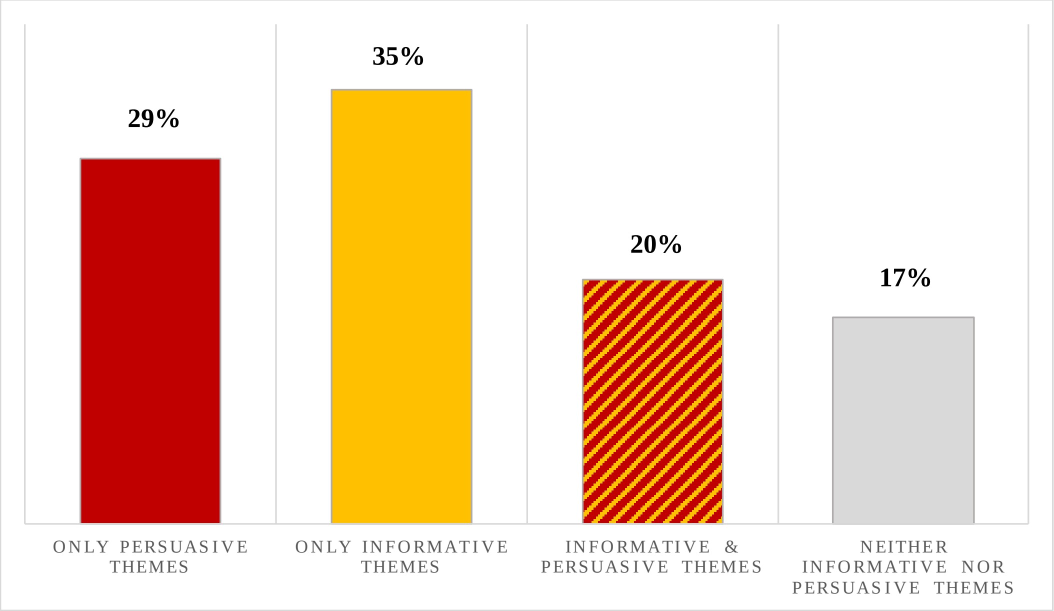
Additionally, by looking at Figure 6.3.3.2, we can inspect the likelihood that specific categories of themes are used in combination. Notably, in line with Figure 6.3.3.1, most of the re-permission emails collected (35%) highlight the GDPR principles of *control* and *transparency* and did not include any *persuasive* themes, which is in line with what the Regulator wanted to achieve through these particular type of privacy-related communications. However, the histogram also signals that a significant portion of emails (29%) used only *persuasive themes*. This means that companies took advantage of the re-permission emails to prompt customers to disclose their data by exploiting other communicative elements that ease consumers' actions. Interestingly, there is also a sizable portion of emails that were crafted as hybrid solutions combining *control* or *transparency* with *incentives* or *framing*.

Figure 6.3.3.1 – Average Likelihood that a Specific Theme is Used in Re-Permission Emails.



Note
Transparency is measured using a five-level scale variable. The bar represent the likelihood of values above 3.

Figure 6.3.3.2 – Average Likelihood that Combinations of Themes' Categories are Used in Re-Permission Emails.



Note
Transparency is measured using a five-level scale variable. The bar represent the likelihood of values above 3.

Lastly, Table 6.3.3.1 present a comparison between the results obtained through the models we estimated and the ones obtained by the manual coding procedure. As it is possible to see, the percentages of emails containing the different topics are not too distant from one another, suggesting a good fit of the estimation to the manual coding procedure.

Table 6.3.3.1 – Comparison between the Frequencies of the Presence of the "Themes": Manual Coding vs. Estimation.

	Description	Manual Coded Variables	Estimated Variables	Delta
Control				
	Yes	50.65%	51.86%	1%
	No	49.35%	48.14%	
Transparency				
	1	18.51%	11.75%	7%
	2	33.77%	36.12%	2%
	3	20.45%	24.10%	4%
	4	21.43%	20.78%	1%
	5	5.84%	7.24%	1%
Framing				
	Yes	37.66%	20.72%	17%
	No	62.34%	79.28%	
Monetary Incentives				
	Yes	33.12%	34.40%	1%
	No	66.88%	65.60%	
Non-Monetary Incentives				
	Yes	29.55%	34.60%	5%
	No	70.45%	65.40%	

6.4. Validation Between Data-Driven & Theory-Based Techniques

As previously highlighted, we used two different methodologies to content-analyze the re-permission emails that companies sent out on the occasion of the GDPR enforcement.

Firstly, we used a data-driven approach. This was done by implementing an unsupervised content analysis technique (LDA modeling) which can uncover the latent topics present in a predefined collection of texts. Thanks to this approach, we were able to identify three main topics that characterize our sample of re-permission emails. We defined two of the topics as more *informative* in nature – since they contained more GDPR related words (e.g., control, transparency, processing, update) – while the last one can be seen as more persuasive and more related to the strategies' companies implemented to prompt consumers' consent behavior (e.g., click, consent, offers, marketing).

Secondly, we employed a theory-based approach. We turned to literature to detect the constructs that previous works suggest being influential in prompting consumers' disclosure behaviors, identifying six main themes: control, transparency, incentives (monetary and non-monetary), and framing (either in terms of gain and loss and in terms of time orientation). Then, we created a coding protocol and asked two independent judges to code the emails accordingly. Notably, the coders also highlight three other recurring elements that characterize the sample of re-permission emails. Table 6.4.1 describes these additional variables.

Table 6.4.1 – Operationalization of the Additional "Themes" Identified by the Coders.

Additional Coded Themes	Levels	Definition
Security	0-1	Coded as 1 if the email contains referrals to the companies' data security practices and protection standards (e.g., "we guarantee the security of your data") and 0 otherwise.
Care	0-1	Coded as 1 if the email contains referrals to the importance of consumers' privacy for the company (e.g., "We care about your privacy") and 0 otherwise.
Clarity	0-1	Coded as 1 if the email describes in simple and understandable way the novelties introduced by the GDPR (e.g., by using iconographic or bullet-points) and 0 otherwise.

The use of the first method allowed us to get an overall sense of the content of these emails and provides a first empirical evidence that companies coupled the GDPR requirement - to inform and make aware customers about data collection and procedures – with other communicative elements. However, this approach has left us with little in-depth information about the variety of arguments used by companies in the re-permission emails. Thanks to the second procedure, instead, we were able to get more granular insights about the type of elements that companies used in their privacy-related communications and to classify them according to their final aim: while transparency and control were mainly used to *inform* customers about the GDPR novelties, the use of incentives and the framing of the text inherently have the aim to alter consumers' behaviors and *persuade* them to take action.

The availability of results from two different content analysis methods allows us to cross-validate them one against the other. Therefore, we modeled the three themes identified by the LDA on the manually coded variables to see whether there are differences in the way in which manually coded themes load on LDA topics. Table 6.4.2 presents the results of the three models.

As it is possible to see, we found empirical evidence that the LDA topics are related to the manually coded themes in different ways. The first and second topics can be seen as complementing each other in terms of significant *themes*. While the latter is positively related to more transparent and clearer communications and negatively related to the presence of incentives or the text's framing in terms of gains and losses, the former is the opposite. Interestingly, the third topic is not related to the themes identified by the literature about privacy-related communications. It is positively related to communications that stress the relevance of data security and negatively associated with the presence of customers' privacy care statements highlighting a type of communication that is more technical and focused on data processing and data storage procedures.

Therefore, thanks to this analysis, we provide additional evidence that themes can be grouped into macro-categories: marketing, framing, and care are mainly associated with the second topic; transparency and clarity are characteristics of the first topic; security is the central theme characterizing the third topic.

Notably, in Chapter 6.1, we labeled the first and the third topic as more "*informative*" in nature, while the second one as more "*persuasive*", only by looking at the most relevant words per topic. Additionally, in Chapter 4, we also grouped the six theory-based themes into the "*persuasive*" and "*informative*" macro-categories only by distinguishing among the main goal that the specific themes were trying to achieve (e.g., increase knowledge vs. altering behaviors). Therefore, by combining two different approaches, the findings of this analysis provide additional support for the categorization of the six theory-based themes into the *persuasive* and *informative* macro classes.

Table 6.4.2 – Results from the Logit models – DV = Topics identified by the LDA; IV = Manual Coded variables (N = 308).

		Topic 1 Informative (Transparency)			**Topic 2 Persuasive**			**Topic 3 Neither Highly Informative nor Highly Persuasive**		
		Coef.	Std. Err.	Sig.	Coef.	Std. Err.	Sig.	Coef.	Std. Err.	Sig.
Manual Coded Variables	Control	-0.118	0.243		0.001	0.237		0.020	0.224	
	Transparency	0.842	0.239	***	-0.960	0.236	***	0.124	0.213	
	Clarity	0.884	0.443	**	-0.624	0.434		0.592	0.497	
	Security	-0.039	0.394		-0.033	0.386		0.816	0.384	**
	Marketing [1]	-0.925	0.378	**	1.118	0.375	***	-0.568	0.435	
	Framing [2]	-2.504	0.418	***	2.171	0.410	***	0.237	0.448	
	Care	-0.345	0.400		0.663	0.401	*	-0.992	0.387	**
	Constant	-0.928	0.502	*	1.234	0.496	**	-2.007	0.547	***

Note:
*** p < 0.01; ** p < 0.05; * p < 0.1
[1] We grouped the presence of any monetary and non-monetary incentives into the variable "Marketing".
[2] We grouped the presence of a gain or loss type of framing into the variable "Framing".

6.5. Summary of the Results

As shown in Table 6.5.1, we used two content analysis approaches – data-driven and theory-based – to analyze our sample of emails. While the use of the first method (e.g., LDA) allowed us to get an overall sense of the content of these emails – and provides first empirical evidence that companies coupled the informative nature of the GDPR re-permission emails with persuasive communicative elements – the second procedure (e.g., manual content analysis) provides more in-depth information about the type of elements that companies used in their privacy-related communications and to classify them according to their final aim. The availability of results from two different content analysis methods allowed us to cross-validate them one against the other and to show that we were able to reach consistency in highlighting the peculiarities of the content of re-permission emails. Additionally, we also executed another corroboration check by performing a factor analysis on the theory-based constructs (estimated through our three-step procedure), and we found that two latent factors characterize our emails: one more informative and one more persuasive in nature (see Appendix E for details).

For simplicity, in Table 6.5.2, we present the results regarding the content of the GDPR re-permission emails collected to get an immediate sense of how firms have crafted and designed these particular and delicate types of communication.

The content analysis showed that many of the re-permission emails collected were designed as merely informative tools. However, there is also another considerable percentage of them that was centered on persuasive cues only. Indeed, a substantial number of companies inserted either type of incentives (monetary and/or non-monetary) and used a particular framing of the text to enhance the probability of getting access to customers' data. Interestingly, we also found that many of these emails were designed as *hybrid* solutions, suggesting that firms tried to make an effort to balance the informative nature that these specific types of communications should have with the degree of persuasion they need to reach their purposes.

Table 6.5.1 – Summary of the Two Procedures and Cross-Validation Results.

			Cross-Validation	
Method	**Type of Content Analysis**	**Results of the Methods**	**LDA**	**Manual Coding**
Latent Dirichlet Allocation (LDA)	Data-Driven & Unsupervised	Three Topics: • Informative • Persuasive • Neither Highly Informative Nor Highly Persuasive	Informative	Transparency Control
Manual Content Analysis	Theory-Based & Supervised	Six Main Themes: • Transparency • Control • Monetary Incentives • Non-Monetary Incentives • Framing: Gain and Loss • Framing: Time Orientation	Persuasive	Monetary Incentives Non-Monetary Incentives Framing (Gain/Loss) Framing (Time Orientation)

Table 6.5.2 – Summary of the Main Results obtained from the Content Analysis.

	Variable	Description	Main Descriptive Results
			Our sample is composed by communications:
Informative	Control	Binary variable indicating if companies are providing details with regards to users' control over their data.	That do talk about users' *control* on their data.
	Transparency	Ordered categorical variable with five levels determining the degre to which companies have been transparent in the description of data practices and data security procedures in their re-permission emails.	That have a medium level of *transparency* about data protection and privacy procedures.
Persuasive	Framing: Gain/Loss	Binary variable assuming value 1 if the company has used some kind of framing in the design of the communication - either in terms of gain or losses.	That present some kind of *framing* - more frequently highlighting the gains rather than losses.
	Monetary Incentives	Binary variable assuming value 1 if the company has used monetary incentives in the re-permission email and 0 otherwise.	That use both *monetary* and *non-monetary* incentives
	Non-Monetary Incentives	Binary variable assuming value 1 if the company has used non-monetary incentives in the re-permission email and 0 otherwise.	
	Framing: Time Orientation	Continuous variable (0-100) giving the percentage of text that show some kind of time orientation (either past, present or future focused).	With low levels of *past* and *future* orientation in the text.

7. Firms' Self Interest & Re-Permission Emails' Content

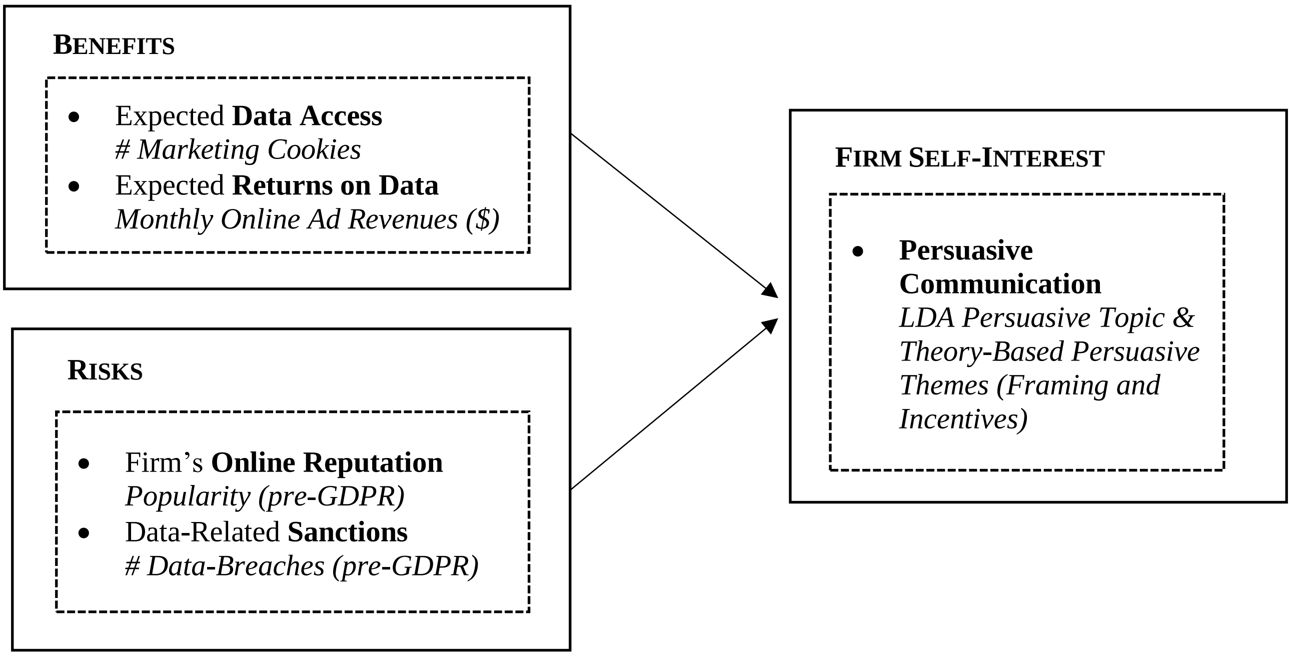
In this chapter, we are interested in addressing the second research question of this thesis, which tries to understand whether and how the themes used in the GDPR re-permission emails – identified in Chapter 6 – are driven by firms' self-interest.

As we theorized in our conceptual framework (Chapter 4), we are particularly interested in understanding whether firms designed their re-permission emails opportunistically by carefully evaluating the benefits and risks associated with their data collection. Consequently, we aim at answering questions such as the following one: "*How self-interested were firms? Did the benefits (from the data usage) as well as the risks (of non-compliance) drive request content and intent?*"

In other words, we tried to study if firms' online strategies – which are based on data collection and exploitation – as well as their financial and reputational costs from not adhering to the scope of the regulation, somehow dictate how companies decide to strategically communicate with their consumers by turning their data request from merely informative to persuasive.

To verify this, we collected information about the benefits (e.g., data harvesting and monetization strategies) and the risks (e.g., online reputation and sanctions) firms may incur when handling data. We then modeled the degree of persuasion of the re-permission emails' content – measured in terms of the LDA persuasive topic and the theory-based persuasive themes – on the firms' number of marketing cookies, expected ad revenue, past data-breached experienced, and online popularity to test for the presence of a connection between the design of the data request and the interests of the company crafting it. Figure 7.1.1 shows how the main theoretical constructs used in our conceptual framework (Chapter 4) map on the variables we have collected (Chapter 5).

Figure 7.1.1 – Conceptual Framework Translated into a Model



7.1. Summary of the Main Dependent Variables

For convenience, in this paragraph, we summarize the main metrics we will use as dependent variables for the models described in this chapter.

In our conceptual framework (described in Chapter 4), we argued that firms might have used privacy-related communication strategically depending on the tradeoff they face between the benefits and the risks associated with the data collection and exploitation. Moreover, we also contended that firms' self-interest might have influenced their data request content, shifting it from a merely informative type of communication to a more persuasive one. Therefore, as it is possible to see in Figure 7.1.1, we measured firms' self-interest by evaluating the degree of persuasion of their privacy-related communications (e.g., GDPR re-permission emails).

As described in Chapter 6, we evaluate the content of the re-permission emails collected through two main alternative methods:

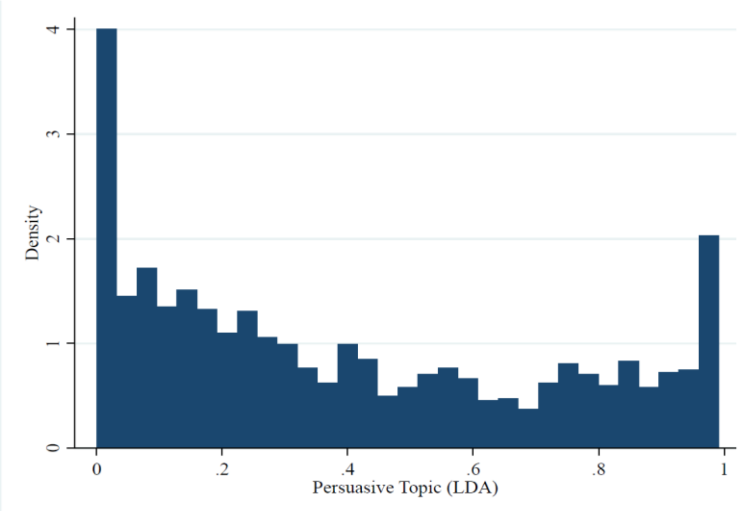
(i) *Data-Based Unsupervised Technique* (e.g., LDA Model) which have allowed us to detect three main latent topics without imposing any pre-defined structure on the data collected.

(ii) *Theory-Based Approach* (e.g., Protocol-Based Content Analysis) which have allowed us to get more granular insights on the content of this specific type of privacy-related communications by using the constructs that privacy literature highlights as relevant in driving users' disclosure behavior.

Following, we report the distribution plots and the summary statistics for the main variables that the above-cited approaches have been identified as signaling a persuasive type of content and that will then be used in the models described in the next sub-chapters:

- **Data-Driven**: Persuasive LDA Topic

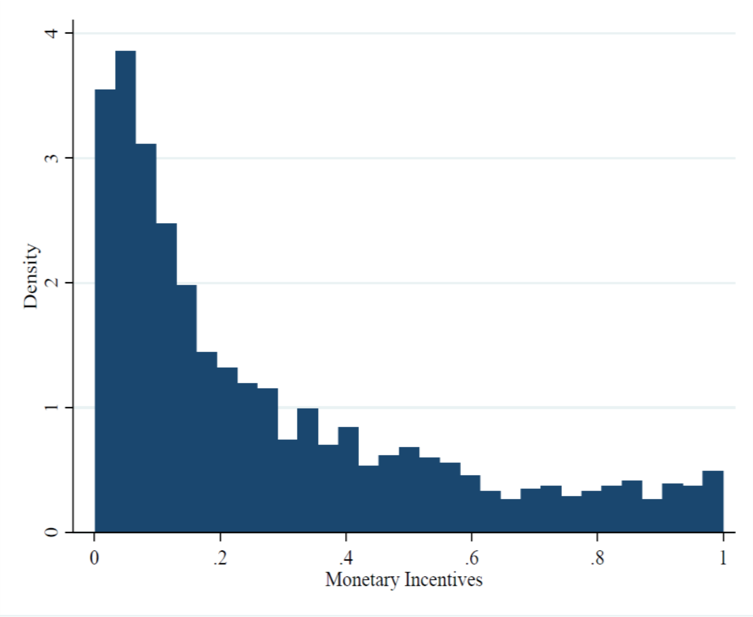
This variable represents the degree of persuasion included in the re-permission emails by evaluating the text's words. It is essential to highlight that we identified three latent topics from the LDA procedures. We were able to assign three different probabilities – which correspond to the three latent topics identified – to the emails' text we have collected, which, taken together, sum to one. Indeed, the LDA method mainly aims to detect the hidden constructs that describe a set of documents and maximize the differentiation between the discovered topics (see Chapter 6.1 for details).



Obs.	=	1,506
Average	=	0.400
SD	=	0.323
Min	=	0.000
Median	=	0.312
Max	=	0.989
Presence of Zeros	=	75
Presence of Ones	=	0

- **Theory-Based**: Monetary Incentives

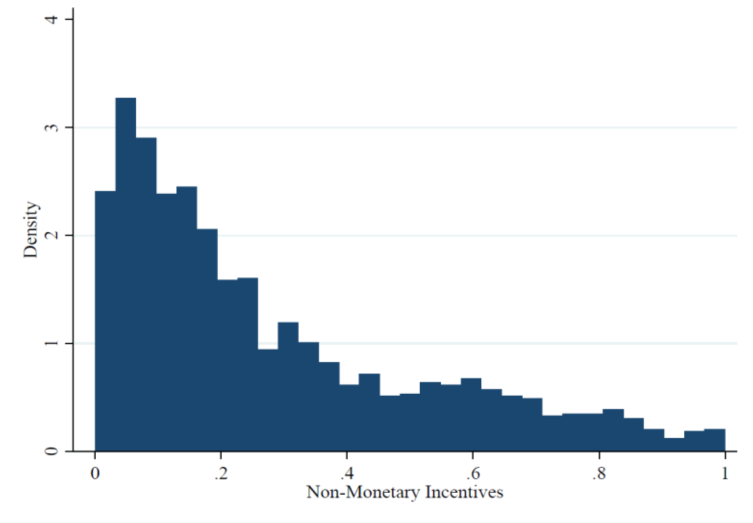
This variable represents the probability that the GDPR re-permission email includes monetary incentives such as discounts or coupons (see Appendix C for examples).



Obs.	=	1,506
Average	=	0.278
SD	=	0.266
Min	=	0.001
Median	=	0.175
Max	=	1.000
Presence of 0	=	0
Presence of 1	=	1

- **Theory-Based**: Non-Monetary Incentives

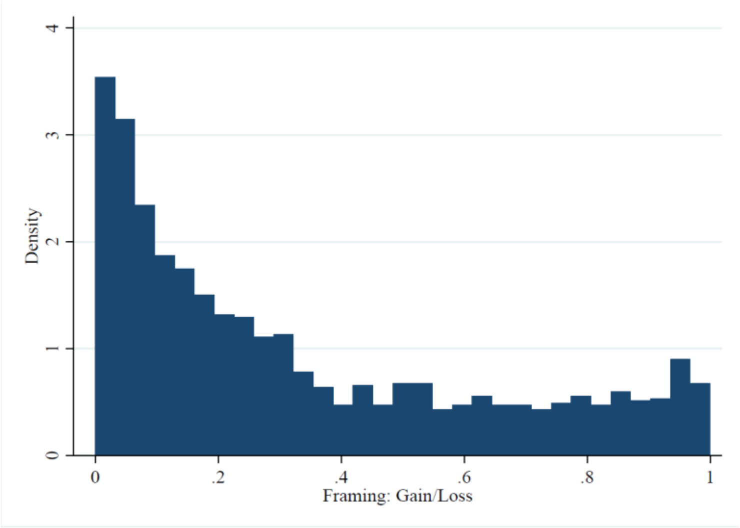
This variable represents the probability that the GDPR re-permission email includes non-monetary incentives such as participation in events or lotteries (see Appendix C for examples).



Obs.	=	1,506
Average	=	0.280
SD	=	0.243
Min	=	0.000
Median	=	0.194
Max	=	1.000
Presence of 0	=	0
Presence of 1	=	0

- **Theory-Based**: Framing: Gain/Loss

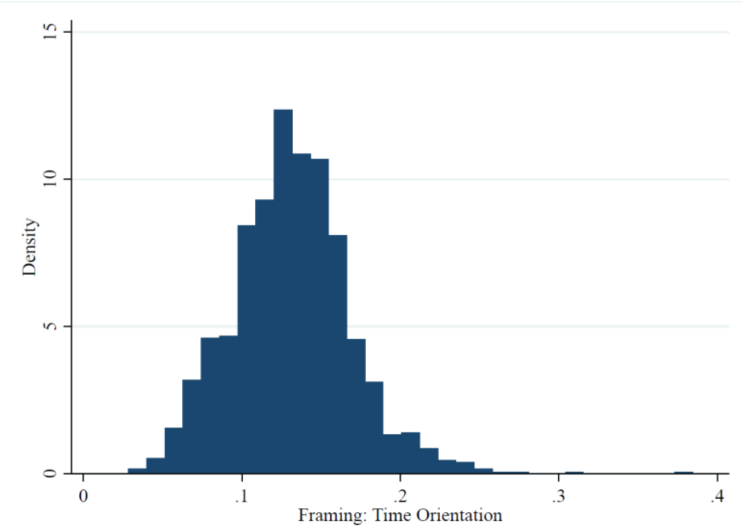
This variable represents the probability that the text of the GDPR re-permission email is framed in terms of gain from data disclosure and/or losses from data denial.



Obs.	=	1,506
Average	=	0.332
SD	=	0.300
Min	=	0.000
Median	=	0.226
Max	=	1.000
Presence of 0	=	0
Presence of 1	=	0

- **Theory-Based**: Framing: Time Orientation

This variable represents the percentage of words contained in the GDPR re-permission email that has a specific time orientation both in terms of past (e.g., past tenses or referral to past events) and in terms of future (e.g., future tenses or referral to future events).



Obs.	=	1,506
Average	=	0.13127
SD	=	0.03736
Min	=	0.0284
Median	=	0.13035
Max	=	0.3846
Presence of 0	=	0
Presence of 1	=	0

The availability of two sets of results – which are convergent and robust (see Chapter 6) – is crucial for our analysis since it allows us to reach two main objectives: (i) get valuable insights from the different levels of granularity of the variables gained from the content analysis, and (ii) check the robustness of our results to changes in our dependent variables.

7.2. Model Specification

All of our dependent variables are *probabilities* since they have been estimated either through unsupervised latent models (e.g., Persuasive LDA Topic) or through predictive models grounded on a theory-based procedure that has been used to scale larger datasets (e.g., Incentives and Framing). Therefore, as highlighted in previous studies (Buis 2010; Loch, Boxall, and Wheeler 2016), the use of linear regression is not suitable because of its four main assumptions:

(i) OLS is supposed to be used when the model's dependent variable assumes values on the whole Real line, and it predicts values on the entire Real line.

(ii) OLS requires the relation between the dependent and independent variables to be linear.

(iii) OLS requires the normality of the residuals.

(iv) OLS requires homoscedasticity.

In contrast, when dealing with proportions and probabilities, we have that:

(i) The dependent variable is bounded between zero and one.

(ii) The effect of the independent variables on the dependent variable tends not to be linear.

(iii) Residuals are not normally distributed.

(iv) Data are usually heteroskedastic, meaning that the dependent variable's variance tends to increase around the mean and decrease when reaching the boundaries.

Therefore, we had to find other models' specifications to be applied to our data. One of the most widespread models used in literature when researchers deal with proportions is the Beta-regression since it is a very flexible type of model based on two parameters that allow the distribution to have a wide variety of shapes (e.g., Buckley 2003; Ferrari and Cribari-Neto 2004; Hardin and Hilbe 2014;

Mebane 2000; Paolino 2001; Smithson and Verkuilen 2006). One of the main limitations of this model is that the dependent variable should be strictly greater than zero and smaller than one, meaning that the interval's extremes should be excluded from the support of the variable to be able to use the Beta-regression. However, as it is possible to see from the distribution plots and summary statistics of our dependent variables, we have cases in which zeros and ones are, instead included. Since it is reasonably frequent that, in real situations, proportions assume zeros and ones (Baum 2008), other models' specifications have been proposed to take into account also the extremes of the interval.

One first option, which has been employed in various studies, is to employ the zero-one inflated Beta-regression (e.g., Cook, Kieschnick, and McCullough 2008; Loch, Boxall, and Wheeler 2016; Ospina and Ferrari 2010, 2011; Williams 2019). This model specification is used when the proportion of zeros or ones is not negligible, assuming a mixed continuous-discrete distribution: the Beta-regression is used to model the continuous part, whereas the Bernoulli distribution takes care of the discrete component. However, also this model specification is not entirely suitable in our situation since the proportion of zeros and ones needs to be "*considerable*" to be modeled as distinct processes – which is not our case.

Therefore, we opted for another model specification that has been proposed when the dependent variable is bounded between zero and one and when zeros and ones are also present in the dataset: the Fractional Response Model proposed by Papke and Wooldridge (1996). This type of model has been employed in different settings (e.g., Adegbesan and Higgins 2011; Buis 2010; Gallani, Krishnan, and Wooldridge 2015; Williams 2019) and can be considered as a standard approach when handling proportion type of data (Adegbesan and Higgins 2011). It is an extension of the Generalized Linear Model with a nonlinear functional form (e.g., the logistic link function). As defined by the authors, the model's assumption is as follows:

$$E(y_i|\boldsymbol{x_i}) = G(\boldsymbol{x_i\beta})\ \ \forall i \tag{1}$$

where $G(\cdot)$ is a known function satisfying $0 < G(z) < 1$ for all $z \in \mathbb{R}$, such as the logistic function of this form $G(z) \equiv \Lambda\,(z) \equiv \frac{\exp\,(z)}{1+\exp\,(z)}$.

The model is then estimated through a quasi-likelihood method by maximizing a Bernoulli log-likelihood function of this form:

$$l_i(b) \equiv y_i \log[G(\boldsymbol{x_i\beta})] + (1 - y_i) \log\,[1 - G(\boldsymbol{x_i\beta})] \tag{2}$$

Therefore, we define our models as in Equation (3):

$$\begin{aligned} E(Persuasive_i|\boldsymbol{x_i}) &= G(\boldsymbol{x_i\beta}) \\ &= \frac{\exp(\boldsymbol{x_i\beta})}{1+\exp(\boldsymbol{x_i\beta})} \\ &= \frac{\exp\left(\alpha+\sum_{k=1}^{K}\beta_k Benefits_{ki}+\sum_{j=1}^{J}\gamma_j Risks_{ji}+\sum_{m=1}^{M}\delta_m OtherContent_{mi}+\sum_{l=1}^{L}\tau_l Controls_{li}\right)}{1+\exp\left(\alpha+\sum_{k=1}^{K}\beta_k Benefits_{ki}+\sum_{j=1}^{J}\gamma_j Risks_{ji}+\sum_{m=1}^{M}\delta_m OtherContent_{mi}+\sum_{l=1}^{L}\tau_l Controls_{li}\right)} \end{aligned} \tag{3}$$

where $Persuasive_i$ represents the percentage of persuasiveness for the GDPR re-permission email of company i. This is measured through two set of variables: (i) *LDA Persuasive Topic* – which is the probability that the specific re-permission email of company i contains predominantly persuasive type of words – and (ii) *Theory-Based Persuasive Themes* – which are the probabilities that the specific re-permission email of company i contains each of the four main themes literature suggest being influential in prompting disclosure. Chapter 7.3 deals with the result of the model in which the dependent variable is the "*LDA Persuasive Topic*". Chapter 7.4 presents the results of the models estimated on the four different theory-based themes: "*Monetary Incentives*", "*Non-Monetary Incentives*", "*Framing: Gain/Loss*", "*Framing: Time Orientation*".

Benefits$_{ki}$ represents the K variables related to the benefits that companies can achieve from using persuasion in their communications, which are data access and returns on data – measured, respectively, as "# Marketing cookies" and "Expected Monthly Online Ad Revenue" for company i. *Risks$_{ji}$* represents the J variables dealing with the risks that companies can experience from not being completely compliant with the data protection law, which may be possible losses in reputation or financial sanctions – measured, respectively, as "Online Popularity" and "# Data-Breaches" experienced prior the GDPR enforcement by company i. Consequently, β_k and γ_k are the key parameters of interest, helping in evaluating the delicate tradeoff that companies faced when deciding whether to insert persuasive cues inside their GDPR re-permission emails.

OtherContent$_{mi}$ takes into consideration the additional type of content included in the re-permission email of company i. This variable will comprehend different measures depending on the type of content analysis considered.

When dealing with LDA results, *OtherContent* will take into account the probability that the specific re-permission email of company i contains a predominantly informative type of words – e.g., *OtherContent = LDA Informative Topic*. The inclusion of this variable will allow controlling for the effect of the presence of informative content in the re-permission email of company i, and isolate the characteristics of the companies that mainly used communications with a highly persuasive type of content.

In contrast, when dealing with results from the theory-based content analysis, *OtherContent* represents the additional M percentages of themes that the re-permission emails contain according to our estimation procedure. Therefore, *OtherContent* will take into consideration both (i) the probabilities that the specific re-permission email of company i contains the two main themes related to the main GDPR principles (e.g., Transparency and Control) and (ii) the probabilities that the specific re-permission email of company i contains the additional three persuasive themes that may be included into GDPR re-permission emails (e.g., Monetary Incentives, Non-Monetary Incentives, Framing, Time Orientation). In other words, *OtherContent* will take into account both the informative nature and the additional part of persuasiveness of the communication not included in the dependent variable considered.

Controls$_{li}$ represents the set of L control variables specific for company i (e.g., size of the firm, industry, country, age, type of online business – publisher vs. advertiser) and τ_l represents its set of parameters. α indicate the intercept of the model.

We estimated the model by using the *glm* command in STATA in combination with link(logit), *family(binomial)*, and *vce(robust)* as suggested by literature (Baum 2008; Buis 2010; McDowell and Cox 2004).

Additionally, in Appendices F and G, we provided robustness checks to changes in the measurement of the variable "# Marketing Cookies". As highlighted in Chapter 5, we get a double measure for this variable from Cookiebot. Therefore, we tested whether the choice of the Cookiebot extraction – used to get the two measurements for the variable "# Marketing Cookies" –may shift the model's results dramatically. As it is possible to see in Appendices F and G, this is not our case, given the negligible differences achieved in the estimation parameters.

7.3. Results: LDA Persuasive Topic over Firms' Self-Interest

As previously said, we estimated the firm's GDPR re-permission email content by using two alternative approaches. In this sub-chapter, we are going to focus our attention on the results from the data-driven methodology implemented, the LDA modeling approach. Thanks to this approach, we were able to assign a degree of persuasion (in the form of probability) to each of the re-permission emails in our sample by considering all the persuasive elements present in the firm's textual communications.

Therefore, as described in the previous section, we regressed the percentage of text that we identified as "*Persuasive*" – by analyzing the LDA results – on both the benefits and the risks firms had to evaluate when designing their data requests. Moreover, we also inserted, in the models, a covariate that allows controlling for the percentage of text which is mainly informative (by adding the "LDA Informative Theme" as a covariate) in order to be able to identify and isolate the characteristics of the companies which decided to be predominantly persuasive. This will allow us to understand which companies behaved opportunistically and turned to persuasion when deciding about the content of GDPR re-permission emails. Column 6 of Table 7.3.1 summarizes the parameter estimates of the complete model (robustness checks are provided in Appendix F).

By focusing our attention on the ***benefits*** only – data collection and monetization – the results indicate that companies are more likely to be extremely persuasive when they know to have the potential to extract value out of the data collection (β_{2_Model6} =0.023). Additionally, the findings also highlight that it is the value of the data *per se* and not the intensity of the data collection that makes firms more self-interested since the number of marketing cookies is not associated with more intense use of persuasive elements in their re-permission emails.

Concerning the ***risks*** only – reputation and sanction – the most interesting result is that it does seem that companies are not particularly worried about their reputation and popularity when communicating about their data practices. In contrast, it seems that the number of data-breaches experienced before the GDPR enforcement is relevant in driving re-permission emails' content, since the more the data-security exposures the company faced in the past, the less the likelihood that the same company decided to turn to persuasion and craft its data request opportunistically (γ_{2_Model6} = -0.390).

Another compelling result regards the ***control variables*** inserted in the model. The country and the industry of the company are significantly associated with the inclusion of just persuasive cues in its re-permission email. Firstly, firms based in Europe are more careful in using high levels of persuasion, trying to be more compliant with the GDPR law (τ_{1_Model6}= -0.266). This is probably an effect of the introduction of the new GDPR reform, acting more firmly on the European companies and suggesting that companies based in the EU are less likely to use persuasion to a great degree in their GDPR re-permission emails. Secondly, the model also suggests that firms operating in the "*Travel, Tourism and Hospitality*" sector are more likely to use persuasion to a great degree (τ_{8_Model6}= 0.212), while firms operating in the "*Media and Entertainment*" sector are less likely to include persuasive elements in their data related communications (τ_{4_Model6}= -0.361).

A detailed discussion of the results presented is provided in Chapter 7.5.

Table 7.3.1 – Fractional Logit Models' Results - DV = LDA Persuasive Topics (Equation(3)).

		DV = LDA Persuasive Topic											
		(1)		(2)		(3)		(4)		(5)		(6)	
Benefits	# Marketing Cookies (1)			0.001	(0.001)							0.000	(0.001)
	ln(Expected Monthly Online Ad Revenue) [1]					0.023**	(0.011)					0.023**	(0.012)
Risks	Website Popularity (3 Months pre-GDPR)							-0.373**	(0.156)			-0.390**	(0.154)
	# Data-Breaches (pre-GDPR)									-0.000	(0.000)	-0.000	(0.000)
Email's Content	LDA Informative Topic	-4.496***	(0.111)	-4.497***	(0.112)	-4.566***	(0.118)	-4.485***	(0.112)	-4.507***	(0.112)	-4.559***	(0.117)
Controls	EU	-0.276***	(0.050)	-0.274***	(0.050)	-0.246***	(0.053)	-0.296***	(0.050)	-0.273***	(0.050)	-0.266***	(0.052)
	Firm Size	-0.004	(0.014)	-0.005	(0.014)	-0.015	(0.016)	0.002	(0.014)	-0.006	(0.014)	-0.010	(0.016)
	Firm Age	-0.001	(0.001)	-0.001	(0.002)	-0.000	(0.002)	-0.001	(0.002)	-0.001	(0.001)	-0.000	(0.002)
	Sectors:												
	Media and Entertainment	-0.293**	(0.122)	-0.298**	(0.122)	-0.378***	(0.131)	-0.276**	(0.123)	-0.296**	(0.122)	-0.361***	(0.131)
	Professional Services	0.071	(0.085)	0.068	(0.085)	0.061	(0.090)	0.081	(0.085)	0.068	(0.085)	0.071	(0.090)
	Retail Trade	0.086	(0.092)	0.083	(0.092)	0.085	(0.096)	0.093	(0.092)	0.087	(0.092)	0.093	(0.095)
	Software and IT Services	0.062	(0.104)	0.060	(0.104)	0.063	(0.107)	0.070	(0.104)	0.059	(0.104)	0.071	(0.107)
	Travel, Tourism and Hospitality	0.246**	(0.102)	0.243**	(0.101)	0.189*	(0.110)	0.268***	(0.101)	0.245**	(0.101)	0.212*	(0.110)
	Advertiser (0/1)=1 [3]	-0.006	(0.098)	0.011	(0.099)	0.002	(0.106)	-0.006	(0.098)	0.000	(0.098)	0.011	(0.107)
	Country Missing	-0.603***	(0.219)	-0.601***	(0.219)	-0.424*	(0.246)	-0.618***	(0.219)	-0.601***	(0.220)	-0.435*	(0.244)
	Firm Size Missing	-0.304**	(0.140)	-0.302**	(0.140)	-0.405**	(0.172)	-0.292**	(0.140)	-0.306**	(0.140)	-0.394**	(0.172)
	Firm Age Missing	0.263**	(0.115)	0.265**	(0.115)	0.324**	(0.135)	0.258**	(0.115)	0.264**	(0.116)	0.318**	(0.135)
	Advertiser Missing	-0.004	(0.107)	0.020	(0.111)	0.056	(0.128)	-0.004	(0.107)	0.008	(0.108)	0.072	(0.130)
	Constant	1.828***		1.796***		1.617***		1.819***		1.847***		1.609***	
	Observations	1506		1506		1364 [2]		1506		1506		1364 [2]	
	Log-Pseudolikelihood	-556.37		-556.32		-499.35		-555.94		-556.29		-498.86	
	AIC	1142.74		1144.64		1030.71		1143.88		1144.59		1035.72	
	BIC	1222.50		1229.72		1114.20		1228.96		1229.66		1134.86	

Notes:

Standard errors in parentheses

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

[1] We add a small constant term to get around instances of zeroes in this variable as we take logs (Pattabhiramaiah, Sriram and Manchanda 2019).

[2] The sample size has been reduced because there are companies for which the expected online ad revenue figure was not available.

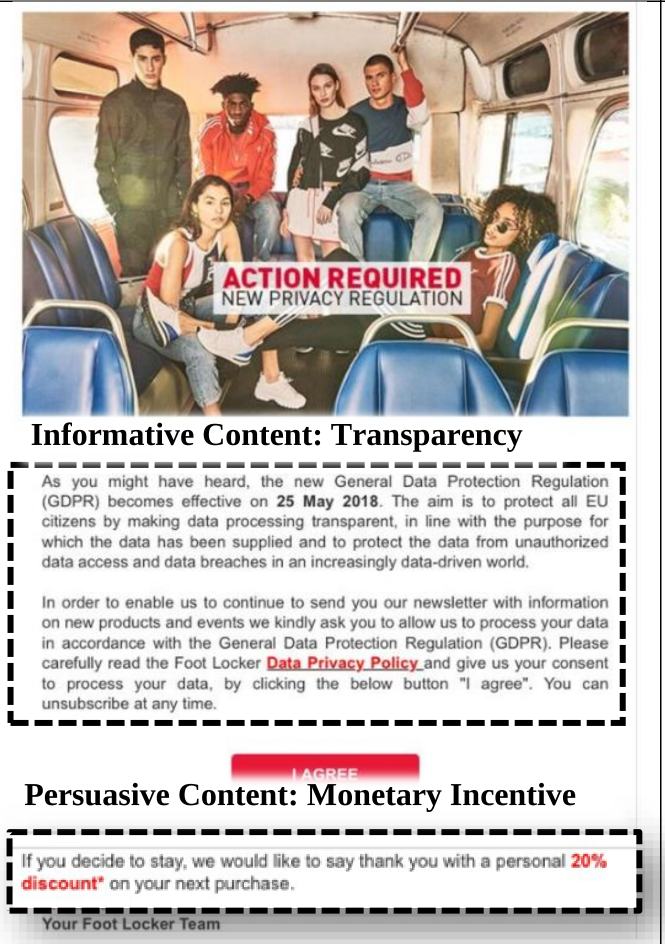
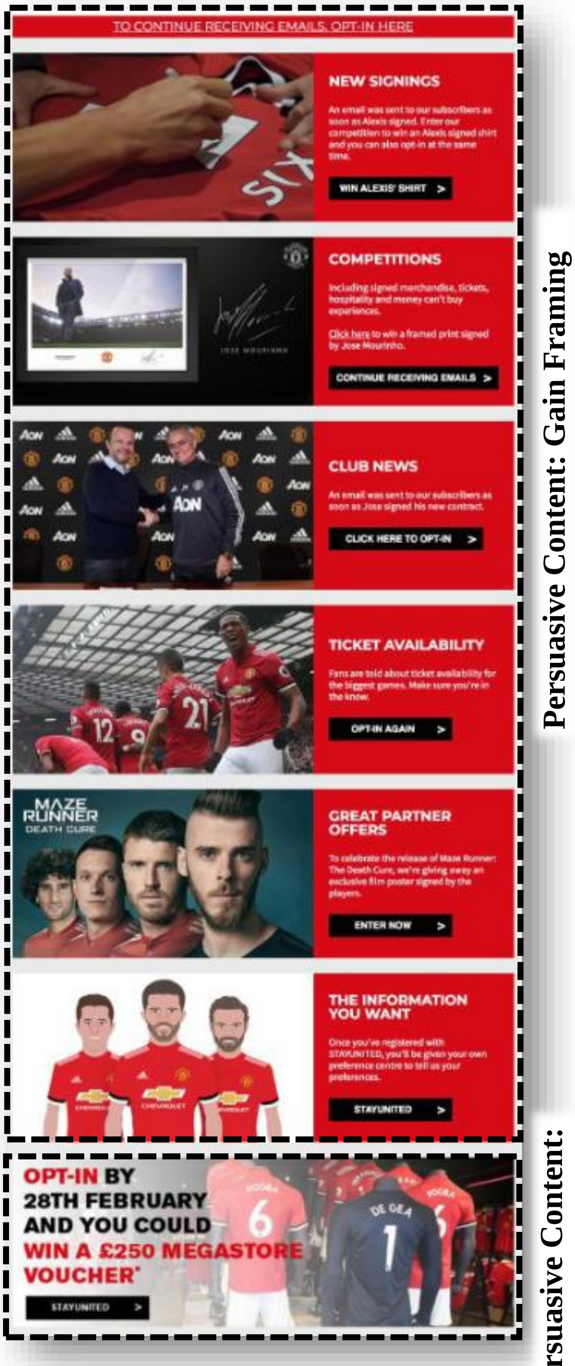
[3] The "Advertiser" dummy variable is equal to 1 if the company is only an advertiser (N = 962), and 0 if the company is (i) a publisher and advertiser (N = 105); (ii) only a publisher (N = 12); or (iii) missing (N=427). We controlled for the missing values by adding a dummy variable which assume value equal to 1 if the information is missing and 0 otherwise.

7.4. Results: Theory-Based Persuasive Themes over Firms' Self-Interest

As previously said, we estimated the firm's GDPR re-permission email content by using two alternative approaches. In this sub-chapter, we will focus our attention on the results achieved from the theory-driven methodology, which has been done using a manual coding procedure combined with predictive models (see Chapter 6.2 for details). Therefore, similarly to what we have done in the previous sub-chapter, we tested whether the singular themes that theory suggests being influential in altering consumer behavior – incentives and framing – are related to both the benefits and the risks firms evaluated when designing their GDPR data request.

This will allow us to reach other relevant results by unveiling which specific persuasive theme identified in the literature drives the results presented in Chapter 7.3 and identifying the characteristics of the companies that opted to use just one individual persuasive theme in their re-permission emails. Indeed, it is essential to highlight that the "*LDA Persuasive Topic*" variable used in Chapter 7.3 comprises the entire degree of persuasion of the emails, and it assumes higher values when the re-permission email contains more persuasiveness than information (e.g., more persuasive themes). Therefore, not only it signals whether the email is predominantly persuasive, but also it gives an indication about the number of specific persuasive elements which have been included in the data request. Figure 7.4.1 presents two examples of email that differ with regards to the degree of persuasion and the number of persuasive themes used. Foot Locker opted for a communication that is not predominantly persuasive and utilized as a persuasive element, just the inclusion of a monetary incentive (Panel A). In contrast, Manchester United decided to craft a strongly persuasive communication based on multiple persuasive cues: monetary incentives and the gain type of frame in its data request (Panel B). By looking at their scores in terms of persuasiveness, indeed, we can see that LDA Persuasive Topic is equal to 0.49 for the Foot Looker email, while it is 0.96 for the Manchester United communication. This, once again, suggests that the "*LDA Persuasive Topic*" variable can be considered as a total degree of persuasion of the email by considering not only the predominance of the persuasive text in the GDPR communication but also the variety of the persuasive themes that literature suggests being influential in prompting behavior.

Figure 7.4.1 – Examples of Degree of Persuasion in GDPR Re-Permission Emails.

Panel A: Foot Locker	Panel B: Manchester United
 ACTION REQUIRED NEW PRIVACY REGULATION **Informative Content: Transparency** As you might have heard, the new General Data Protection Regulation (GDPR) becomes effective on **25 May 2018**. The aim is to protect all EU citizens by making data processing transparent, in line with the purpose for which the data has been supplied and to protect the data from unauthorized data access and data breaches in an increasingly data-driven world. In order to enable us to continue to send you our newsletter with information on new products and events we kindly ask you to allow us to process your data in accordance with the General Data Protection Regulation (GDPR). Please carefully read the Foot Locker Data Privacy Policy and give us your consent to process your data, by clicking the below button "I agree". You can unsubscribe at any time. I AGREE **Persuasive Content: Monetary Incentive** If you decide to stay, we would like to say thank you with a personal 20% discount* on your next purchase. Your Foot Locker Team	 TO CONTINUE RECEIVING EMAILS, OPT-IN HERE NEW SIGNINGS An email was sent to our subscribers as soon as Alexis signed. Enter our competition to win an Alexis signed shirt and you can also opt-in at the same time. WIN ALEXIS' SHIRT > COMPETITIONS Including signed merchandise, tickets, hospitality and money can't buy experiences. Click here to win a framed print signed by Jose Mourinho. CONTINUE RECEIVING EMAILS > CLUB NEWS An email was sent to our subscribers as soon as Jose signed his new contract. CLICK HERE TO OPT-IN > TICKET AVAILABILITY Fans are told about ticket availability for the biggest games. Make sure you're in the know. OPT-IN AGAIN > GREAT PARTNER OFFERS To celebrate the release of Maze Runner: The Death Cure, we're giving away an exclusive film poster signed by the players. ENTER NOW > THE INFORMATION YOU WANT Once you've registered with STAYUNITED, you'll be given your own preference centre to tell us your preferences. STAYUNITED > OPT-IN BY 28TH FEBRUARY AND YOU COULD WIN A £250 MEGASTORE VOUCHER* STAYUNITED > **Persuasive Content: Gain Framing** **Persuasive Content: Monetary Incentive**
LDA Persuasive Topic = 0.49	**LDA Persuasive Topic = 0.96**

In contrast, this section deals with the analysis of the specific type of persuasive theme used in the firms' GDPR re-permission emails. Indeed, as just shown in the example in Figure 7.4.1, companies may choose between different options in terms of persuasive elements to include in their communications and may also decide to use just one persuasive cue. In this section, we want to explore the benefits and cost tradeoff that companies may have experienced in crafting their messages from another point of view by analyzing which one of the persuasive themes used singularly may drive the results we have seen in the previous chapter. The results from this analysis may also provide different insights from the ones presented in Chapter 7.3. Companies that have turned to communications that are mainly persuasive and that used multiple persuasion elements simultaneously may be different – both in terms of strategies and characteristics – from those that have opted for the inclusion of just one persuasive element. In addition, the models of this section will also allow us to both (i) corroborate the results achieved in Chapter 7.3 about the influence that benefits (of data access) and risks (of non-compliance) may have on the re-permission emails' content and (ii) get additional insights about the interactions which may exist among the specific themes included in these communications. Table 7.4.1 summarizes the parameter estimates of the models (robustness checks are provided in Appendix G).

We found results that closely resemble what we found in Chapter 7.3. By looking at the ***benefits***, these additional models highlight that companies that are expected to extract more value out of data are those that are also more likely to use persuasive themes ($\beta_{2\_Monetary}$= 0.042; $\beta_{2\_TimeOrientation}$= 0.013) and that there is no relationship between the amount of data collected and the use of persuasive themes. By analyzing the ***risks***, we found that companies that experienced more data breaches in the pre-GDPR period are less likely to shift their communications towards persuasion ($\gamma_{2\_Framing}$=-0.458) and that firm's reputation is not related to the use of persuasive cues. Additionally, we find consistent results also regarding the interaction between the use of persuasive and ***informative themes***: re-permission emails designed to be perceived as more informative in nature are less likely to rely on high levels of persuasion ($\delta_{6\_Monetary}$=-0.112; $\delta_{6\_Framing}$= -0.069; $\delta_{6\_TimeOrientation}$= -0.038).

However, the set of results presented in Figure 7.4.1 allows getting additional insights about how the specific persuasive theme has been used and the characteristics of the companies that decided to insert in their re-permission emails just one element of persuasion.

Firstly, it is possible to observe a consistent and relevant difference in the use of ***incentives*** in re-permission email communications. If, on the one hand, monetary incentives are used by companies that are generating more ad revenues from their websites ($\beta_{2\_Monetary}$= 0.042), on the other hand, firms opting for inserting non-monetary incentives in their re-permission emails are those which are less able to monetize the data collected ($\beta_{2\_NonMonetary}$= -0.038) and have experienced more data breaches in the past ($\gamma_{2\_NonMonetary}$= 0.268).

Secondly, relevant insights can be provided by looking at the ***control variables*** inserted in the model. Similarly to what we found in Chapter 7.3, companies that belong to the "*Media and Entertainment*" sector use persuasion to a lower degree ($\tau_{4\_Monetary}$= -0.223). However, thanks to the availability of more granular information about the specific persuasive theme a company can decide to use, we also find other interesting and slightly different results than those obtained in Chapter 7.3. We found that companies in the "*Retail Trade*" industry assign higher importance to the use of monetary incentives ($\tau_{6\_Monetary}$= 0.338) than to the type of frame of the message ($\tau_{6\_Framing}$= -0.233) to achieve their data-related goals. Interestingly, our results suggest that also companies operating in the "*Professional Services*" and in the "*Software and IT Services*" are less inclined to framing their re-permission emails in terms of gain or loss ($\tau_{5\_Framing}$= -0.246; $\tau_{7\_Framing}$= -0.397). However, in sharp contrast with what we have obtained in Chapter 7.3, companies based in the EU seems to be the ones that use only monetary incentives or framing to a great degree ($\tau_{1\_Monetary}$= 0.196; $\tau_{1\_Framing}$= 0.206).

Lastly, in line with our expectations and previous results, our results also prove that companies often opted for simultaneous use of persuasive themes in their GDPR re-permission emails. For example, the adoption of a specific time-related vocabulary is associated with the firm's choice of also using non-monetary incentives in its communication ($\delta_{2\_TimeOrientation}$= 0.184); firms tend to craft emails using a gain/loss type of framing and also to provide some sort of incentive in order to prompt disclosure behavior ($\delta_{1\_TimeOrientation}$= 1.036; $\delta_{2\_TimeOrientation}$= 3.035).

A detailed discussion of the results presented is provided in Chapter 7.5.

Table 7.4.1 – Fractional Logit Models' Results - DV = Theory-Based Persuasive Themes (Equation(3)).

		DV: Monetary Incentives		DV: Non-Monetary Incentives		DV: Framing: Gain/Loss		DV:Framing: Time Orientation	
Benefits	# Marketing Cookies (1)	0.000	(0.001)	0.001	(0.001)	0.000	(0.001)	0.000	(0.000)
	ln(Expected Monthly Online Ad Revenue) [1]	0.042***	(0.013)	-0.038***	(0.009)	-0.004	(0.013)	0.013***	(0.004)
Risks	Website Popularity (3 Months pre-GDPR)	-0.000	(0.000)	-0.000	(0.000)	0.000*	(0.000)	0.000	(0.000)
	# Data-Breaches (pre-GDPR)	0.012	(0.147)	0.268**	(0.114)	-0.458***	(0.117)	-0.034	(0.033)
Email's Content	Monetary Incentives			1.579***	(0.123)	1.036***	(0.172)	0.021	(0.055)
	Non-Monetary Incentives	2.349***	(0.192)			3.035***	(0.190)	0.184***	(0.060)
	Framing: Gain/Loss	0.960***	(0.154)	1.849***	(0.109)			0.071	(0.043)
	Framing: Time Orientation	0.387	(1.180)	2.392***	(0.842)	1.748*	(0.992)		
	Control	-0.045	(0.048)	-0.064	(0.041)	-0.036	(0.048)	-0.013	(0.013)
	Transparency	-0.112***	(0.041)	0.014	(0.034)	-0.069*	(0.041)	-0.038***	(0.011)
Controls	EU	0.196***	(0.075)	0.033	(0.057)	0.206***	(0.072)	-0.013	(0.020)
	Firm Size	0.018	(0.018)	0.008	(0.015)	-0.023	(0.018)	-0.012**	(0.005)
	Firm Age	-0.001	(0.001)	0.001	(0.001)	0.001	(0.001)	0.000	(0.000)
	Sectors:								
	Media and Entertainment	-0.223*	(0.119)	-0.032	(0.095)	0.028	(0.131)	-0.025	(0.034)
	Professional Services	0.088	(0.097)	0.100	(0.080)	-0.246**	(0.101)	0.013	(0.026)
	Retail Trade	0.338***	(0.104)	0.001	(0.093)	-0.233**	(0.115)	0.039	(0.030)
	Software and IT Services	0.070	(0.123)	0.152	(0.098)	-0.397***	(0.120)	-0.011	(0.030)
	Travel, Tourism and Hospitality	0.103	(0.127)	0.096	(0.095)	0.006	(0.119)	-0.019	(0.031)
	Advertiser (0/1)=1 [3]	0.224*	(0.126)	-0.059	(0.085)	-0.229*	(0.134)	-0.023	(0.033)
	Country Missing	-0.330	(0.266)	-0.335	(0.208)	0.366	(0.315)	-0.055	(0.065)
	Firm Size Missing	0.340*	(0.182)	-0.110	(0.129)	-0.252	(0.187)	-0.073*	(0.043)
	Firm Age Missing	-0.002	(0.192)	0.082	(0.132)	0.231	(0.187)	0.076*	(0.044)
	Advertiser Missing	0.185	(0.146)	0.076	(0.104)	-0.309**	(0.157)	0.002	(0.039)
	Constant	-2.599***	(0.290)	-2.125***	(0.216)	-1.625***	(0.270)	-1.926***	(0.072)
	Observations	1364 [2]		1364 [2]		1364 [2]		1364 [2]	
	Log-Pseudolikelihood	-528.53		-503.28		-546.54		-378.68	
	AIC	1103.06		1052.56		1139.07		803.36	
	BIC	1223.08		1172.58		1259.09		923.37	

Notes:

Standard errors in parentheses

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

[1] We add a small constant term to get around instances of zeroes in this variable as we take logs (Pattabhiramaiah, Sriram and Manchanda 2019).

[2] The sample size has been reduced because there are companies for which the expected online ad revenue figure was not available.

[3] The "Advertiser" dummy variable is equal to 1 if the company is only an advertiser (N = 962), and 0 if the company is (i) a publisher and advertiser (N = 105); (ii) only a publisher (N = 12); or (iii) missing (N=427). We controlled for the missing values by adding a dummy variable which assume value equal to 1 if the information is missing and 0 otherwise.

7.5. Discussion

In the left-hand side of our conceptual framework (Chapter 4), we claimed that there exists a conflict of interest for firms that have to adhere to the GDPR requirement – e.g., communicating "*clearly and understandably*" about their data practices – while, at the same time, have to find strategies to maintain and sustain their data-based business models. We were, consequently, interested in identifying whether firms' self-interest drove the design of the GDPR re-permission emails and the insertion of persuasive cues inside privacy-related communications that were meant to be merely informative. We contended that online firms designed their re-permission emails by carefully evaluating the benefits they can achieve from data access (e.g., data collection and online ad revenue) against the risks they may incur from not wholly adhere to the GDPR spirit (e.g., reputation and sanctions). In other words, we argued that firms shifted their communications toward persuasion only when they ascribe higher value to users' data and, at the same time, do not entirely perceive the risks related to data collection and management. The results of the first and second analyses (presented in Chapter 7.3 e 7.4) support our hypotheses.

It emerged that only firms significantly benefiting from data collection were willing to craft their communication in a persuasive way – e.g., by inserting incentives and strategically phrasing their re-permission emails – to increase consumers' likelihood to grant them data access. This is in line with the work done by Acquisti, Brandimarte, and Loewenstein (2015), in which they argue that three main themes are relevant in understanding what influences people's privacy decisions: instability and uncertainty of customers' privacy preferences, presence of context-dependencies, and malleability of privacy concerns. This latter factor suggests that marketing plays a relevant role in shaping individuals' privacy decisions by using subtle factors that can alter the perception of the risks related to personal information disclosure, prompting people to provide data that otherwise would be difficult to collect. Our first results add to this by highlighting that companies intentionally adopted persuasive communication elements to convince and cue consumers to provide their consent and that this decision is mainly connected to the degree of data monetization that the company itself is able to generate – either by using data internally or by selling them to other external data collectors.

Moreover, our second set of results highlights that **firms' expected returns on data are positively correlated with the presence of monetary incentives and negatively correlated with the presence of non-monetary incentives in the emails** (for examples of emails including Monetary or Non-Monetary Incentives, see Appendix C). This suggests that firms that are more likely to extract value from customers' data are less likely to use non-monetary incentives to persuade customers as if they considered this type of incentive less effective at granting permission. This may be due to the different psychological construal that monetary and non-monetary incentives generate in the customers. Previous literature suggests that these two types of incentives act separately on consumers' behaviors either because they involve different mental framings – e.g., "*non-monetary promotions are framed as segregated gains rather than reduced losses*" (Lowe and Barnes 2012, p. 2) – or because they are perceived differently in terms of distance in time from their realization – e.g., "*whereas monetary incentives provide an immediate financial benefit, there is only a small chance to win in a lottery*" (Krafft, Arden, and Verhoef 2017, p. 43). Consequently, companies that more urgently need data to get revenues may perceive the use of non-monetary incentives as less adequate in prompting people's actions and be more inclined to use monetary incentives. However, monetary incentives, such as discounts or coupons, also involve companies' financial costs. Our results, therefore, suggest that companies extensively profiting out of data make an effort to compensate customers for the data disclosed, meaning that they know their data value and are ready to share part of the revenue streams that they may generate with the data owners – who make those profits possible.

Additionally, we also showed that **persuasion decreased when companies experienced a higher number of data breaches before the GDPR enforcement**. Literature and business press suggest that companies are increasingly experiencing the negative externalities that data collection inherently bears. Data breaches have increased dramatically in the last decades, and, as a natural consequence, the number of legal actions and litigations that customers have taken against the companies exposed (Bleier, Goldfarb, and Tucker 2020; Carson 2020; Marotti 2020). Therefore, as a result, breached firms may have been more worried about the sanctions promised by the EU Legislator and have behaved more carefully on the occasion of the GDPR enforcement by more strictly adhering to its requirements. Nonetheless, this does not mean that companies with pre-GDPR data breaches have not used persuasion at all in their re-permission emails. In fact, we showed that they did use a persuasive communication element by including non-monetary incentives into their messages. However, as stated above, the literature suggests that non-monetary incentives should be considered more as a future possibility than an expected reality (Krafft, Arden, and Verhoef 2017). Therefore, we may contend that this type of incentive may be perceived as characterized by a lower degree of persuasion by customers.

Our set of results also consistently showed a **negative correlation between the presence of informative and persuasive themes**, meaning that it is possible, from the analysis of the content of the re-permission emails to separate companies which more strictly adhered to the GDPR requirements and firms that used strategically re-permission emails to reach their financial interests. This suggests that policymakers and regulators may use our text analysis procedures to identify and discriminate among the companies that have acted entirely in line with the GDPR principles – "*the good ones*" – and the ones that have tried to turn their communication into an opportunity to reach their interests – "*the bad ones*".

Lastly, our results also allow drafting a very rough but effective **identification of the characteristics of the firms trying to "*game the system*"**.

From the first set of results presented in Chapter 7.3, we detected the main features of firms that are more intensively using persuasive themes in their GDPR re-permission emails (e.g., by turning their message toward merely persuasion or by inserting more persuasive cues). Indeed, thanks to the controls added in this first set of models, it is possible to detect the overall main characteristics that firms that behaved opportunistically have. These are mainly non-European firms operating in the "*Travel, Tourism, and Hospitality*" industry. This first set of results may not be particularly surprising. Companies in the EU may have behaved more "*safely*" because they perceived more closely the Regulator intention and the possible consequences of non-compliance, whereas firms operating in the "*Travel, Tourism and Hospitality*" (e.g., Lastminute.com, Kiwi, Dominos, Pizza Hut) are those which tend to rely more on users' data to promote their business – for example, through the use of newsletters or retargeted emails – and, therefore, may need data more eagerly. Interestingly, companies in the "*Media and Entertainment*" industry seem to be those which are less likely to turn to high levels of persuasion. Firms operating in the "*Media and Entertainment*" industry are surely more known to heavily rely on data collection, which allows them to make profits (e.g., through advertising). However, by inspecting the companies belonging to the "*Media and Entertainment*" sector, we can see that these are companies such as Netflix, Spotify, Twitter, Facebook, or The Economist, which notably require in an *apriori* fashion users' data access for the user to be able to exploit their services (e.g., typically achieved through the user's sig-in). Therefore, they might not need to convince their consumers so strongly by using also persuasive cues to get data because it is their audience that is willing to provide them "*spontaneously*" in order to use the firms' online services.

From the second set of results presented in Chapter 7.4, we could instead discriminate between the differences among the companies that opted for the inclusion of just one persuasive theme in their communications. These are mainly European firms that belong to the "*Retail Trade*" sector. While the former aspect is in line with our expectations – retailing companies, such as Kroger, Nordstrom, or H&M, need users' data and are more used to use persuasion and distort the communications with their consumers to force users' actions (e.g., by sending newsletter or coupons) – the latter seems to be in sharp contrast with the result obtained from the first set of results. However, it suggests that there are differences in the way in which European companies decided to use persuasion. Taken together, the results from the first and the second analyses indicate that European companies use persuasion to a lower degree than non-European ones and focus their attention on just one persuasive element to obtain users' data. This may suggest that European firms may have been more meticulous in selecting the only persuasive element to include in their re-permission emails, and it seems that the choice fell on those elements which may be considered as more effective in prompting users' disclosure behavior: (i) the inclusion of monetary incentives or (ii) the use of a gain/loss frame to catch users' attention and increase their likelihood to achieve data access. This provides initial support to the superiority attributed to these two types of persuasive elements by firms.

All in all, our results indicate that companies crafted their re-permission email communications opportunistically depending on the benefits they can achieve and the risk they may take. We showed that the more the company is reliant on data monetization strategies, the more its communication shifts from being merely informative to be exclusively persuasive by adding, primarily, monetary incentives in the attempt to push their users to relinquish personal information. We also showed that the more the company's experienced data breaches in the past, the less its reliance on persuasion, especially in terms of message framing. However, our findings also suggest a particular use of non-monetary incentives. On the one side, they seem to be considered as less effective by companies that use and profit out of users' data more intensively, whereas, on the other side, they are perceived as a "*safety*" persuasive option by breached companies that may have felt more strongly the legislative pressure imposed by the GDPR Legislator. These results provide support to our stance, highlighting that companies experienced conflict of interest when designing their re-permission email, opting for higher levels of persuasion only when it is of utmost importance to achieve data access. The fact that companies that are intensively using data to achieve profit are also trying to "*game the system*" may have negative connotations as well as positive ones: one may argue that the achievement of an agreement with users on their data management by shifting the attention away from the main communication subject (e.g., data privacy and protection) may not be fair and entirely lawful, partially defeating the "*transparency*" pillar of the GDPR privacy regulation; however, the need to obtain consent from users, generated positive externalities for the data owners themselves, which are rewarded for their data disclosure (e.g., through the provision of a monetary incentive).

8. Conclusions

The present research aims to analyze the phenomenon of GDPR re-permission emails, study how companies developed their privacy-related communications, and examine the possible reasons that have led companies to craft them differently.

To achieve these goals, we content-analyzed a sample of 1506 re-permission emails that 1396 firms sent out on the GDPR enforcement occasion, and we related the firms' privacy-related communications to the *benefits* and the *risks* that firms evaluated when designing their GDPR messages. This approach allowed us to characterize the *themes* that companies used to achieve users' opt-in and uncover mechanisms that could have caused the observed differences in firms' privacy-related communications.

This thesis's first research question aims to better define the peculiarities of this type of privacy-related communications by identifying the *themes* that firms used in their GDPR email campaigns. Notably, the EU Regulator did not mandate how companies should design their re-permission emails. Therefore, we contended that firms used different arguments to craft their communications to obtain customers' opt-in. Previous literature studying consumers' disclosure behavior suggests that firms use different communicative elements to lessen customers' feelings of vulnerability and heighten their perception of security. In particular, two sets of factors have been shown to influence data disclosure: *informative* (e.g., transparency, control) and *persuasive* arguments (e.g., monetary and non-monetary incentives, framing in terms of gain/loss or time orientation). Our results indicate that a large proportion of firms used only *informative* themes in their emails, in that being completely compliant with the GDPR objective. However, we also showed that there is a considerable subset of communications that relied on *persuasion*, in that providing some preliminary evidence that companies tried to overcome the possible negative effects that privacy regulations entail in a very simplistic way. Even if the GDPR has not legally prevented firms from doing so, the use of *persuasive* elements in privacy-related communications surely does not hue to the EU Regulator's intent.

The second research questions of this dissertation aim to show that the tradeoff between the expected benefits from data usage and the expected costs from not completely adhere to the GDPR law's principles influence the design of firms' re-permission emails. We, therefore, argued that only companies that have relevant interests in getting data access are more likely to use *persuasive* arguments. In contrast, companies that more closely perceive the risks of not being fully compliant with the law's spirit are more prone to stick with the *informative* nature that these communications should have. Our results provide support to this contention.

Firstly, we show that firms which are more able to generate online ad revenues from the data collection are more likely to use *persuasive* cues in their data-related communications. This suggests that firms behaved in a self-interested way and designed their re-permission communication, bearing in mind the relevance of data to their marketing activities.

Secondly, our results indicate that the use of *persuasive* cues decreases if firms experienced problems with data-security before the GDPR enforcement (e.g., data-breaches), and they were more likely to design their re-permission communications in a "*clear and understandable*" manner – in line with the "*transparency"* requirement of the GDPR. This suggests that companies that more strongly perceive the harsh penalties from their concealed behavior –in terms of trust, online revenue, and GDPR sanctions – see the value of being more transparent and straightforward about their data practices.

Thirdly, our analysis indicates that firms with more data's financial potential (online ad revenues) use monetary incentives to cue consumers' disclosure. We observe the opposite for the use of non-monetary incentives, that are used by companies that extract little value out of the data collected and that, additionally, are more likely to have experienced data-security problems in the past. This result highlights that specific persuasive tools are perceived differently and are used by different types of firms. Interestingly, only firms strongly oriented to data exploitation are ready to provide monetary compensations to individuals for sharing their personal information and decided to share part of their data revenues with those who make these profits possible (e.g., data owners).

Lastly, our data permit to draw a raw profile of the firms which have opted for persuasion in their GDPR re-permission emails. In particular, we showed that non-European firms operating in the "*Travel, Tourism and Hospitality*" industry are more likely to use simultaneously different persuasive elements in their data-related communications. Instead, if we concentrate our attention only on the companies that used just one persuasive element in their messages, we found that these are mainly European companies that belong to the "*Retail Services*" sector and that this happens primarily by including monetary incentives. This may suggest that European firms may have behaved differently and more conservatively from non-European firms, preferring the choice of just one more efficient element of persuasion (e.g., discounts, coupons) to grant them data access instead than using more persuasive cues, which may completely change the re-permission email's content. This, indeed, may be considered misleading by the EU Regulator, who can also decide to opt for severe fines.

Overall, the findings of this dissertation have important implications for customers, managers, and policymakers.

Consumers need to realize that GDPR re-permission emails are likely driven more by self-interest than by compliance and that they might want to be more cautious when reading firms' privacy-related communications before taking any action.

Policymakers can use our results to understand the effects of re-permission emails on privacy-policy effectiveness. Although there is no doubt that freedom of choice was legally guaranteed to firms, we showed that those who were more interested in using commercial data asked for them in a tendentious fashion, partially defeating the purpose of the policy. Our analysis highlights that GDPR was not effective in reducing the firms' ability to generate value out of the data collected and demonstrates that companies become wiser in leveraging privacy-related communications in order to obtain users opt-in. Therefore, policymakers may use our analysis to detect companies to "*monitor*" as they are more likely interested in data and transforming their communication from *informative* to *persuasive*. Moreover, our research mainly suggests that leaving firms with freedom of choice does not always lead to the expected results (e.g., increased information about firms' data procedures) and indicates that it can be fruitful – for the GDPR Regulator – to define strategies to detect and overcome the "*workarounds*" that companies devised to encourage data disclosure. One potential suggestion may be to standardize the text of privacy-related communication that companies send to data owners. Another option can be to formalize a "*price for data*" that sets a standard exchange rate for the users' data among all the different players of the online sector (e.g., firms and customers). There are, with this regard, numerous real-world examples in which data have already become an exchange currency. One of the more recent ones regards the provision of COVID-19 vaccines by Pfizer to Israel in exchange for access to Israelis' health data (Schwartz and Trofimov 2021), strongly suggesting that data are increasingly used to complete economic transactions.

Our research can also provide valuable insights to managers who can have a better overview of the communication strategies that have been used so far to get users' data access. This can help them reason about the type of data request's content they should use, by carefully evaluating the pros and cons of the different types of design identified.

We believe that this research has theoretical and empirical contributions as well.

From a theoretical standpoint, this dissertation tries to provide increased knowledge to privacy-related literature by demonstrating that companies have been attempting to avoid the "*unintended but unavoidable*" consequences of privacy regulations by exploiting their privacy-related communications. In order to prove this, we leveraged two important streams of research. On the one hand, there are studies on how to request personal data (Acquisti, Brandimarte, and Loewenstein 2015; Martin 2018). On the other hand, literature provides evidence of the negative effects of privacy policy interventions on the whole online advertising ecosystem (Goldberg, Johnson, and Shriver

2019; Marotta, Abhishek, and Acquisti 2019; Peukert et al. 2020). By jointly considering these two pieces of literature, we proved that companies, on the occasion of the GDPR enforcement, tried to strategically craft their re-permission emails by inserting persuasive arguments in their communications to mitigate the expected adverse effects that the EU Regulation inherently entails. We also demonstrated that the rules of persuasive communications which are used in economic, marketing, and electoral settings (DellaVigna and Gentzkow 2010; Mueller and Stratmann 1994; Narayanan, Manchanda, and Chintagunta 2003) could also be applied to the privacy and data security context. The decision to opt for a highly persuasive communication is, indeed, the result of a meticulous cost/benefits analysis for the company, which acts opportunistically and designs its re-permission email such that it can reach its financial data-related interests. We also highlighted that this happens only when the benefits from data usage outweigh the expected costs from not completely adhere to the GDPR spirit.

From an empirical point of view, we believe that this dissertation provides some relevant advancements. Firstly, we provide one of the first practical attempts to apply Natural Language Processing techniques – particularly the Latent Dirichlet Allocation – to privacy-related communications and texts. The use of these types of unsupervised text analysis techniques is recent in academic literature and marketing works, and there is still a scarcity of empirical studies applying automatic machine learning algorithms for content analysis. Secondly, we developed a reliable approach to content-analyze privacy communications, which integrates both human and automated interventions and can be easily scaled to larger datasets to immediately get a sense of the degree of persuasion vs. information present in texts. Lastly, we provide a novel measure for the company's expected online advertising revenue. Previous research on the effects of privacy regulations on the online ecosystem has mainly looked at the impacts that a data regulation has on the company's ability to track its website users through marketing cookies. However, research on the influence that the new privacy regimes have on the companies' ability to generate positive ad revenue streams is still scarce.

Marotta, Abhishek, and Acquisti (2019) and Beales and Eisenach (2014) made some seminal attempts in this direction by focusing their attention on the effect that the absence of cookies may have on the impressions' CPM. Our measure elaborates on this by also considering the ad Click-Through-Rate and the total traffic that the website generates, becoming an overall estimate of the potential value of the data that a specific website may generate.

8.1. Limitations and Future Research

Despite the relevance of the results described above, this dissertation also suffers from several limitations. We did our best to cope with them, but data availability issues did not allow us to draw the complete picture of the GDPR re-permission email phenomenon.

Causality

Our study does not allow us to make any causal statements. With our data, we were able to describe the characteristics of the firms that have crafted these emails more aggressively: we observed that companies that have higher potential to generate online ad revenues are also more likely to insert persuasive elements into their communications and that the experience of a data-breach in the pre-GDPR period make companies more cautious and compliant with the law. We made an effort to control for the companies' pre-GDPR online strategy by inserting the pre-GDPR Alexa Ranking scores into our model. However, even if the presence of the pre-GDPR Alexa Rank may partially account for the firms' data reliance before the re-permission email phenomenon[14], we cannot directly state that the fact that firms were collecting more data or making more money out of them was the main reason to insert persuasive arguments into their privacy-related communications. Indeed, we miss data about the number of cookies and the online ad revenues that firms were generating before the GDPR enforcement. The availability of this type of pre-GDPR data would have allowed us to design a Diff-in-Diff model and solve the issue of causality, providing more insights about the

[14] More popularity means more traffic and engagement; this may lead to the collection of more data and, hence, a greater potential to extract value form users' information.

mechanisms behind companies' communication strategies and offering policymakers with real-based evidence about the effectiveness of the implemented privacy regulation. Therefore, future research can address this first issue by collecting additional information about companies’ websites just before the GDPR enforcement.

Opt-In Decisions

To have an overall view of the phenomenon, we need to have information about the *users’ opt-in behavior* to prove that the use of particular themes in the firms’ emails generates an increase (or a decrease) in the customers’ likelihood to relinquish data. Even if the literature has demonstrated that the use of persuasion is effective in prompting customers’ disclosure behavior (e.g., Athey, Catalini, and Tucker 2017; Grossklags and Acquisti 2007; Krafft, Arden, and Verhoef 2017), it would have been interesting to complement our framework with information about the opt-in rate of each communication we have collected. The availability of opt-in rates would, indeed, consent to draw conclusions about the most effective re-permission email design by suggesting which themes or interaction of themes (e.g., informative vs. persuasive) are more useful to prompt consumers to share their personal information. Additionally, since the GDPR also requires companies to collect opt-in permissions every year, companies can learn how to shape their subsequent re-permission email campaigns in order to achieve higher opt-in rates. Lastly, it would be interesting to explore any complementary/substitution effects between the two strategies to provide evidence about the perfect match of themes that privacy-related communication should incorporate to maximize the likelihood that users grant data access. The availability of research on these topics would generate essential implications for both managers – who can use the research to make their interaction with customers more effective and more instrumental to their purposes – and academics – who can learn how persuasive and informative themes may work together.

In Appendix H, we present some preliminary results from two randomized post-test experiments, where we provide corroboration results for the arguments presented in this thesis. We, indeed, show that (i) companies constructed their data requests in a way such that their consumers perceived them as highly informative or persuasive – suggesting an intentional behavior of the companies crafting GDPR re-permission emails – and (ii) mainly informative communication prevent consumers from disclosing – providing support for the view that companies were rightly warned about the effect that the GDPR "*Transparency*" requirement could have had on their online business. However, it could be fruitful to increase the generalizability of the results obtained from our experiments by running a field experiment in collaboration with a company that allows examining both individual-level revealed opt-in rates as well as the ROI of using incentives to obtain data.

Interaction effects between the presence of Informative and Persuasive Themes

Another limitation of our study regards the fact that, through our models, we did not produce any insights about the companies which have decided to use both informative and persuasive themes in their re-permission emails. Indeed, the analyses presented in this thesis describe only those companies that have designed their data requests in an extremely persuasive way by also combining more persuasive stimuli. This means that this study's results just describe online strategies and characteristics of companies that completely shifted their messages' content from information to merely persuasion (and vice versa). In our analyses, we indeed mapped only the extremes – e.g., messages with exclusively persuasive content (e.g., values of LDA Persuasive Topic close to 1) and messages with exclusively informative content (e.g., values of LDA Informative Topic close to 1) – but we were not able to provide any insight on the type of companies which decided to compose their emails with both an informative and persuasive content (e.g., values of LDA Persuasive Topic close to 0.5 and values of LDA Informative Topic close to 0.5). Therefore, our models did not give us the possibility to provide a complete overview of the re-permission email phenomenon and discriminate between the variety of communications that firms could have possibly designed.

Consequently, a further experimental investigation is needed to study the combined use of persuasive and informative elements in data-related communications and to clarify which may be the reasons that may have led firms to opt for this choice. Moreover, it could be fruitful to analyze how the inclusion of both stimuli may prompt different users' disclosure behavior since this may provide additional and valuable results, for both academics and policymakers, about the effectiveness that privacy-related regulations have on customers' welfare – in terms of security and protection.

Online Ad Revenue Measure

We also think that our ad revenue measure suffers from multiple limitations.

Firstly, we believe that the collection of a *single overall measure of the global expected online ad revenue* for the company's website may be limiting our analysis, driving us to possible misleading results. It would have been interesting to discern the ad revenues that companies may generate in different countries and to be able to compare the revenue streams generated in the EU – which were potentially affected by the GDPR – towards those produced in other parts of the world. In principle, this may isolate and measure the economic effect that the GDPR had on the European websites' ad revenues. Consequently, future research can try to dig deeper into the way in which the GDPR economically affected the monetization of European websites by retrieving the online ad revenue estimation at a country level.

Secondly, as highlighted in previous chapters, our *ad revenue measure should be considered as a sub-portion of the total ad revenue* that a website can generate. The platform that provided us with this information (SEMrush) bases its estimates on the Google AdSense Network only, in that losing information about all the ad revenues that websites may generate on other ad exchange platforms. Additionally, by looking at the last statistics provided by eMarketer, it seems that Google is losing market share in the online advertising industry in favor of Facebook and Amazon (EMarketer 2020). Therefore, an avenue for future research would be to generate a more comprehensive measure of a website's online ad revenue by also considering ad revenues from other ad exchange providers.

Thirdly, we also stressed how the *ad revenue measure* we collected should be interpreted as a *potential future revenue stream generated from the advertisement*. The measure we collected is mainly based on three components: the website's traffic, the average CPC, and the average CTR of the industry to which the website belongs. As can be noted, there is no component taking into consideration the fact that the revenue is actually generated and cashed by companies. Consequently, the interpretation of this measure *per se* can be seen as independent from the decision of the company to really monetize the website through ad placement. However, if we couple this estimation with additional information about how companies use their websites (e.g., publisher and/or advertiser type of website), we can produce a more in-depth analysis of the communication strategies implemented by companies that have different ways to monetize the data collection. Although we were able to retrieve information about the website usage, we could not perform any separate analyses because most of the companies composing our sample were classified as advertisers. Future research may build on this research and try to overcome this limitation by collecting additional re-permission emails sent out by publishers.

Users' Data Disclosure Compensation

Another potential limitation regards the lack of *information about the real monetary discount proposed* by companies in exchange for users' data. Our research indicates that firms leveraging behavioral advertising and targeting are willing to subsidize their customers financially, sharing some of the benefits they generated by customers' information with them, as this is likely to be a "win-win" scenario. The question that remains is to what extent this compensation is perceived as fair and desirable. Managers should calculate the optimal price for their customers while accounting for the risk of triggering reactance in data owners or negatively affecting their companies' reputation. Therefore, a research's possible development would be to set an empirical study to detect the "*correct*" compensation companies should give to their customers, not to create tensions with them and perpetuate their data-based strategies.

Additional Data Request Strategies

In this thesis, we studied the phenomenon of re-permission emails to analyze how companies ask for data. However, even if this is one of the most prominent ways companies used to obtain consumers' consent, especially in the GDPR enforcement, this is surely not the only one. There may be other strategies that companies may have implemented to influence users' disclosure behavior that may be worth a detailed investigation (e.g., default options, display of cookies' information). This can be an avenue for future research that may elaborate on this work and provide additional evidence about all the different ways companies have implemented to easily access users' data and their subsequent effectiveness in terms of users' opt-in decisions. The availability of comprehensive studies on firms' data request strategies would provide insights for both policy makers – who derive additional pieces of information about all the strategies used by companies to "game the system" which may be worth monitoring and regulating – and academics – which may be able to extensively describe firms' data request strategies and their consequences on the users' data disclosure behavior.

BIBLIOGRAPHY

Acquisti, Alessandro, Idris Adjerid, and Laura Brandimarte (2013), "Gone in 15 Seconds: The Limits of Privacy Transparency and Control," *IEEE Security & Privacy*, 11 (4), 72–74.

Acquisti, Alessandro, Laura Brandimarte, and George Loewenstein (2015), "Privacy and human behavior in the age of information," *Science*, 347 (6221), 509–14.

Acquisti, Alessandro, Allan Friedman, and Rahul Telang (2006), "Is There a Cost to Privacy Breaches? An Event Study," in *ICIS 2006 Proceedings - Twenty Seventh International Conference on Information Systems*.

Acquisti, Alessandro, Leslie K. John, and George Loewenstein (2013), "What Is Privacy Worth?," *The Journal of Legal Studies*, 42 (2), 249–74.

Adegbesan, J. Adetunji and Matthew J. Higgins (2011), "The intra-alliance division of value created through collaboration," *Strategic Management Journal*, 32 (2), 187–211.

Adjerid, Idris, Alessandro Acquisti, Laura Brandimarte, and George Loewenstein (2013), "Sleights of privacy: Framing, Disclosures, and the Limits of Transparency," in *Proceedings of the Ninth Symposium on Usable Privacy and Security - SOUPS '13*, New York, New York, USA: ACM Press, 1.

Aguirre, Elizabeth, Dominik Mahr, Dhruv Grewal, Ko De Ruyter, and Martin Wetzels (2015), "Unraveling the Personalization Paradox: The Effect of Information Collection and Trust-Building Strategies on Online Advertisement Effectiveness," *Journal of Retailing*, 91 (1), 34–49.

Allen, Mike (2017), *The SAGE Encyclopedia of Communication Research Methods*, SAGE.

AlSumait, Loulwah, Daniel Barbará, James Gentle, and Carlotta Domeniconi (2009), "Topic Significance Ranking of LDA Generative Models," in *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, 67–82.

Aridor, Guy, Yeon-Koo Che, and Tobias Salz (2020), "The Economic Consequences of Data Privacy Regulation: Empirical Evidence from GDPR," *SSRN Electronic Journal*.

Athey, Susan, Christian Catalini, and Catherine E. Tucker (2017), "The Digital Privacy Paradox: Small Money, Small Costs, Small Talk," *SSRN Electronic Journal*.

Aziz, Arslan and Rahul Telang (2016), "What Is a Digital Cookie Worth?," *SSRN Electronic Journal*.

Baum, Christopher F. (2008), "Stata tip 63: Modeling proportions," *Stata Journal*, 8 (2), 299–303.

Beales, Howard and Jeffrey A. Eisenach (2014), "An Empirical Analysis of the Value of Information Sharing in the Market for Online Content," *SSRN Electronic Journal*.

Benson, Vladlena, George Saridakis, and Hemamaali Tennakoon (2015), "Information disclosure of social media users," *Information Technology & People*, 28 (3), 426–41.

Berry, Steven, Martin Gaynor, and Fiona Scott Morton (2019), "Do Increasing Markups Matter? Lessons from Empirical Industrial Organization," *Journal of Economic Perspectives*, 33 (3), 44–68.

Blattberg, Robert C., Byung-Do Kim, and Scott A. Neslin (2008), “Customer Privacy and Database Marketing,” in *Database Marketing,* Springer, New York, 75–101.

Blei, David M., Andrew Y. Ng, and Michael I. Jordan (2003), “Latent Dirichlet Allocation,” *Journal of Machine Learning Research,* 3, 993–1022.

Bleier, Alexander and Maik Eisenbeiss (2015), “The Importance of Trust for Personalized Online Advertising,” *Journal of Retailing,* 91 (3), 390–409.

Bleier, Alexander, Avi Goldfarb, and Catherine E. Tucker (2020), “Consumer privacy and the future of data-based innovation and marketing,” *International Journal of Research in Marketing,* 37 (3), 466–80.

Boerman, Sophie C., Sanne Kruikemeier, and Frederik J. Zuiderveen Borgesius (2017), “Online Behavioral Advertising: A Literature Review and Research Agenda,” *Journal of Advertising,* 46 (3), 363–76.

Brandimarte, Laura, Alessandro Acquisti, and George Loewenstein (2013), “Misplaced Confidences,” *Social Psychological and Personality Science,* 4 (3), 340–47.

Brill, Julie (2011), “The intersection of consumer protection and competition in the new world of privacy,” *Competition Policy International,* 7 (1), 6–23.

Buckley, Jack (2003), “Estimation of Models with Beta-Distributed Dependent Variables: A Replication and Extension of Paolino’s Study,” *Political Analysis*, 11 (2), 204–5.

Budak, Ceren, Sharad Goel, Justin M Rao, and Georgios Zervas (2014), “Do-Not-Track and the Economics of Third-Party Advertising,” *SSRN Electronic Journal,* 1–38.

Bughin, Jaques, Tanguy Catlin, Martin Hirt, and Paul Willmott (2018), “Why Digital Strategies Fail,” *McKinsey Quarterly,* (accessed November 11, 2020), [available at https://www.mckinsey.com/business-functions/mckinsey-digital/our-insights/why-digital-strategies-fail].

Buis, Maarten L. (2010), “Analyzing Proportions,” in *Eighth German Stata Users Group Meeting*.

Cacioppo, John T., Richard E. Petty, Feng Kao Chuan, and Regina Rodriguez (1986), “Central and Peripheral Routes to Persuasion. An Individual Difference Perspective,” *Journal of Personality and Social Psychology,* 51 (5), 1032–43.

Campbell, James, Avi Goldfarb, and Catherine Tucker (2015), “Privacy Regulation and Market Structure,” *Journal of Economics & Management Strategy,* 24 (1), 47–73.

Carson, Angelique (2020), “GDPR ushers in civil litigation claims across the EU,” (accessed March 1, 2021), [available at https://iapp.org/news/a/gdpr-ushers-in-civil-litigation-claims-across-the-eu/].

Castro, Daniel (2010), “Stricter Privacy Regulations for Online Advertising Will Harm the Free Internet,” ITIF.

Chandon, Pierre, Brian Wansink, and Gilles Laurent (2000), “A Benefit Congruency Framework of Sales Promotion Effectiveness,” *Journal of Marketing,* 64 (4), 65–81.

Chellappa, Ramnath K. and Raymond G. Sin (2005), “Personalization versus Privacy: An Empirical Examination of the Online Consumer’s Dilemma,” *Information Technology and Management,* 6 (2–3), 181–202.

Chen, Brian X. (2018), “Getting a Flood of G.D.P.R.-Related Privacy Policy Updates? Read Them,” *The New York Times*, (accessed November 12, 2020), [available at https://www.nytimes.com/2018/05/23/technology/personaltech/what-you-should-look-for-europe-data-law.html].

Chung, Tuck Siong, Michel Wedel, and Roland T. Rust (2016), “Adaptive personalization using social networks,” *Journal of the Academy of Marketing Science*, 44 (1), 66–87.

Cook, Douglas O., Robert Kieschnick, and B.D. McCullough (2008), “Regression analysis of proportions in finance with self selection,” *Journal of Empirical Finance*, 15 (5), 860–67.

D’Annunzio, Anna and Antonio Russo (2020), “Ad Networks and Consumer Tracking,” *Management Science*, 66 (11), 5040–58.

Davis, Ben (2018), “GDPR: 15 good & bad examples of repermissioning emails & campaigns,” *Econsultancy*, (accessed November 12, 2020), [available at https://econsultancy.com/gdpr-examples-repermissioning-emails-campaigns/].

Degeling, Martin, Christine Utz, Christopher Lentzsch, Henry Hosseini, Florian Schaub, and Thorsten Holz (2018), “We Value Your Privacy ... Now Take Some Cookies: Measuring the GDPR’s Impact on Web Privacy,” *Proceedings 2019 Network and Distributed System Security Symposium*, (February).

DellaVigna, Stefano and Matthew Gentzkow (2010), “Persuasion: Empirical Evidence,” *Annual Review of Economics*, 2 (1), 643–69.

Detweiler, Jerusha B., Brian T. Bedell, Peter Salovey, Emily Pronin, and Alexander J. Rothman (1999), “Message framing and sunscreen use: Gain-framed messages motivate beach-goers.,” *Health Psychology*, 18 (2), 189–96.

Dinev, Tamara and Paul Hart (2004), “Internet privacy concerns and their antecedents - measurement validity and a regression model,” *Behaviour & Information Technology*, 23 (6), 413–22.

Droge, Cornelia (1989), “Shaping the Route to Attitude Change: Central versus Peripheral Processing through Comparative versus Noncomparative Advertising,” *Journal of Marketing Research*, 26 (2), 193.

Edelman, David C. and Marc Singer (2015), “Competing on Customer Journeys,” *Harvard Business Review*, (accessed November 11, 2020), [available at https://hbr.org/2015/11/competing-on-customer-journeys].

Elving, Wim J. L., Ursa Golob, Klement Podnar, Anne Ellerup - Nielsen, and Christa Thomson (2015), “The bad, the ugly and the good: new challenges for CSR communication,” *Corporate Communications: An International Journal*, (D. Wim J.L. Elving, Dr Ursa Golob, Dr, ed.), 20 (2), 118–27.

EMarketer (2020), “Google Ad Revenues to Drop for the First Time,” (accessed November 13, 2020), [available at https://www.emarketer.com/content/google-ad-revenues-drop-first-time].

European Union (1995), “Directive 95/ /EC of the European parliament and of the council of 24 October 1995 on the protection of individuals with regard to the processing of personal data and on the free movement of such data,” *Official Journal of European Union*, L 281/31.

European Union (2016), “Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 on the Protection of Natural Persons with Regard to the Processing of

Personal Data and on the Free Movement of Such Data, and Repealing Directive 95/46/EC (General Da," *Official Journal of European Union*, L 119/1.

Farahat, Ayman and Michael Bailey (2013), "How effective is targeted advertising?," in *2013 American Control Conference*, IEEE, 6014–21.

Farrell, Joseph (2012), "Can privacy be just another good?," *Journal on Telecommunication & High Technology Law*, 10, 251–61.

Ferrari, Silvia and Francisco Cribari-Neto (2004), "Beta Regression for Modelling Rates and Proportions," *Journal of Applied Statistics*, 31 (7), 799–815.

Fisher, Lauren (2019), "Digital Marketing in Today's Privacy-Conscious World," eMarketer.

Fitzsimons, Gavan J. and Donald R. Lehmann (2004), "Reactance to Recommendations: When Unsolicited Advice Yields Contrary Responses," *Marketing Science*, 23 (1), 82–94.

Foote, Wendy (2019), "GDPR and CCPA, Which Might Pave the Way to Federal Law?," CPO Magazine, (accessed February 5, 2021), [available at https://www.cpomagazine.com/data-protection/will-ccpa-pave-the-way-for-a-national-privacy-law/].

Forward Action (2018), "GDPR : How Changing Your Opt in Language Can Increase Consent Rate by 50 %," *Forward Action*, (accessed November 12, 2020), [available at https://medium.com/@forward_action/gdpr-how-changing-your-opt-in-language-can-increase-consent-rate-by-50-f9cffe1f6f22].

Gallani, Susanna, Ranjani Krishnan, and Jeff Wooldridge (2015), "Applications of Fractional Response Model to the Study of Bounded Dependent Variables in Accounting Research," *SSRN Electronic Journal*.

Gamliel, Eyal and Ram Herstein (2007), "The effect of framing on willingness to buy private brands," *Journal of Consumer Marketing*, 24 (6), 334–39.

Georgiadis, Christos K., Nikolaos Polatidis, Haralambos Mouratidis, and Elias Pimenidis (2017), "A Method for Privacy-preserving Collaborative Filtering Recommendations," *Journal of Universal Computer Science*, 23 (2), 146–66.

Ghosh, Dipayan (2018), "How GDPR Will Transform Digital Marketing," *Harvard Business Review*, (accessed November 11, 2020), [available at https://hbr.org/2018/05/how-gdpr-will-transform-digital-marketing].

Goldberg, Samuel, Garrett A. Johnson, and Scott K. Shriver (2019), "Regulating Privacy Online: The Early Impact of the GDPR on European Web Traffic & E-Commerce Outcomes," *SSRN Electronic Journal*.

Goldfarb, Avi and Catherine E. Tucker (2011), "Privacy regulation and online advertising," *Management Science*, 57 (1), 57–71.

Grewal, Dhruv, Jerry Gotlieb, and Howard Marmorstein (1994), "The Moderating Effects of Message Framing and Source Credibility on the Price-Perceived Risk Relationship," *Journal of Consumer Research*, 21 (1), 145.

Griffiths, Thomas L. and Mark Steyvers (2004), "Finding scientific topics," *Proceedings of the National Academy of Sciences*, 101 (Supplement 1), 5228–35.

Groenfeldt, Tom (2015), "CEOs Are In The Dark About How Their Firms Use Data," *Forbes*, (accessed November 12, 2020), [available at https://www.forbes.com/sites/tomgroenfeldt/2015/02/05/ceos-are-in-the-dark-about-how-their-firms-use-data/?sh=6c97fb5236d8].

Grossklags, Jens and Alessandro Acquisti (2007), "When 25 Cents is too much: An Experiment on Willingness-To-Sell and Willingness-To-Protect Personal Information," in *Workshop on the Economics of Information Security*.

Guo, Tong, Srinivasaraghavan Sriram, and Puneet Manchanda (2020), "'Let the Sunshine In': The Impact of Industry Payment Disclosure on Physician Prescription Behavior," *Marketing Science*, 39 (3), 516–39.

Habermas, Jürgen (1985), *The theory of communicative action: Reason and the rationalization of society (Vol. 1).*, Beacon Press.

Hanson, Nicole and Wonjoo Yun (2018), "Should 'big food' companies introduce healthier options? The effect of new product announcements on shareholder value," *Marketing Letters*, 29 (1), 1–12.

Hardin, James W. and Joseph M. Hilbe (2014), "Estimation and testing of binomial and beta-binomial regression models with and without zero inflation," *The Stata Journal*, 14 (2), 292–303.

Hausman, Angela V. and Jeffrey Sam Siekpe (2009), "The effect of web interface features on consumer online purchase intentions," *Journal of Business Research*, 62 (1), 5–13.

Havlena, William J. and Susan L. Holak (1991), "'The Good Old Days': Observations on Nostalgia and Its Role in Consumer Behavior," *NA - Advances in Consumer Research*, 18 (1), 323–29.

Holtrop, Niels, Jaap E. Wieringa, Maarten J. Gijsenberg, and Peter C. Verhoef (2017), "No future without the past? Predicting churn in the face of customer privacy," *International Journal of Research in Marketing*, 34 (1), 154–72.

Hui, Teo, and Lee (2007), "The Value of Privacy Assurance: An Exploratory Field Experiment," *MIS Quarterly*, 31 (1), 19.

Iordanou, Costas, Georgios Smaragdakis, Ingmar Poese, and Nikolaos Laoutaris (2018), "Tracing Cross Border Web Tracking," in *Proceedings of the Internet Measurement Conference 2018*, New York, NY, USA: ACM, 329–42.

Janakiraman, Ramkumar, Joon Ho Lim, and Rishika Rishika (2018), "The Effect of a Data Breach Announcement on Customer Behavior: Evidence from a Multichannel Retailer," *Journal of Marketing*, 82 (2), 85–105.

Jay, Cline (2017), "GDPR Compliance Top Data Protection Priority for 92% of US Organizations in 2017, According to PwC Survey," *PwC*, (accessed November 25, 2020), [available at https://www.pwc.com/us/en/press-releases/2017/pwc-gdpr-compliance-press-release.html].

Jia, Jian, Ginger Zhe Jin, and Liad Wagman (2019), "GDPR and the Localness of Venture Investment," *SSRN Electronic Journal*.

Jin, Ginger Zhe and Liad Wagman (2020), "Big data at the crossroads of antitrust and consumer protection," *Information Economics and Policy*, (xxxx), 100865.

Johnson, Garrett A. (2013), "The Impact of Privacy Policy on the Auction Market for Online Display Advertising," *SSRN Electronic Journal*.

Johnson, Garrett A. and Scott K. Shriver (2019), "Privacy & Market Concentration: Intended &

Unintended Consequences of the GDPR," *SSRN Electronic Journal*.

Johnson, Garrett A., Scott K. Shriver, and Shaoyin Du (2020), "Consumer Privacy Choice in Online Advertising: Who Opts Out and at What Cost to Industry?," *Marketing Science*, 39 (1), 33–51.

Kahneman, Daniel and Amos Tversky (1979), "Prospect Theory: An Analysis of Decision under Risk," *Econometrica*, 47 (2), 263.

Kapadia, Shashank (2019), "Evaluate Topic Models: Latent Dirichlet Allocation (LDA)," *Towards Data Science*, (accessed December 4, 2020), [available at https://towardsdatascience.com/evaluate-topic-model-in-python-latent-dirichlet-allocation-lda-7d57484bb5d0].

Krafft, Manfred, Christine M. Arden, and Peter C. Verhoef (2017), "Permission Marketing and Privacy Concerns — Why Do Customers (Not) Grant Permissions?," *Journal of Interactive Marketing*, 39, 39–54.

Ku, Hsuan-Hsuan, Po-Hsiang Yang, and Chia-Lun Chang (2018), "Reminding customers to be loyal: does message framing matter?," *European Journal of Marketing*, 52 (3/4), 783–810.

Lambrecht, Anja (2017), "E-Privacy Provisions and Venture Capital Investments in the EU."

Lambrecht, Anja and Catherine Tucker (2015), "Can Big Data Protect a Firm from Competition?," *SSRN Electronic Journal*, 1–20.

Leffler, Keith B. (1981), "Persuasion or Information? The Economics of Prescription Drug Advertising," *The Journal of Law & Economics*, 24 (1), 45–74.

Lehmann, Donald R., Sunil Gupta, and Joel H. Steckel (1998), *Marketing Research*, Prentice Hall.

Lerner, Josh (2011), "The Impact of Copyright Policy Changes on Venture Capital Investment in Cloud Computing Companies," *Architecture*.

Levin, Irwin P., Sandra L. Schneider, and Gary J. Gaeth (1998), "All Frames Are Not Created Equal: A Typology and Critical Analysis of Framing Effects," *Organizational Behavior And Human Decision Processes*, 76 (2), 149–88.

Lewis, Randall A. and David H. Reiley (2014), "Online ads and offline sales: Measuring the effect of retail advertising via a controlled experiment on Yahoo!," *Quantitative Marketing and Economics*, 12 (3), 235–66.

Libert, Timothy, Lucas Graves, and Rasmus Kleis Nielsen (2018), "Changes in Third-Party Content on European News Websites after GDPR," Oxford: Reuters Institute for the Study of Journalism.

Libert, Timothy and Rasmus Kleis Nielsen (2018), "Third-Party Web Content on EU News Sites: Potential Challenges and Paths to Privacy Improvement," Oxford: Reuters Institute for the Study of Journalism.

Loch, Adam, Peter Boxall, and Sarah Ann Wheeler (2016), "Using proportional modeling to evaluate irrigator preferences for market-based water reallocation," *Agricultural Economics*, 47 (4), 387–98.

Lohr, Steve (2012), "The Age of Big Data," *The New York Times*, (accessed November 11, 2020), [available at https://www.nytimes.com/2012/02/12/sunday-review/big-datas-impact-in-the-world.html].

Lowe, Ben and Bradley R. Barnes (2012), “Consumer perceptions of monetary and non-monetary introductory promotions for new products,” *Journal of Marketing Management*, 28 (5–6), 629–51.

Lu, Bin, Myle Ott, Claire Cardie, and Benjamin K. Tsou (2011), “Multi-aspect Sentiment Analysis with Topic Models,” in *2011 IEEE 11th International Conference on Data Mining Workshops*, IEEE, 81–88.

Malhotra, Naresh K., Sung S. Kim, and James Agarwal (2004), “Internet Users’ Information Privacy Concerns (IUIPC): The Construct, the Scale, and a Causal Model,” *Information Systems Research*, 15 (4), 336–55.

Manchanda, Puneet, Jean-Pierre Dubé, Khim Yong Goh, and Pradeep K Chintagunta (2006), “The Effect of Banner Advertising on Internet Purchasing,” *Journal of Marketing Research*, 43 (1), 98–108.

Marotta, Veronica, Vibhanshu Abhishek, and Alessandro Acquisti (2019), “Online Tracking and Publishers ’ Revenues : An Empirical Analysis,” in *Workshop on the Economics of Information Security*, 1–35.

Marotti, Ally (2020), “After Facebook agrees to pay $ 550 million settlement to Illinois users , Google sued over same law,” *Chicago Tribune*, (accessed March 1, 2021), [available at https://www.chicagotribune.com/business/ct-biz-google-facial-recognition-lawsuit-20200211-3pzwk7qwznhfvgdv5lk2bdyqu4-story.html].

Martin, Kelly D., Abhishek Borah, and Robert W. Palmatier (2017), “Data Privacy: Effects on Customer and Firm Performance,” *Journal of Marketing*, 81 (1), 36–58.

Martin, Kelly D. and Patrick E. Murphy (2017), “The role of data privacy in marketing,” *Journal of the Academy of Marketing Science*, 45 (2), 135–55.

Martin, Kirsten (2018), “The penalty for privacy violations: How privacy violations impact trust online,” *Journal of Business Research*, 82, 103–16.

McAfee, Andrew and Erik Brynjolfsson (2012), “Big data: The management revolution,” *Harvard Business Review*, (accessed November 12, 2020), [available at https://hbr.org/2012/10/big-data-the-management-revolution].

McDowell, Allen and Nicholas J Cox (2004), “Logit transformation Author,” *STAT Corporation*, (accessed March 18, 2021), [available at https://www.stata.com/support/faqs/statistics/logit-transformation/].

Mebane, Walter R. (2000), “Coordination, Moderation, and Institutional Balancing in American Presidential and House Elections,” *American Political Science Review*, 94 (1), 37–57.

Meyerowitz, Beth E. and Shelly Chaiken (1987), “The effect of message framing on breast self-examination attitudes, intentions, and behavior.,” *Journal of Personality and Social Psychology*, 52 (3), 500–510.

Mikkelsen, Daniel, Kayvaun Rowshankish, Henning Soller, and Kalin Stamenov (2017), “Tackling GDPR compliance before time runs out,” *McKinsey & Company*, (accessed November 11, 2020), [available at https://www.mckinsey.com/business-functions/risk/our-insights/tackling-gdpr-compliance-before-time-runs-out].

Miller, Amalia R. and Catherine E. Tucker (2009), “Privacy Protection and Technology Diffusion: The Case of Electronic Medical Records,” *Management Science*, 55 (7), 1077–93.

Miller, Amalia R. and Catherine E. Tucker (2011), "Can Healthcare IT Save Babies?," *SSRN Electronic Journal*.

Miller, Klaus and Bernd Skiera (2017), "Economic Damage of Cookie Lifetime Restrictions," *SSRN Electronic Journal*, (October), 1998–2004.

Milne, George R. and Mary J. Culnan (2004), "Strategies for reducing online privacy risks: Why consumers read (or don't read) online privacy notices," *Journal of Interactive Marketing*, 18 (3), 15–29.

Mimno, David, Hanna M. Wallach, Edmund Talley, Miriam Leenders, and Andrew McCallum (2011), "Optimizing semantic coherence in topic models," *EMNLP 2011 - Conference on Empirical Methods in Natural Language Processing, Proceedings of the Conference*, (2), 262–72.

Mohan, Bhavya, Ryan W. Buell, and Leslie K. John (2019), "Lifting the Veil: The Benefits of Cost Transparency," *SSRN Electronic Journal*.

Mueller, Dennis C. and Thomas Stratmann (1994), "Informative and persuasive campaigning," *Public Choice*, 81 (1–2), 55–77.

Murphy, Hannan (2020), "California's Privacy Law Arrives to Confusion and Costs for Businesses," *Financial Time*, (accessed February 5, 2021), [available at https://www.ft.com/content/7b541808-2bdf-11ea-bc77-65e4aa615551].

Narayanan, Sridhar, Puneet Manchanda, and Pradeep K. Chintagunta (2003), "The Informative versus Persuasive Role of Marketing Communication in New Product Categories: An Application to the Prescription Antihistamines Market," *SSRN Electronic Journal*.

Neumann, Nico, Catherine E. Tucker, and Timothy Whitfield (2019), "Frontiers: How effective is third-party consumer profiling? evidence from field studies," *Marketing Science*, 38 (6), 918–26.

Nysveen, Herbjørn, Per E. Pedersen, and Helge Thorbjørnsen (2005), "Intentions to Use Mobile Services: Antecedents and Cross-Service Comparisons," *Journal of the Academy of Marketing Science*, 33 (3), 330–46.

Ospina, Raydonal and Silvia L. P. Ferrari (2010), "Inflated beta distributions," *Statistical Papers*, 51 (1), 111–26.

Ospina, Raydonal and Silvia L. P. Ferrari (2011), "A general class of zero-or-one inflated beta regression models," *Computational Statistics and Data Analysis*, 56 (6), 1609–23.

Paolino, Philip (2001), "Maximum Likelihood Estimation of Models with Beta-Distributed Dependent Variables," *Political Analysis*, 9 (4), 325–46.

Papke, Leslie E. and Jeffrey M. Wooldridge (1996), "Econometric methods for fractional response variables with an application to 401(k) plan participation rates," *Journal of Applied Econometrics*, 11 (6), 619–32.

Pattabhiramaiah, Adithya, S. Sriram, and Puneet Manchanda (2019), "Paywalls: Monetizing Online Content," *Journal of Marketing*, 83 (2), 19–36.

Peppers, Don and Martha Rogers (2016), *Managing Customer Experience and Relationships: A Strategic Framework*, John Wiley & Sons Inc.

Petty, Richard E. and John T. Cacioppo (1984), “The effects of involvement on responses to argument quantity and quality: Central and peripheral routes to persuasion,” *Journal of Personality and Social Psychology*, 46 (1), 69–81.

Petty, Richard E. and John T. Cacioppo (1986), “The elaboration likelihood model of persuasion,” *Advances in Experimental Social Psychology*, 19 (C), 123–205.

Petty, Richard E, Jamie Barden, and S. C. Wheeler (2002), “The Elaboration Likelihood Model of persuasion: health promotions that yield sustained behavioral change,” *Emerging Theories in Health Promotion Practive and Research*, 71–99.

Peukert, Christian, Stefan Bechtold, Michail Batikas, and Tobias Kretschmer (2020), “European Privacy Law and Global Markets for Data,” *SSRN Electronic Journal*.

Phelps, Joseph, Glen Nowak, and Elizabeth Ferrell (2000), “Privacy Concerns and Consumer Willingness to Provide Personal Information,” *Journal of Public Policy & Marketing*, 19 (1), 27–41.

Prince, Christine (2018), “Do consumers want to control their personal data? Empirical evidence,” *International Journal of Human-Computer Studies*, 110, 21–32.

Puranam, Dinesh, Vishal Narayan, and Vrinda Kadiyali (2017), “The Effect of Calorie Posting Regulation on Consumer Opinion: A Flexible Latent Dirichlet Allocation Model with Informative Priors,” *Marketing Science*, 36 (5), 726–46.

Rothman, Alexander J., Roger D. Bartels, Jhon Wlaschin, and Peter Salovey (2006), “The Strategic Use of Gain- and Loss-Framed Messages to Promote Healthy Behavior: How Theory Can Inform Practice,” *Journal of Communication*, 56 (suppl_1), S202–20.

Rutz, Oliver J., Michael Trusov, and Randolph E. Bucklin (2011), “Modeling Indirect Effects of Paid Search Advertising: Which Keywords Lead to More Future Visits?,” *Marketing Science*, 30 (4), 646–65.

Sanchez-Rola, Iskander, Matteo Dell’Amico, Platon Kotzias, Davide Balzarotti, Leyla Bilge, Pierre-Antoine Vervier, and Igor Santos (2019), “Can I Opt Out Yet?,” in *Proceedings of the 2019 ACM Asia Conference on Computer and Communications Security*, New York, NY, USA: ACM, 340–51.

SanJosé-Cabezudo, Rebeca, Ana M. Gutiérrez-Arranz, and Jesús Gutiérrez-Cillán (2009), “The Combined Influence of Central and Peripheral Routes in the Online Persuasion Process,” *CyberPsychology & Behavior*, 12 (3), 299–308.

Schwartz, Felicia and Yaroslav Trofimov (2021), “How Israel Delivered the World ’ s Fastest Vaccine Rollout,” *Wall Street Journal*, (accessed April 8, 2021), [available at https://www.wsj.com/articles/how-israel-delivered-the-worlds-fastest-vaccine-rollout-11616080968].

Sen, Sankar, Zeynep Gürhan-Canli, and Vicki Morwitz (2001), “Withholding Consumption: A Social Dilemma Perspective on Consumer Boycotts,” *Journal of Consumer Research*, 28 (3), 399–417.

Sharma, Priyanka, Yidan Sun, and Liad Wagman (2019), “The Differential Effects of New Privacy Protections on Publisher and Advertiser Profitability,” *SSRN Electronic Journal*.

Sheehan, Kathleen M., Irene Kostin, Diane Napolitano, and Michael Flor (2014), “The TextEvaluator Tool: Helping Teachers and Test Developers Select Texts for Use in Instruction and Assessment," *The Elementary School Journal*, 115 (2), 184–209.

Shokri, Reza and Vitaly Shmatikov (2015), "Privacy-Preserving Deep Learning," in *Proceedings of the 22nd ACM SIGSAC Conference on Computer and Communications Security*, 1310–21.

Shrimsley, Robert (2018), "Brexit : the Conservatives and their thirty years ' war over Europe," *Financial Times*, (accessed November 11, 2020), [available at https://www.ft.com/content/0dee56c0-fdfa-11e8-ac00-57a2a826423e].

Sievert, Carson and Kenneth Shirley (2014), "LDAvis: A method for visualizing and interpreting topics," in *Proceedings of the Workshop on Interactive Language Learning, Visualization, and Interfaces*, Stroudsburg, PA, USA: Association for Computational Linguistics, 63–70.

Smithson, Michael and Jay Verkuilen (2006), "A better lemon squeezer? Maximum-likelihood regression with beta-distributed dependent variables.," *Psychological Methods*, 11 (1), 54–71.

Soleymanian, Miremad, Charles B. Weinberg, and Ting Zhu (2019), "Sensor Data and Behavioral Tracking: Does Usage-Based Auto Insurance Benefit Drivers?," *Marketing Science*, 38 (1), 21–43.

Solove, Daniel J. and Paul Schwartz (2014), *Information privacy law*, Wolters Kluwer Law & Business.

Son, Jai Yeol and Sung S. Kim (2008), "Internet users' information privacy-protective responses: A Taxonomy and a nomological model," *MIS Quarterly: Management Information Systems*, 32 (3), 503–29.

Sørensen, Jannick and Sokol Kosta (2019), "Before and After GDPR: The Changes in Third Party Presence at Public and Private European Websites," in *The World Wide Web Conference on - WWW '19*, New York, New York, USA: ACM Press, 1590–1600.

Stahl, Bernd Carsten and David Wright (2018), "Ethics and Privacy in AI and Big Data: Implementing Responsible Research and Innovation," *IEEE Security and Privacy*, 16 (3), 26–33.

Stutzman, Fred, Ralph Gross, and Alessandro Acquisti (2013), "Silent Listeners: The Evolution of Privacy and Disclosure on Facebook," *Journal of Privacy and Confidentiality*, 4 (2), 7–41.

Taillard, Marie-Odile (2000), "Persuasive communication : The case of Marketing."

Tang, Chuanyi and Lin Guo (2015), "Digging for gold with a simple tool: Validating text mining in studying electronic word-of-mouth (eWOM) communication," *Marketing Letters*, 26 (1), 67–80.

Tausczik, Yla R. and James W. Pennebaker (2010), "The Psychological Meaning of Words: LIWC and Computerized Text Analysis Methods," *Journal of Language and Social Psychology*, 29 (1), 24–54.

TextEvaluator (2017), "About the TextEvaluator® Technology," ETS.

Thaler, Richard (1980), "Toward a positive theory of consumer choice," *Journal of Economic Behavior & Organization*, 1 (1), 39–60.

The Guardian (2021), "Judge approves $ 650m settlement of privacy lawsuit against Facebook," *The Guardian*, (accessed March 1, 2021), [available at https://www.theguardian.com/technology/2021/feb/27/facebook-illinois-privacy-lawsuit-settlement].

Thibaut, J. W. and H. H. Kelley (1959), *The social psychology of groups*, John Wiley & Sons Inc, New York.

Thompson, Barney (2018), "Most UK small businesses unprepared for new EU data rules," *Financial Times*, (accessed November 12, 2020), [available at https://www.ft.com/content/87a11d2c-1e35-11e8-aaca-4574d7dabfb6].

Tikkinen-Piri, Christina, Anna Rohunen, and Jouni Markkula (2018), "EU General Data Protection Regulation: Changes and implications for personal data collecting companies," *Computer Law & Security Review*, 34 (1), 134–53.

Tucker, Catherine E. (2014), "Social Networks, Personalized Advertising, and Privacy Controls," *Journal of Marketing Research*, 51 (5), 546–62.

Utz, Christine, Martin Degeling, Sascha Fahl, Florian Schaub, and Thorsten Holz (2019), "(Un)informed Consent: Studying GDPR Consent Notices in the Field," in *Proceedings of the 2019 ACM SIGSAC Conference on Computer and Communications Security*, ACM, 973–90.

ValeoNetworks (2020), "How CCPA & GDPR Affect Your Business' Data Systems," (accessed February 5, 2021), [available at https://www.valeonetworks.com/how-ccpa-gdpr-affect-your-business-data-systems/].

Walsh, Joe (2021), "TikTok Settles Privacy Lawsuit For $ 92 Million," *Forbes*, (accessed March 1, 2021), [available at https://www.forbes.com/sites/joewalsh/2021/02/25/tiktok-settles-privacy-lawsuit-for-92-million/?sh=36a8d2b14872].

Webb, Nicholas J. (2017), *What Customers Crave: How to Create Relevant and Memorable Experiences at Every Touchpoint*, AMACOM.

Wedel, Michel and P. K. Kannan (2016), "Marketing Analytics for Data-Rich Environments," *Journal of Marketing*, 80 (6), 97–121.

Weiss, Einat (2018), "How to convince customers to share data after GDPR," *Harvard Business Review*, (accessed November 11, 2020), [available at https://hbr.org/2018/05/how-to-convince-customers-to-share-data-after-gdpr].

White, Tiffany Barnett, Debra L. Zahay, Helge Thorbjørnsen, and Sharon Shavitt (2008), "Getting too personal: Reactance to highly personalized email solicitations," *Marketing Letters*, 19 (1), 39–50.

Williams, Richard (2019), "Analyzing Proportions: Fractional Response and Zero One Inflated Beta Models."

Wu, Kuang Wen, Shaio Yan Huang, David C. Yen, and Irina Popova (2012), "The effect of online privacy policy on consumer privacy concern and trust," *Computers in Human Behavior*, 28 (3), 889–97.

Xu, Heng, Hock-Hai Teo, Bernard C. Y. Tan, and Ritu Agarwal (2012), "Research Note —Effects of Individual Self-Protection, Industry Self-Regulation, and Government Regulation on Privacy Concerns: A Study of Location-Based Services," *Information Systems Research*, 23 (4), 1342–63.

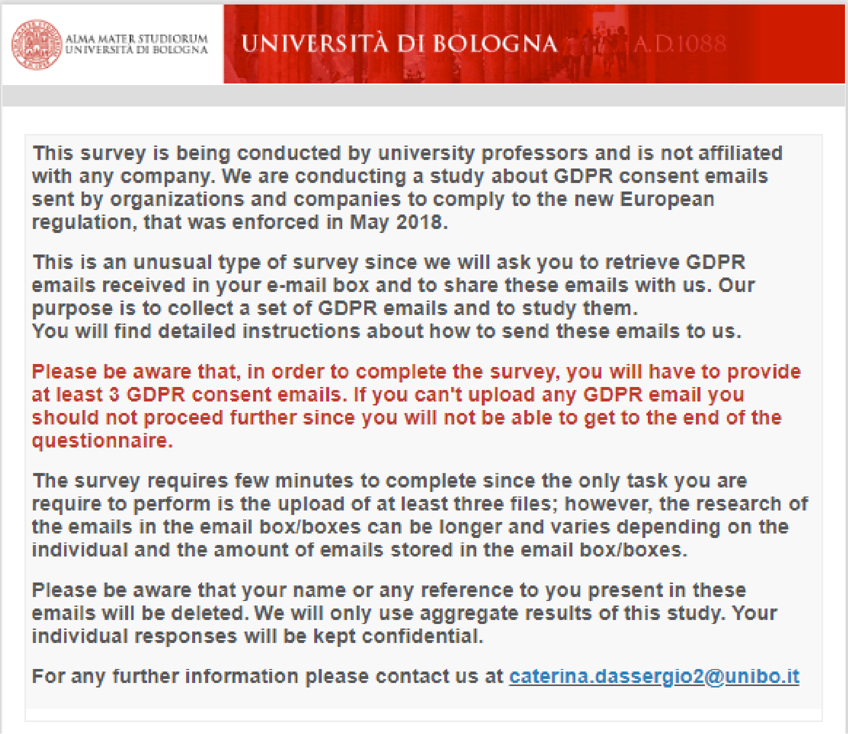
Zaharias, Chris (2010), "Why The Number One Issue In Retargeting Is Not Privacy," *AdExchanger*, (accessed November 11, 2020), [available at https://www.adexchanger.com/data-driven-thinking/counterpoint-why-the-number-one-issue-in-retargeting-is-not-privacy/#:~:text=Ms.,the targeting towards her specifically.&text=There again%2C the issue is that of excessive frequency of targeting.].

Zarum, Lara (2018), “Some Viewers Think Netflix Is Targeting Them by Race. Here’s What to Know.,” *The New York Times*, (accessed November 11, 2020), [available at https://www.nytimes.com/2018/10/23/arts/television/netflix-race-targeting-personalization.html].

APPENDIX A – PROLIFIC STUDY OVERVIEW

The study was created and distributed starting from August 8, 2019, and it runs for seven days. It was conducted in the following way. After an initial description of the study's aim – the collection of GDPR re-permission emails – respondents were required to upload at least three re-permission emails to participate in the survey and get paid for the task (Figure A.1). We then provided three examples of the type of communications we were trying to collect (Figure A.2), and we clearly offer guidance of the operations they were going to perform to be able to provide us with the emails – e.g., we suggested several keywords to retrieve these emails from the users' email boxes (Figure A.3). Respondents who were able to retrieve the emails were then required to upload the PDF with the communication and to answer some demographics questions, while those who were not willing to provide the emails or were not able to retrieve any of them in their email accounts were prevented from proceeding further in the study.

Figure A.1 – Prolific Study Introduction.



This survey is being conducted by university professors and is not affiliated with any company. We are conducting a study about GDPR consent emails sent by organizations and companies to comply to the new European regulation, that was enforced in May 2018.

This is an unusual type of survey since we will ask you to retrieve GDPR emails received in your e-mail box and to share these emails with us. Our purpose is to collect a set of GDPR emails and to study them.
You will find detailed instructions about how to send these emails to us.

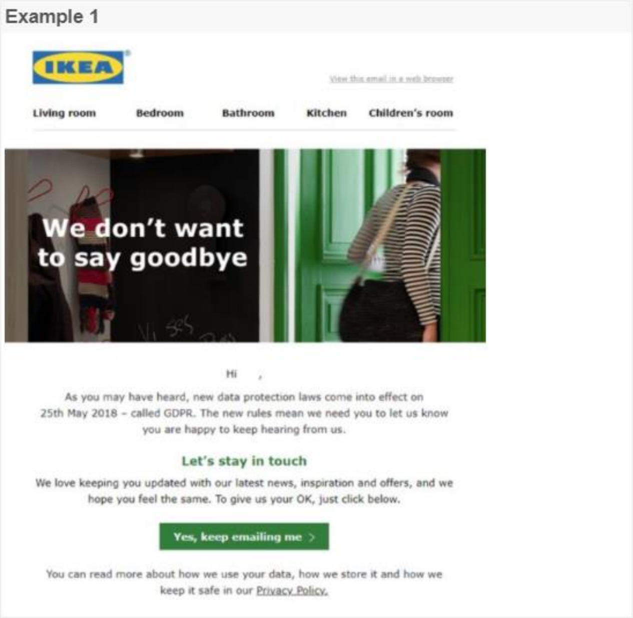
Please be aware that, in order to complete the survey, you will have to provide at least 3 GDPR consent emails. If you can't upload any GDPR email you should not proceed further since you will not be able to get to the end of the questionnaire.

The survey requires few minutes to complete since the only task you are require to perform is the upload of at least three files; however, the research of the emails in the email box/boxes can be longer and varies depending on the individual and the amount of emails stored in the email box/boxes.

Please be aware that your name or any reference to you present in these emails will be deleted. We will only use aggregate results of this study. Your individual responses will be kept confidential.

For any further information please contact us at caterina.dassergio2@unibo.it

Figure A.2 – GDPR Re-Permission Email Examples Shown to the Prolific Panel.



Example 1

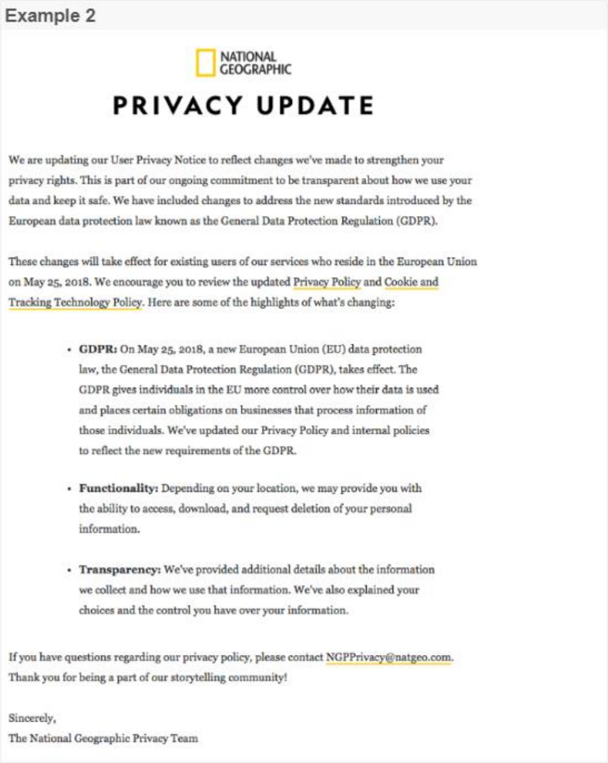
Hi ,

As you may have heard, new data protection laws come into effect on 25th May 2018 – called GDPR. The new rules mean we need you to let us know you are happy to keep hearing from us.

Let's stay in touch

We love keeping you updated with our latest news, inspiration and offers, and we hope you feel the same. To give us your OK, just click below.

You can read more about how we use your data, how we store it and how we keep it safe in our Privacy Policy.



Example 2

PRIVACY UPDATE

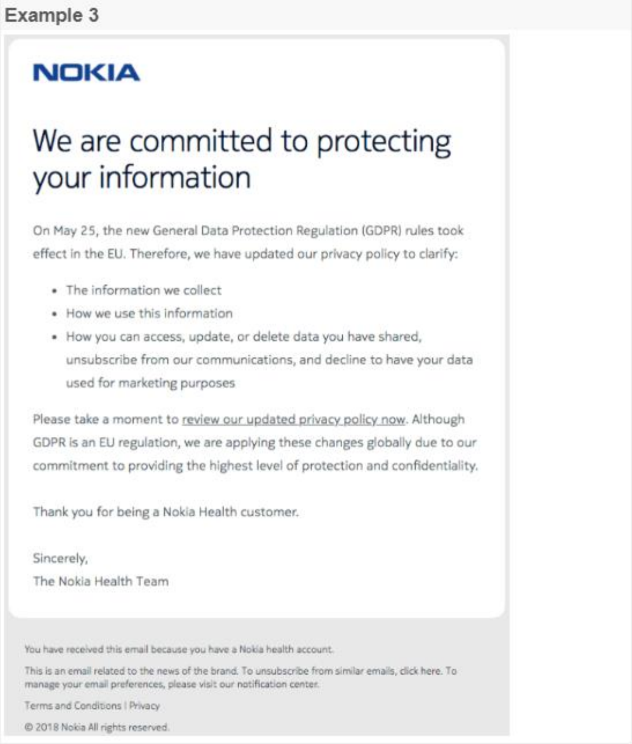
We are updating our User Privacy Notice to reflect changes we've made to strengthen your privacy rights. This is part of our ongoing commitment to be transparent about how we use your data and keep it safe. We have included changes to address the new standards introduced by the European data protection law known as the General Data Protection Regulation (GDPR).

These changes will take effect for existing users of our services who reside in the European Union on May 25, 2018. We encourage you to review the updated Privacy Policy and Cookie and Tracking Technology Policy. Here are some of the highlights of what's changing:

- **GDPR:** On May 25, 2018, a new European Union (EU) data protection law, the General Data Protection Regulation (GDPR), takes effect. The GDPR gives individuals in the EU more control over how their data is used and places certain obligations on businesses that process information of those individuals. We've updated our Privacy Policy and internal policies to reflect the new requirements of the GDPR.
- **Functionality:** Depending on your location, we may provide you with the ability to access, download, and request deletion of your personal information.
- **Transparency:** We've provided additional details about the information we collect and how we use that information. We've also explained your choices and the control you have over your information.

If you have questions regarding our privacy policy, please contact NGPPrivacy@natgeo.com. Thank you for being a part of our storytelling community!

Sincerely,
The National Geographic Privacy Team



Example 3

We are committed to protecting your information

On May 25, the new General Data Protection Regulation (GDPR) rules took effect in the EU. Therefore, we have updated our privacy policy to clarify:

- The information we collect
- How we use this information
- How you can access, update, or delete data you have shared, unsubscribe from our communications, and decline to have your data used for marketing purposes

Please take a moment to review our updated privacy policy now. Although GDPR is an EU regulation, we are applying these changes globally due to our commitment to providing the highest level of protection and confidentiality.

Thank you for being a Nokia Health customer.

Sincerely,
The Nokia Health Team

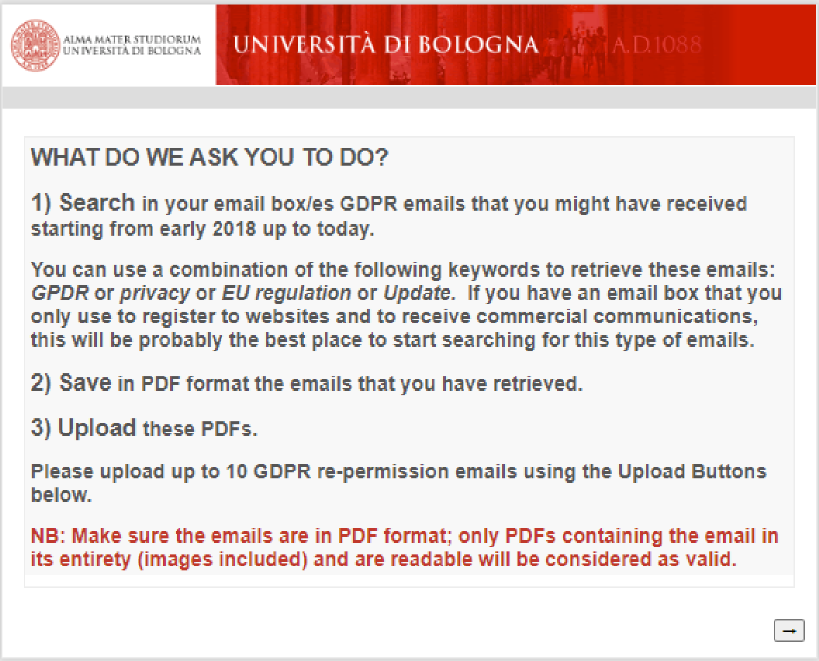
You have received this email because you have a Nokia health account.

This is an email related to the news of the brand. To unsubscribe from similar emails, click here. To manage your email preferences, please visit our notification center.

Terms and Conditions | Privacy

© 2018 Nokia All rights reserved.

Figure A.3 – Prolific Study Guidelines Provided.



WHAT DO WE ASK YOU TO DO?

1) Search in your email box/es GDPR emails that you might have received starting from early 2018 up to today.

You can use a combination of the following keywords to retrieve these emails: *GPDR* or *privacy* or *EU regulation* or *Update*. If you have an email box that you only use to register to websites and to receive commercial communications, this will be probably the best place to start searching for this type of emails.

2) Save in PDF format the emails that you have retrieved.

3) Upload these PDFs.

Please upload up to 10 GDPR re-permission emails using the Upload Buttons below.

NB: Make sure the emails are in PDF format; only PDFs containing the email in its entirety (images included) and are readable will be considered as valid.

We collected answers from 465 respondents who were evenly split between males (53.6%) and females (45.4%). They were mainly young adults – between 18 and 34 years old – highly educated, with full or part-time jobs, and were from United Kingdom (23.7%), Poland (16.6%), and Portugal (13.6%).

They uploaded a total of 2016 files in our online survey. However, after an initial check, we found that 54% of the emails collected was not usable, given that we had respondents who uploaded:

- Duplicates
- Non-GDPR related emails
- Blurred or not readable emails

Consequently, we were able to retrieve and use only 931 unique GDPR re-permission emails from the study.

APPENDIX B – CORROBORATION CHECKS ON SEMRUSH DATA

In order to corroborate the reliability of our primary data source for the website's online performance metrics, we conducted other tests.

First, we looked at some of the companies in the news sector, and we compared their financial statement figures with the expected online advertising revenues estimated by SEMrush. We selected this industry since its main revenue source is from online advertising, and they mainly use Google Ad Network to find advertisements to place on their websites (Libert and Nielsen 2018). This, in principle, helped us to find comparable data. We extracted firms expected online ad revenues in June 2020, and we multiply the resulting estimate by 12 to get the annual total estimate. We reported the results in Table B.1 below. As it is possible to see, the estimates are very close to the real numbers disclosed by the companies at the end of 2019. There are still differences in the figures that may be ascribed to the presence of advertising exchanged on ad networks other than Google or to the fact that the calculation used to get the yearly expected online ad revenue have been performed on a month (June 2020), which may be not representative for the real average of the year.

Table B.1 – Comparison between SEMrush Yearly Expected Online Ad Revenue (Jun 2020) and the Ad Revenue declared by Companies in their Financial Statements (Dec 2019).

Company	SEMrush Monthly Estimate	SEMrush Yearly Estimate	Balance Sheet	Source
The Economist	$2,165,623	$25,987,476	$27,000,000	https://www.economistgroup.com/pdfs/Annual_Report_2019.pdf
The New York Times	$37,484,597	$449,815,164	$530,680,000	https://www.statista.com/statistics/192907/advertising-revenue-of-the-new-york-times-company-since-2006/
Il sole 24 Ore	$3,567,906	$42,814,872	$45,488,000	https://www.gruppo24ore.ilsole24ore.com/it-it/investors/bilanci-e-relazioni/#2019

Second, we compared the traffic data on a randomly chosen subsample of 80 companies. This was done to establish that the data produced by SEMrush are in line with other online marketing platform leaders in the sector, such as SimilarWeb. We collected the global number of sessions – both mobile and desktop – for the month of May 2020 using both SimilarWeb and SEMrush. As it is possible to establish from the correlation table below (Table B.2), there is a high correlation between these data meaning that SEMrush produces figures comparable to the ones provided by SimilarWeb. The existence of some small differences may be attributed to the different panel of users used by the two platforms to get the number of sessions on the websites.

Table B.2 – Correlation Table between the Number of Sessions Declared by SEMrush and those Declared by SimilarWeb (# Companies = 80).

	# Sessions SEMrush (May 2020)	# Sessions SimilarWeb (May 2020)
# Sessions SEMrush (May 2020)	1	
# Sessions SimilarWeb (May 2020)	0.79	1

With regards to the possible differences existing between the different data sources, it is also important to highlight that SEMrush "*do not have any access to anyone's internal Google Ads or Google Analytics data"* and can only make estimates based *"on the keywords that they have in their database, which contains millions of keywords and the top 7 positions for Google Ads*"[15].

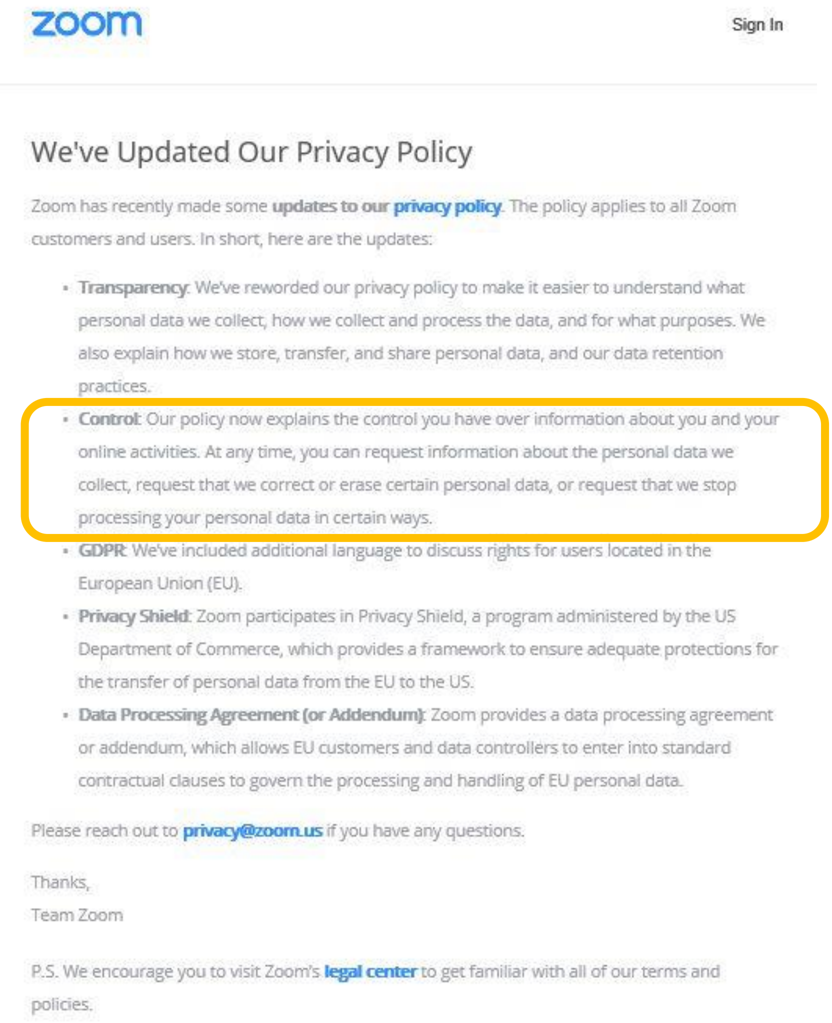
[15] https://www.SemRush.com/kb/508-ad-research-positions-report

Appendix C – Examples of the Manual Coding of the Re-Permission Emails

For clarity, we included some examples of coding of re-permission emails for the different themes considered in the protocol. Two independent coders manually content analyzed the emails. For each theme, we provide an example of email and the relative measure.

Informative: Control

This is a dummy variable indicating if, in the email, the possibility of controlling personal information was explicitly and clearly mentioned.



We've Updated Our Privacy Policy

Zoom has recently made some **updates to our privacy policy**. The policy applies to all Zoom customers and users. In short, here are the updates:

- **Transparency**: We've reworded our privacy policy to make it easier to understand what personal data we collect, how we collect and process the data, and for what purposes. We also explain how we store, transfer, and share personal data, and our data retention practices.
- **Control**: Our policy now explains the control you have over information about you and your online activities. At any time, you can request information about the personal data we collect, request that we correct or erase certain personal data, or request that we stop processing your personal data in certain ways.
- **GDPR**: We've included additional language to discuss rights for users located in the European Union (EU).
- **Privacy Shield**: Zoom participates in Privacy Shield, a program administered by the US Department of Commerce, which provides a framework to ensure adequate protections for the transfer of personal data from the EU to the US.
- **Data Processing Agreement (or Addendum)**: Zoom provides a data processing agreement or addendum, which allows EU customers and data controllers to enter into standard contractual clauses to govern the processing and handling of EU personal data.

Please reach out to **privacy@zoom.us** if you have any questions.

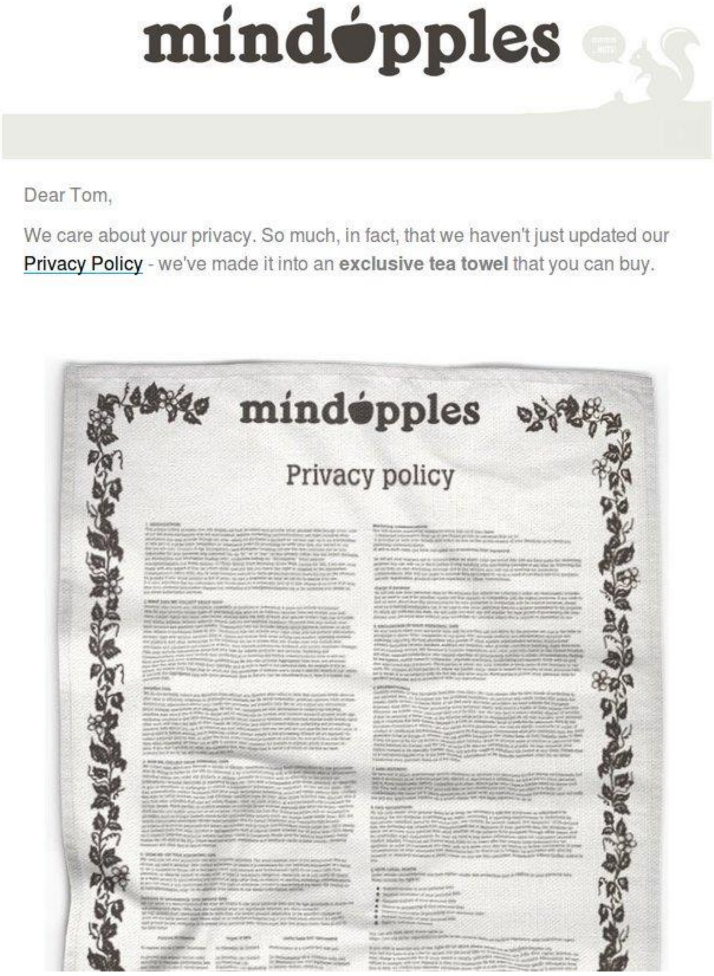
Thanks,
Team Zoom

P.S. We encourage you to visit Zoom's **legal center** to get familiar with all of our terms and policies.

Informative: Transparency

Transparency was coded using five levels to better capture this theme's nature, ranging from communications that are not focusing on this dimension to communications that provide precise details and specific information about the news of the GDPR and the rights customer have according to the new principles of data protection.

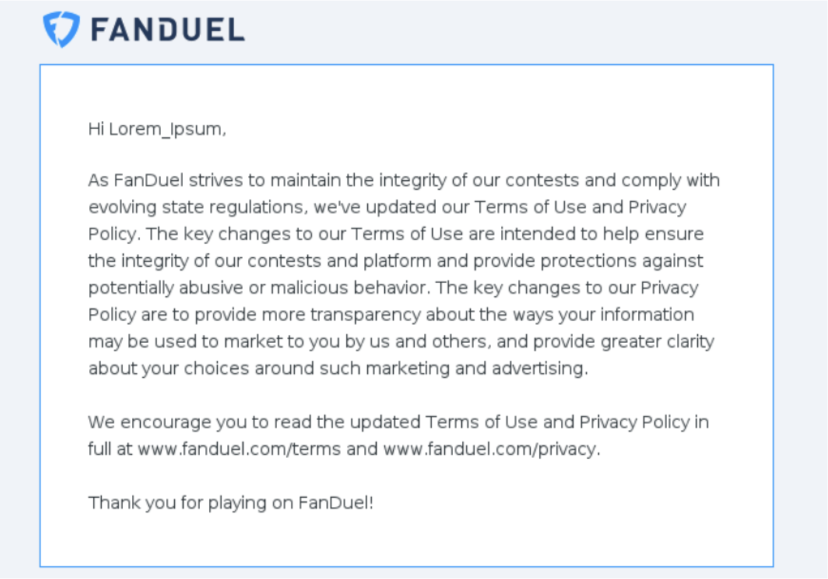
Level 1: Minimum Level of Transparency



Dear Tom,

We care about your privacy. So much, in fact, that we haven't just updated our Privacy Policy - we've made it into an **exclusive tea towel** that you can buy.

Level 2: Low Level of Transparency



FANDUEL

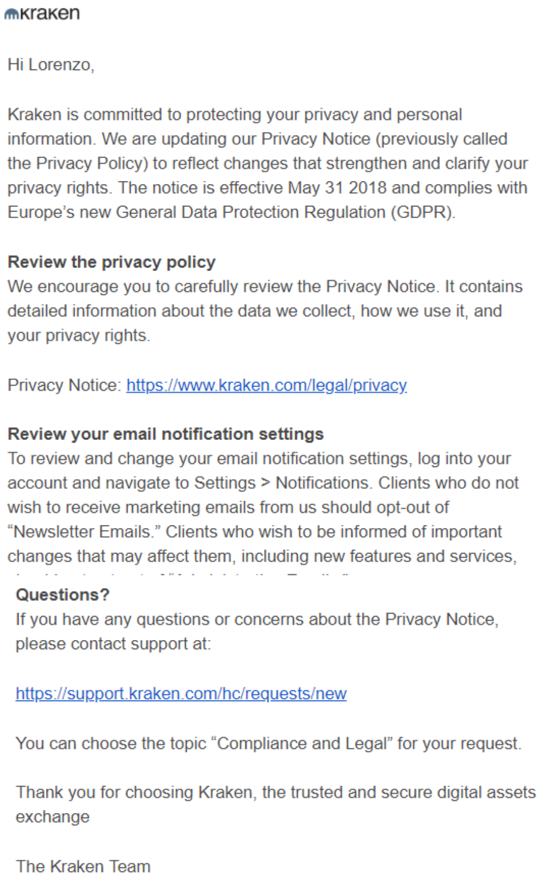
Hi Lorem_Ipsum,

As FanDuel strives to maintain the integrity of our contests and comply with evolving state regulations, we've updated our Terms of Use and Privacy Policy. The key changes to our Terms of Use are intended to help ensure the integrity of our contests and platform and provide protections against potentially abusive or malicious behavior. The key changes to our Privacy Policy are to provide more transparency about the ways your information may be used to market to you by us and others, and provide greater clarity about your choices around such marketing and advertising.

We encourage you to read the updated Terms of Use and Privacy Policy in full at www.fanduel.com/terms and www.fanduel.com/privacy.

Thank you for playing on FanDuel!

Level 3: Average Level of Transparency



kraken

Hi Lorenzo,

Kraken is committed to protecting your privacy and personal information. We are updating our Privacy Notice (previously called the Privacy Policy) to reflect changes that strengthen and clarify your privacy rights. The notice is effective May 31 2018 and complies with Europe’s new General Data Protection Regulation (GDPR).

Review the privacy policy
We encourage you to carefully review the Privacy Notice. It contains detailed information about the data we collect, how we use it, and your privacy rights.

Privacy Notice: https://www.kraken.com/legal/privacy

Review your email notification settings
To review and change your email notification settings, log into your account and navigate to Settings > Notifications. Clients who do not wish to receive marketing emails from us should opt-out of “Newsletter Emails.” Clients who wish to be informed of important changes that may affect them, including new features and services,

Questions?
If you have any questions or concerns about the Privacy Notice, please contact support at:

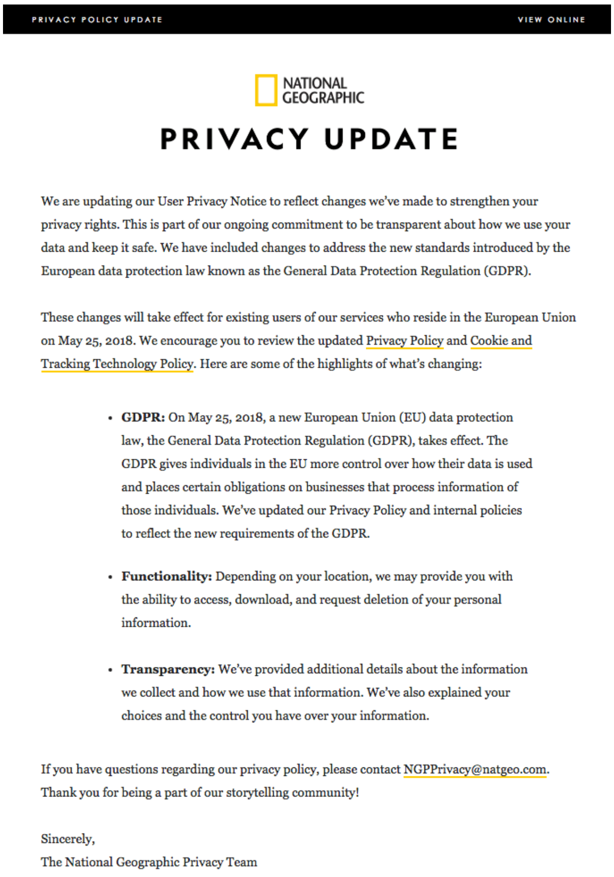
https://support.kraken.com/hc/requests/new

You can choose the topic “Compliance and Legal” for your request.

Thank you for choosing Kraken, the trusted and secure digital assets exchange

The Kraken Team

Level 4: High Level of Transparency



PRIVACY UPDATE

We are updating our User Privacy Notice to reflect changes we've made to strengthen your privacy rights. This is part of our ongoing commitment to be transparent about how we use your data and keep it safe. We have included changes to address the new standards introduced by the European data protection law known as the General Data Protection Regulation (GDPR).

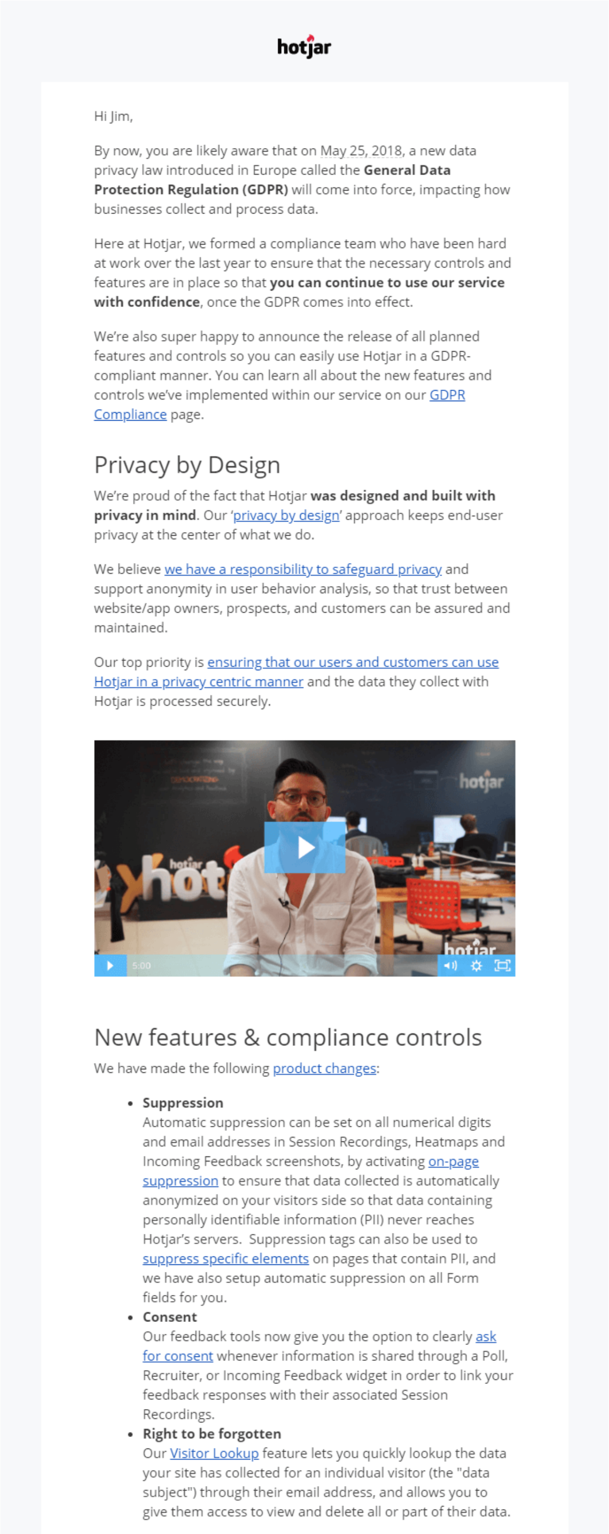
These changes will take effect for existing users of our services who reside in the European Union on May 25, 2018. We encourage you to review the updated Privacy Policy and Cookie and Tracking Technology Policy. Here are some of the highlights of what's changing:

- **GDPR:** On May 25, 2018, a new European Union (EU) data protection law, the General Data Protection Regulation (GDPR), takes effect. The GDPR gives individuals in the EU more control over how their data is used and places certain obligations on businesses that process information of those individuals. We've updated our Privacy Policy and internal policies to reflect the new requirements of the GDPR.
- **Functionality:** Depending on your location, we may provide you with the ability to access, download, and request deletion of your personal information.
- **Transparency:** We've provided additional details about the information we collect and how we use that information. We've also explained your choices and the control you have over your information.

If you have questions regarding our privacy policy, please contact NGPPrivacy@natgeo.com. Thank you for being a part of our storytelling community!

Sincerely,
The National Geographic Privacy Team

Level 5: Very High Level of Transparency



hotjar

Hi Jim,

By now, you are likely aware that on May 25, 2018, a new data privacy law introduced in Europe called the **General Data Protection Regulation (GDPR)** will come into force, impacting how businesses collect and process data.

Here at Hotjar, we formed a compliance team who have been hard at work over the last year to ensure that the necessary controls and features are in place so that **you can continue to use our service with confidence**, once the GDPR comes into effect.

We're also super happy to announce the release of all planned features and controls so you can easily use Hotjar in a GDPR-compliant manner. You can learn all about the new features and controls we've implemented within our service on our GDPR Compliance page.

Privacy by Design

We're proud of the fact that Hotjar **was designed and built with privacy in mind**. Our 'privacy by design' approach keeps end-user privacy at the center of what we do.

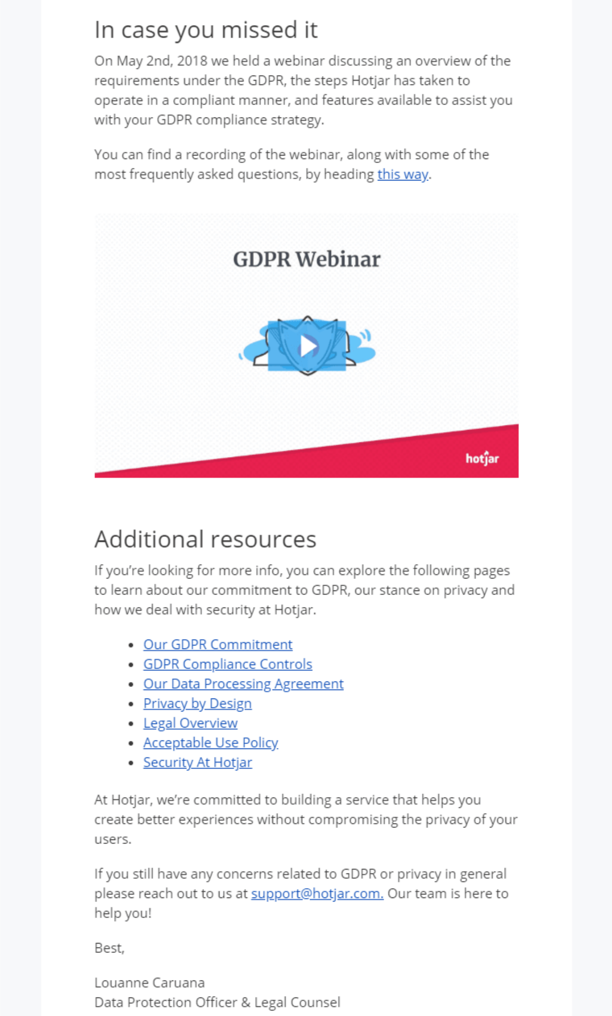
We believe we have a responsibility to safeguard privacy and support anonymity in user behavior analysis, so that trust between website/app owners, prospects, and customers can be assured and maintained.

Our top priority is ensuring that our users and customers can use Hotjar in a privacy centric manner and the data they collect with Hotjar is processed securely.

New features & compliance controls

We have made the following product changes:

- **Suppression**
  Automatic suppression can be set on all numerical digits and email addresses in Session Recordings, Heatmaps and Incoming Feedback screenshots, by activating on-page suppression to ensure that data collected is automatically anonymized on your visitors side so that data containing personally identifiable information (PII) never reaches Hotjar's servers. Suppression tags can also be used to suppress specific elements on pages that contain PII, and we have also setup automatic suppression on all Form fields for you.
- **Consent**
  Our feedback tools now give you the option to clearly ask for consent whenever information is shared through a Poll, Recruiter, or Incoming Feedback widget in order to link your feedback responses with their associated Session Recordings.
- **Right to be forgotten**
  Our Visitor Lookup feature lets you quickly lookup the data your site has collected for an individual visitor (the "data subject") through their email address, and allows you to give them access to view and delete all or part of their data.



In case you missed it

On May 2nd, 2018 we held a webinar discussing an overview of the requirements under the GDPR, the steps Hotjar has taken to operate in a compliant manner, and features available to assist you with your GDPR compliance strategy.

You can find a recording of the webinar, along with some of the most frequently asked questions, by heading this way.

Additional resources

If you're looking for more info, you can explore the following pages to learn about our commitment to GDPR, our stance on privacy and how we deal with security at Hotjar.

- Our GDPR Commitment
- GDPR Compliance Controls
- Our Data Processing Agreement
- Privacy by Design
- Legal Overview
- Acceptable Use Policy
- Security At Hotjar

At Hotjar, we're committed to building a service that helps you create better experiences without compromising the privacy of your users.

If you still have any concerns related to GDPR or privacy in general please reach out to us at support@hotjar.com. Our team is here to help you!

Best,

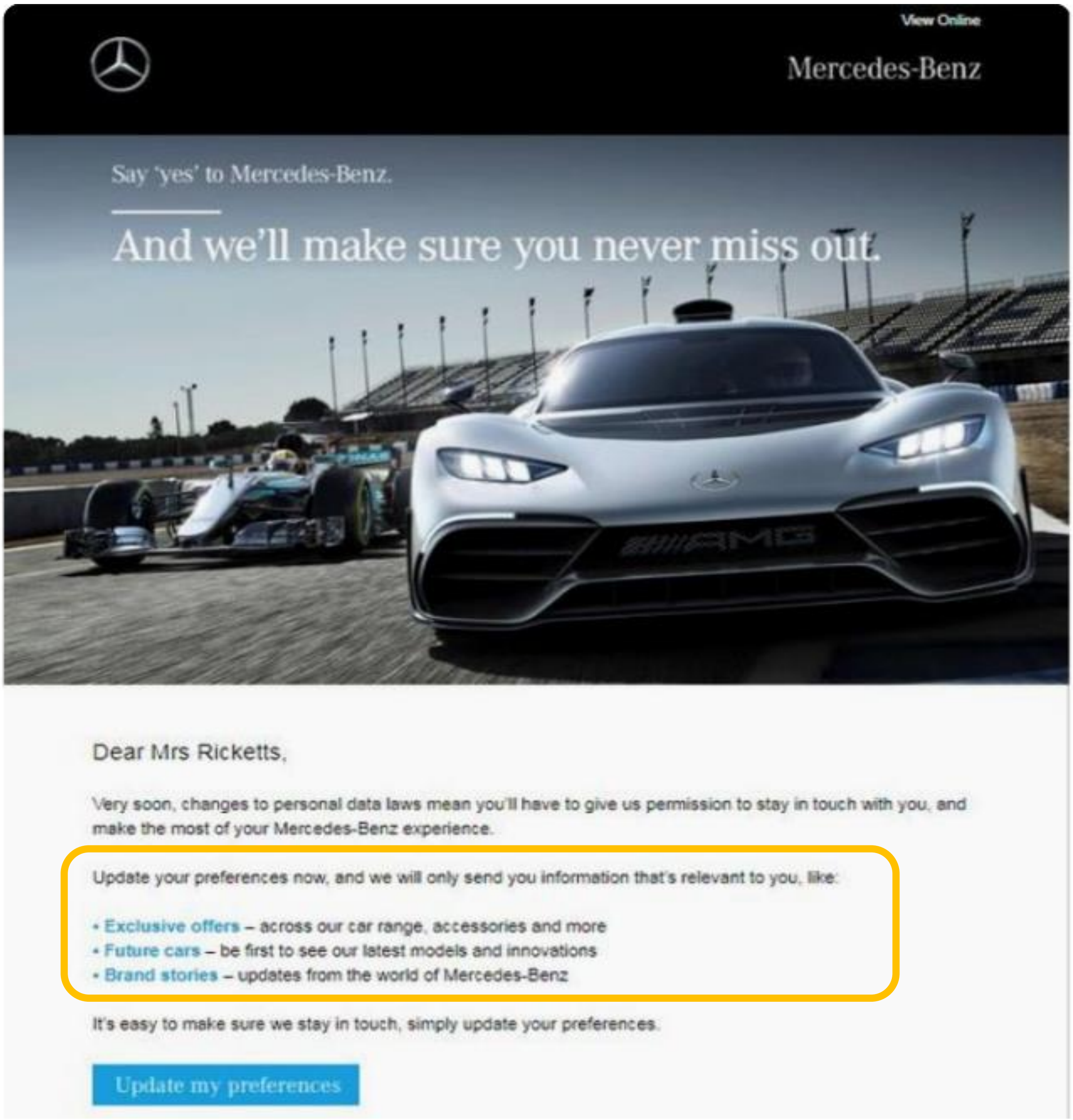
Louanne Caruana
Data Protection Officer & Legal Counsel

Framing: Gain & Loss

This is a dummy variable indicating the presence of a gain frame, a loss frame, or both.

Gain Frame

These communications may be characterized by a gain-type of framing such as: "*if you continue giving us access to your data, we'll be able to send you personalized offers...*".



View Online

Mercedes-Benz

Say 'yes' to Mercedes-Benz.

And we'll make sure you never miss out.

Dear Mrs Ricketts,

Very soon, changes to personal data laws mean you'll have to give us permission to stay in touch with you, and make the most of your Mercedes-Benz experience.

Update your preferences now, and we will only send you information that's relevant to you, like:

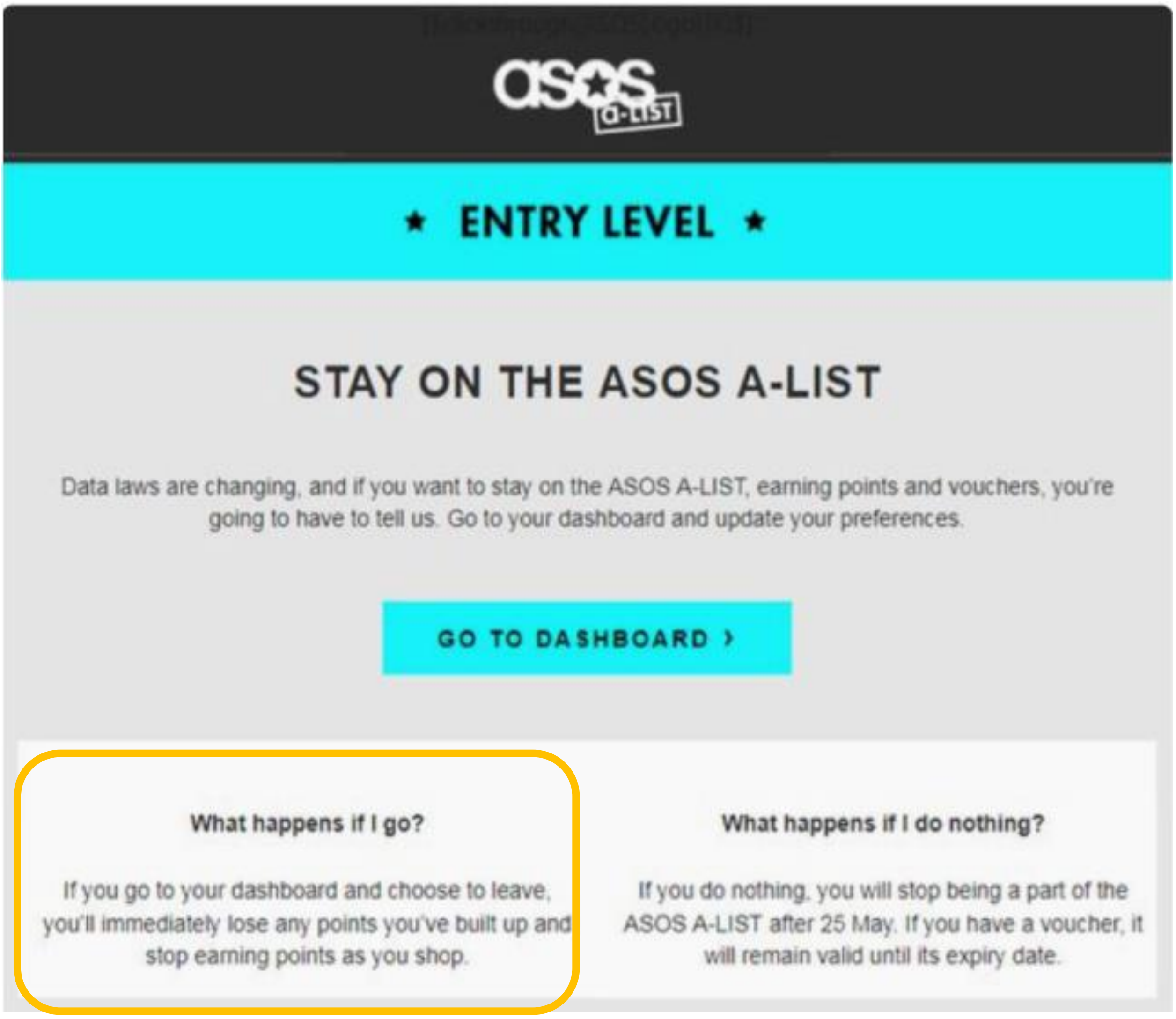
- Exclusive offers – across our car range, accessories and more
- Future cars – be first to see our latest models and innovations
- Brand stories – updates from the world of Mercedes-Benz

It's easy to make sure we stay in touch, simply update your preferences.

Update my preferences

Loss Frame

These communications may be characterized by a loss-type of framing such as: “*if you do not continue giving us access to your data, you will miss out on special offers...*”.



★ ENTRY LEVEL ★

STAY ON THE ASOS A-LIST

Data laws are changing, and if you want to stay on the ASOS A-LIST, earning points and vouchers, you're going to have to tell us. Go to your dashboard and update your preferences.

What happens if I go?

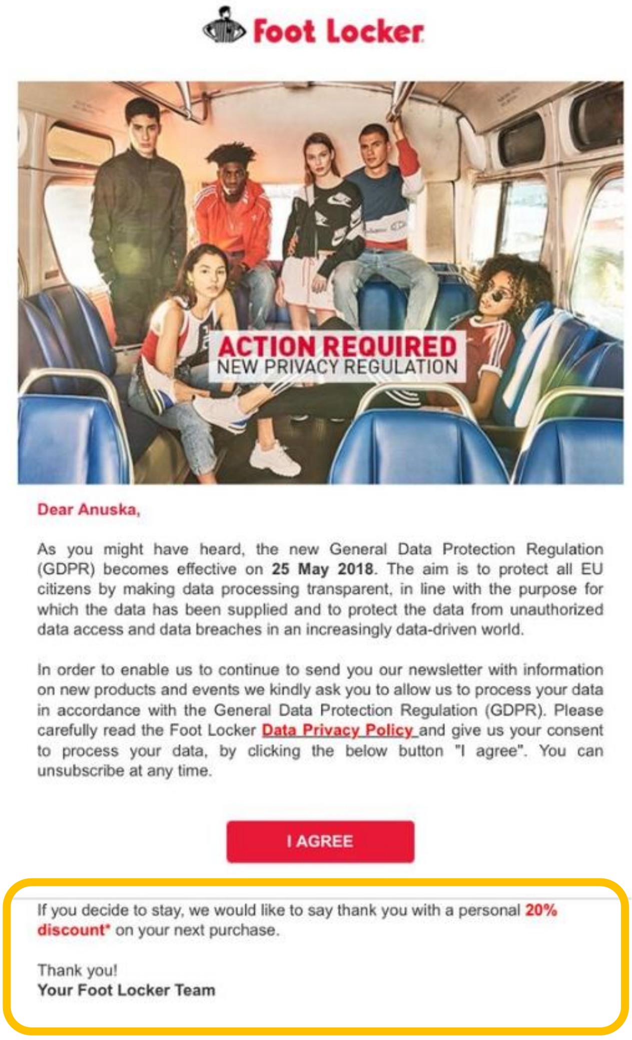
If you go to your dashboard and choose to leave, you'll immediately lose any points you've built up and stop earning points as you shop.

What happens if I do nothing?

If you do nothing, you will stop being a part of the ASOS A-LIST after 25 May. If you have a voucher, it will remain valid until its expiry date.

Monetary Incentives

This is a dummy variable indicating the presence of monetary incentives in the email (e.g., discounts, coupons).



Foot Locker

ACTION REQUIRED
NEW PRIVACY REGULATION

Dear Anuska,

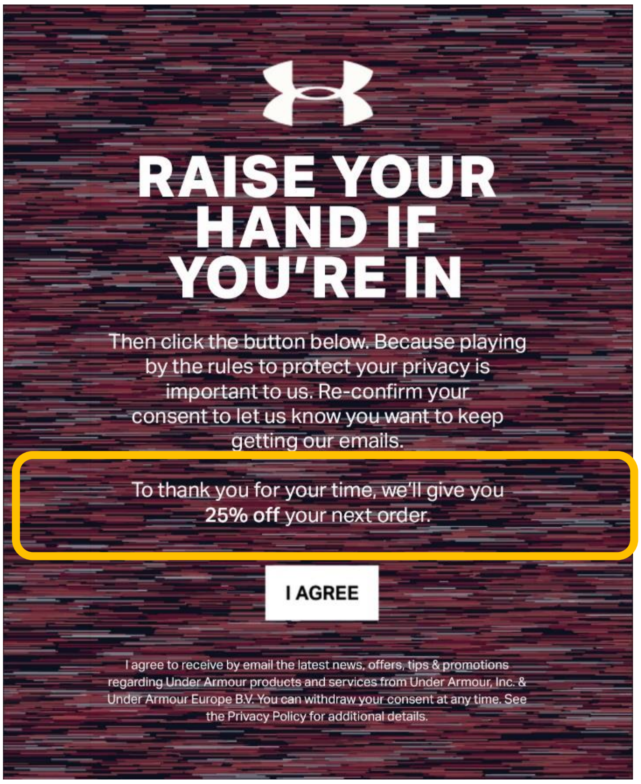
As you might have heard, the new General Data Protection Regulation (GDPR) becomes effective on **25 May 2018**. The aim is to protect all EU citizens by making data processing transparent, in line with the purpose for which the data has been supplied and to protect the data from unauthorized data access and data breaches in an increasingly data-driven world.

In order to enable us to continue to send you our newsletter with information on new products and events we kindly ask you to allow us to process your data in accordance with the General Data Protection Regulation (GDPR). Please carefully read the Foot Locker **Data Privacy Policy** and give us your consent to process your data, by clicking the below button "I agree". You can unsubscribe at any time.

I AGREE

If you decide to stay, we would like to say thank you with a personal **20% discount*** on your next purchase.

Thank you!
Your Foot Locker Team



RAISE YOUR HAND IF YOU'RE IN

Then click the button below. Because playing by the rules to protect your privacy is important to us. Re-confirm your consent to let us know you want to keep getting our emails.

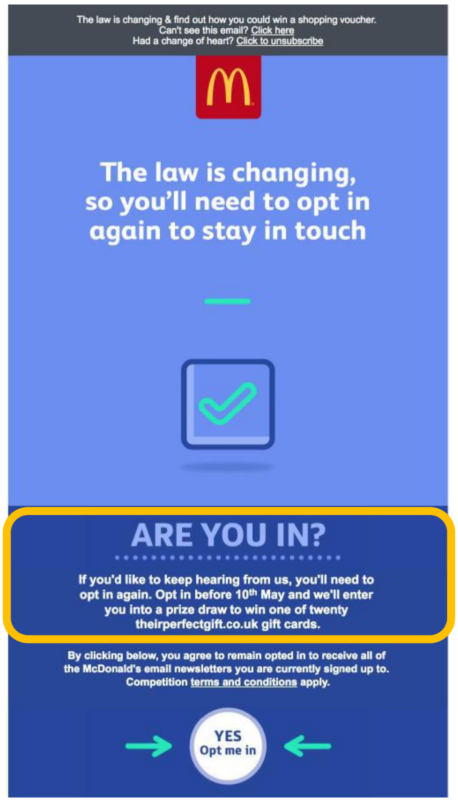
To thank you for your time, we'll give you **25% off** your next order.

I AGREE

I agree to receive by email the latest news, offers, tips & promotions regarding Under Armour products and services from Under Armour, Inc. & Under Armour Europe B.V. You can withdraw your consent at any time. See the Privacy Policy for additional details.

Non-Monetary Incentives

This is a dummy variable indicating the presence of non-monetary incentives in the email (e.g., lottery, free trials, invitations to events).



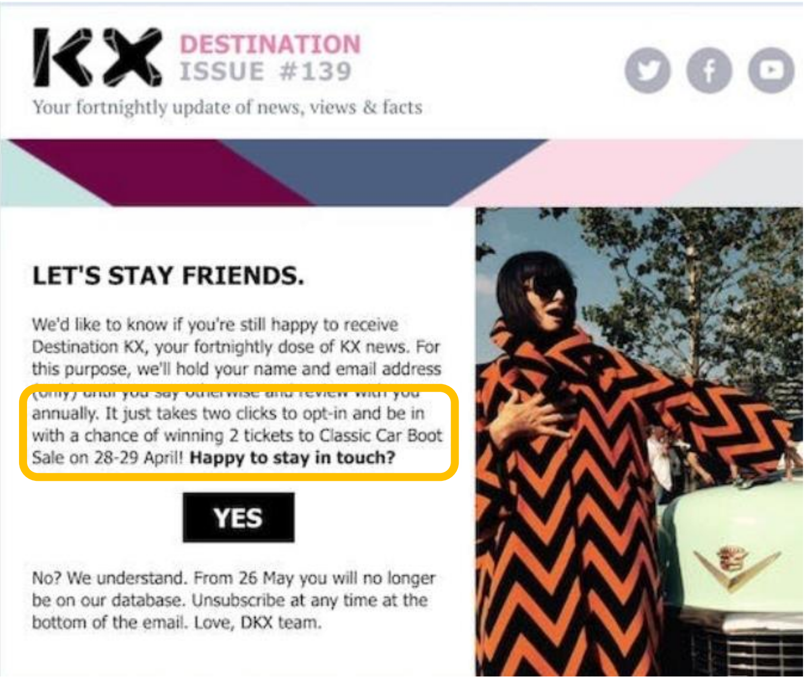
The law is changing & find out how you could win a shopping voucher.
Can't see this email? Click here
Had a change of heart? Click to unsubscribe

The law is changing, so you'll need to opt in again to stay in touch

ARE YOU IN?

If you'd like to keep hearing from us, you'll need to opt in again. Opt in before 10th May and we'll enter you into a prize draw to win one of twenty theirperfectgift.co.uk gift cards.

By clicking below, you agree to remain opted in to receive all of the McDonald's email newsletters you are currently signed up to. Competition terms and conditions apply.



DESTINATION
ISSUE #139
Your fortnightly update of news, views & facts

LET'S STAY FRIENDS.

We'd like to know if you're still happy to receive Destination KX, your fortnightly dose of KX news. For this purpose, we'll hold your name and email address (only) until you say otherwise and review with you annually. It just takes two clicks to opt-in and be in with a chance of winning 2 tickets to Classic Car Boot Sale on 28-29 April! **Happy to stay in touch?**

No? We understand. From 26 May you will no longer be on our database. Unsubscribe at any time at the bottom of the email. Love, DKX team.

APPENDIX D – ESTIMATED CODED VARIABLES

As already stated in Section 6.2, we recruit two independent judges for manually coding 20% of the total sample of re-permission emails collected. We asked them to code them accordingly to a predefined protocol on the following variables:

- Level of Transparency
- Presence of the mention of Control that customers have on their data
- Presence of Monetary Incentives
- Presence of Non-Monetary Incentives
- Usage of a Gain-type of Frame
- Usage of a Loss-type of Frame
- Presence of References to the Past or the Future

In order to estimate the manually coded variables on the whole sample of emails (N = 1506), we decided to try to see if the variables available from LIWC and TextEvaluator were able to predict the value for the variables listed above correctly.

We divided the sub-sample of the manually coded variables into two parts: in-sample (232 observations, which is 75% of the total observations) and out-sample (76 observations, which is 25% of the total observations). This has been done to see whether the predicted values are in line with the actual values for the variables of interest; this will be assessed through lift charts and ROC curves.

Following, it is possible to find the results of the predictive models estimated (Table D.1) as well as the lift charts for the in- and out-sample observations (Figure D.1) and the ROC curves (Figure D.2). Additionally, we also present, for each dummy variable, the Youden-Index plot and the cutoff used to dichotomize the variables (Figure D.3).

Table D.1 - Results from the Estimation Procedure (N = 232; 75% of the 308 manually coded variables).

	Variable	Control			Transparency			Monetary Incentives			Non-Monetary Incentives			Framing		
LIWC	wc	0.013	***	(0.003)	0.038	***	(0.004)	0.000		(0.001)	0.000		(0.001)	0.000		(0.001)
	analytic	-0.010		(0.015)	-0.016		(0.012)	-0.013		(0.014)	-0.006		(0.014)	0.036	**	(0.016)
	clout	0.239	**	(0.104)	0.105	**	(0.049)	0.071		(0.066)	0.104		(0.09)	-0.104	**	(0.045)
	authentic	-0.005		(0.021)	0.013		(0.016)	0.026		(0.019)	-0.022		(0.018)	-0.035	*	(0.02)
	tone	-0.006		(0.019)	-0.021		(0.015)	-0.040	**	(0.019)	-0.012		(0.018)	0.008		(0.021)
	wps	0.011		(0.038)	0.002		(0.025)	0.065	**	(0.032)	0.059	*	(0.032)	0.000		(0.032)
	function	-0.047		(0.072)	-0.097	*	(0.056)	0.069		(0.067)	0.060		(0.068)	0.055		(0.074)
	posemo	0.066		(0.192)	-0.027		(0.153)	0.273		(0.189)	0.121		(0.189)	-0.242		(0.224)
	negemo	-0.461		(0.433)	-0.090		(0.325)	-0.910	**	(0.384)	-0.366		(0.367)	0.512		(0.412)
	Social	-0.042		(0.072)	-0.050		(0.054)	0.077		(0.063)	0.022		(0.063)	0.331	***	(0.077)
	cogproc	0.107		(0.083)	0.095		(0.063)	-0.071		(0.073)	0.027		(0.073)	0.071		(0.081)
	percept	-0.088		(0.185)	-0.181		(0.156)	0.248		(0.177)	0.299	*	(0.177)	0.283		(0.179)
	affiliation	-0.166	*	(0.094)	0.128	*	(0.074)	-0.083		(0.089)	-0.044		(0.09)	-0.195	**	(0.095)
	achieve	0.063		(0.215)	0.009		(0.168)	0.405	*	(0.21)	0.027		(0.211)	0.059		(0.22)
	power	0.138		(0.13)	-0.026		(0.1)	0.130		(0.121)	-0.087		(0.122)	-0.157		(0.128)
	reward	-0.552	**	(0.252)	0.252		(0.19)	0.143		(0.241)	-0.306		(0.24)	-0.180		(0.248)
	risk	0.010		(0.255)	-0.089		(0.194)	0.618	**	(0.244)	0.070		(0.225)	0.161		(0.237)
	relativ	-0.163		(0.117)	-0.098		(0.093)	0.003		(0.106)	0.259	**	(0.113)	0.350	***	(0.125)
	work	-0.377	***	(0.115)	0.029		(0.083)	-0.145		(0.108)	-0.184		(0.113)	-0.158		(0.118)
	leisure	0.202		(0.2)	0.041		(0.173)	0.209		(0.21)	-0.022		(0.212)	0.186		(0.236)
	home	-0.259		(0.593)	0.431		(0.388)	0.546		(0.426)	-0.276		(0.455)	-0.004		(0.461)
	money	-0.063		(0.188)	-0.042		(0.141)	0.468	***	(0.178)	0.135		(0.172)	0.140		(0.192)

	death	-1.366	(3.913)	0.276	(3.134)	0.206	(3.902)	1.955	(3.099)	2.783	(3.215)
	informal	-0.166	(0.28)	-0.168	(0.205)	0.144	(0.238)	-0.240	(0.31)	-0.534 *	(0.305)
TextEvaluator	Academic Vocabulary	0.039	(0.118)	0.040	(0.054)	-0.107	(0.111)	-0.065	(0.089)	-0.011	(0.049)
	Concreteness	-0.026	(0.145)	0.012	(0.063)	0.094	(0.126)	0.109	(0.105)	0.080	(0.066)
	Degree of Narrativity	-0.019	(0.038)	-0.041	(0.03)	-0.067 **	(0.032)	0.011	(0.034)	-0.056	(0.036)
	Interactive/ Conversational Style	0.004	(0.044)	0.009	(0.021)	0.008	(0.04)	0.019	(0.033)	0.015	(0.022)
	Level of Argumentation	-0.012	(0.04)	-0.024	(0.018)	-0.031	(0.036)	-0.011	(0.029)	-0.010	(0.018)
	Lexical Cohesion	0.045	(0.045)	0.004	(0.028)	0.008	(0.042)	-0.008	(0.038)	0.046	(0.036)
	Syntactic Complexity	-0.049	(0.177)	-0.117	(0.078)	-0.075	(0.159)	-0.087	(0.13)	-0.107	(0.079)
	TextEvaluator Complexity Score	0.005	(0.024)	0.010	(0.01)	0.011	(0.021)	0.013	(0.017)	0.009	(0.01)
	Word Unfamiliarity	-0.121	(0.134)	-0.102 *	(0.061)	0.014	(0.121)	-0.010	(0.099)	0.000	(0.06)

Model	Logit	Ordered Logit	Logit	Logit	Logit
# Observations:	232	232	232	232	232
Log likelihood:	-95	-177	-104	-105	-94
Pseudo R2:	0.411	0.495	0.278	0.266	0.384

* p < 0.1, ** p < 0.05, *** p < 0.01

Notes: We use 75% of the manual coded sample (N = 232) to predict the expected value for the dependent variable considered. Standard errors are in parentheses.

Figure D.1 – Lift-Charts to Evaluate the Predictive Ability of the Estimated Models (In & Out Sample).

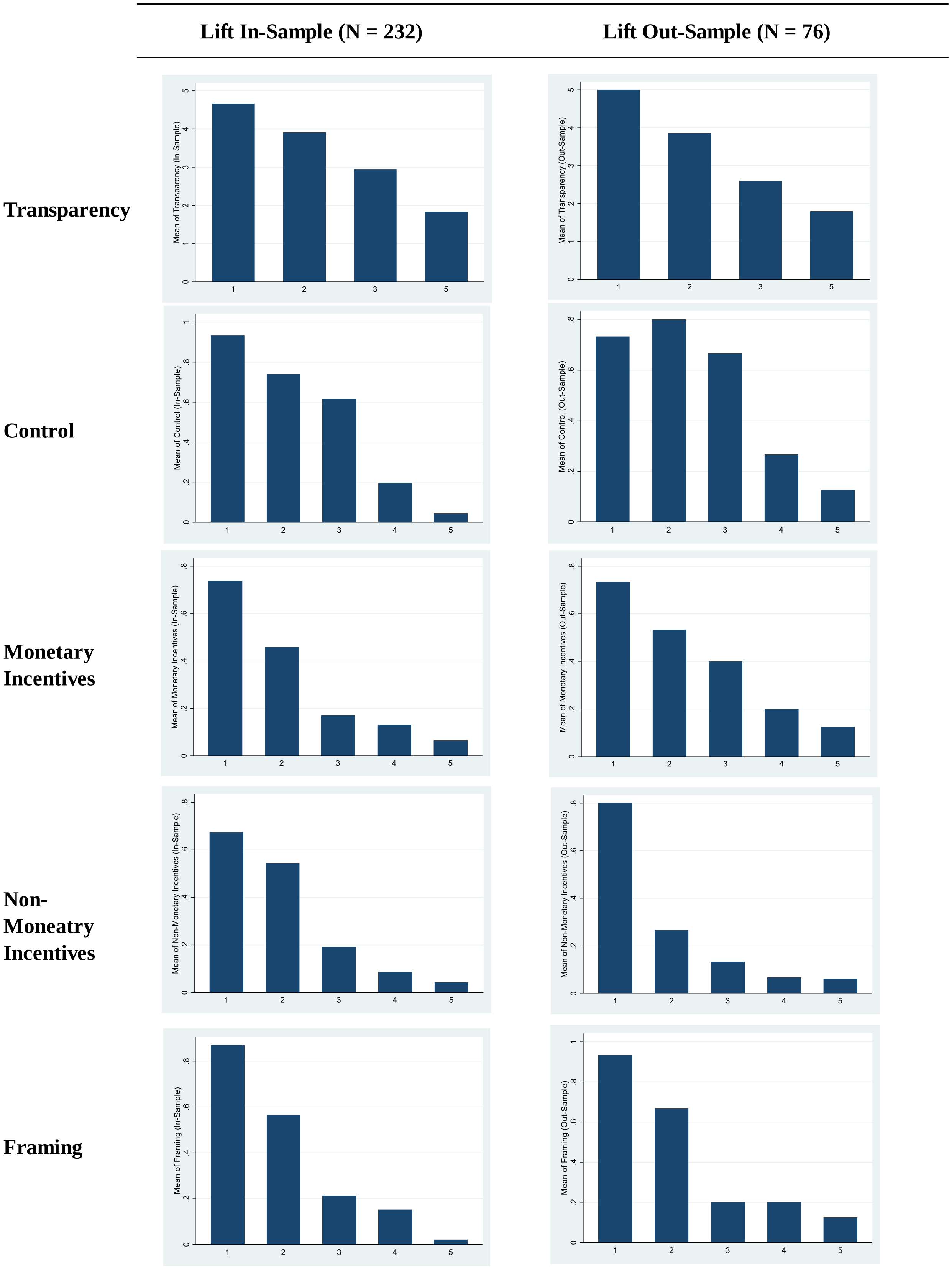
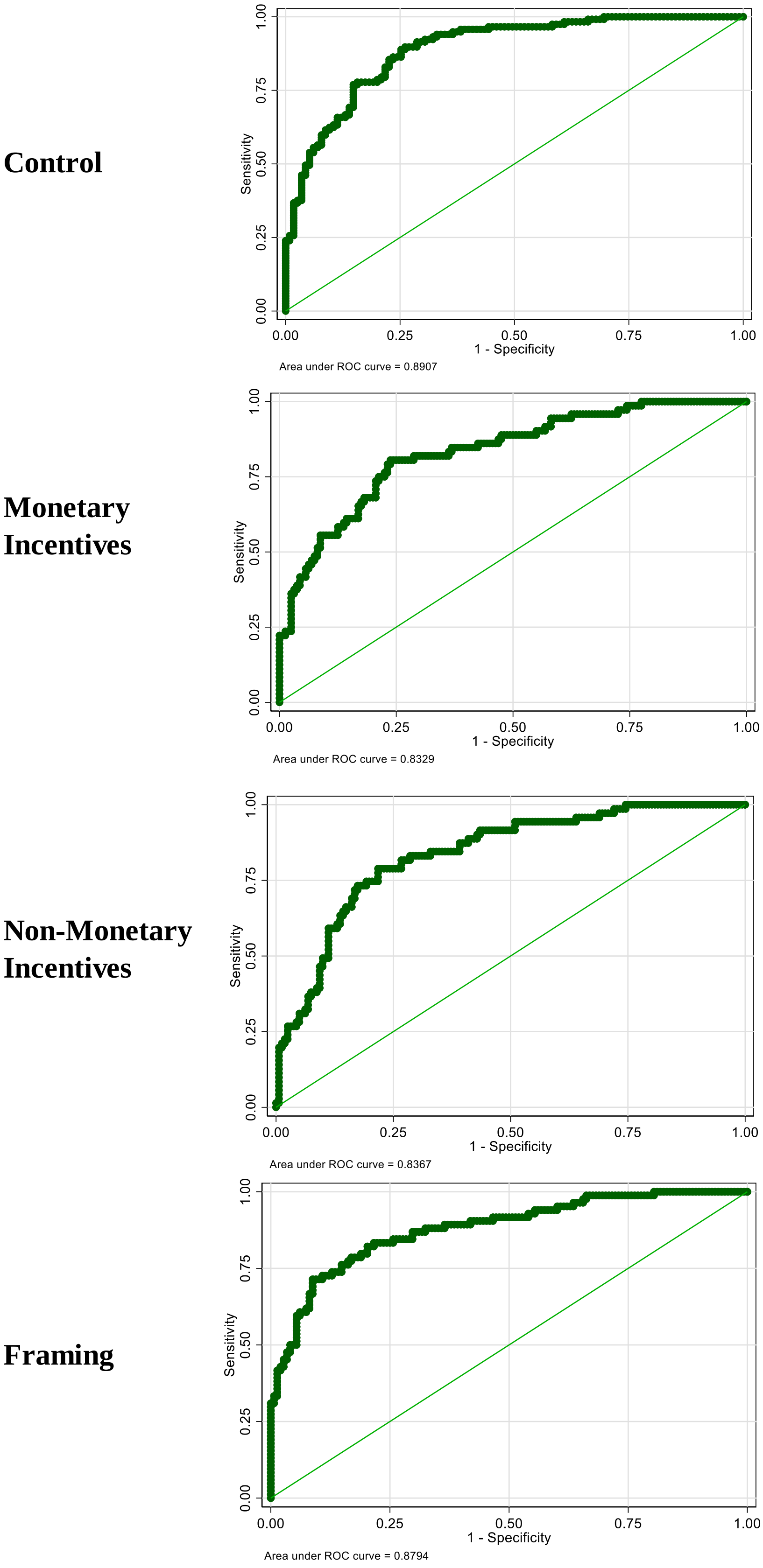


Figure D.2 – ROC Curves (Logit Models).



Control

Area under ROC curve = 0.8907

Monetary Incentives

Area under ROC curve = 0.8329

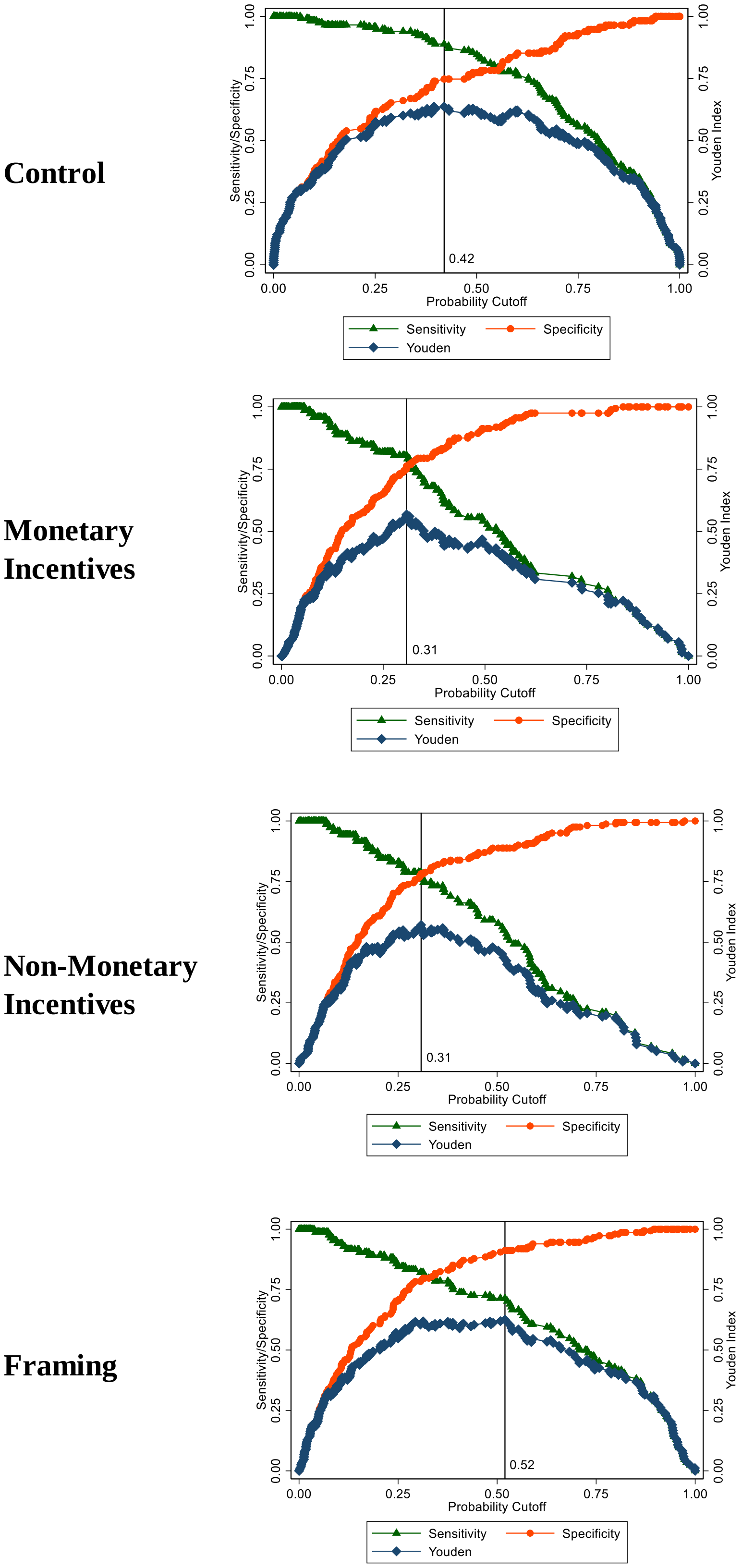
Non-Monetary Incentives

Area under ROC curve = 0.8367

Framing

Area under ROC curve = 0.8794

Figure D.3 – Youden-Index Plots (Logit Models).



APPENDIX E – FACTOR ANALYSIS CORROBORATION CHECK

As described in Chapter 6, our content analysis procedure has been performed by using two methods: (1) data-driven approach – LDA unsupervised technique; (2) theory-based procedure – manual coding of a subsample of the entire collection of email based on the constructs that literature suggests being influential in driving disclosure behavior.

The results of these two procedures allowed us to analyze re-permission emails' content from two perspectives. Thanks to the first procedure, we were able to identify three main topics characterizing our emails by "*letting the data speak*". We can identify them as overarching constructs: Informative, Persuasive and Neither Highly Informative Nor Highly Informative Topics. In addition, through the second approach, we turned to more classical content analysis methods, which are based on the theoretical reasonings of the researcher. Therefore, we captured more granular details about the singular *themes* used by companies in the design of their re-permission emails by estimating the prevalence of the six main themes found in literature in each email collected. Interestingly, the themes identified in literature can be classified into macro-categories by analyzing their communication aim (DellaVigna and Gentzkow 2010). For example, through the theme of transparency, the company tries to increase the knowledge of its user about its security procedure and the users' data rights; therefore, this theme can be classified as *Informative*. In opposition, the use of monetary incentives is a tool that companies use to convince consumers to take action and behave as they would expect (e.g., data disclosure behavior); consequently, this theme may be seen as more *Persuasive* in nature.

Thanks to the availability of two content-analysis methodologies, we were also able to cross-validate them one against the other. Interestingly, we found consistency between the two approaches: on the one side, we found that, in line with our expectations, transparency and control are related to the *Informative Topic*; on the other side, we also find that incentives, framing, and time orientation are related to the *Persuasive Topic*.

In this Appendix, we provide additional support for the distinction of the six themes, identified by the theory-based approach, into the two overarching topics found through the use of the data-driven technique. We use factor analysis on the six themes to identify a smaller number of latent factors that relate to the original variables and can, therefore, classify them into categories of unknown latent constructs.

Before getting into the details of the procedure, Table E.1 shows the descriptive statistics of the six themes identified.

Table E.1 - Descriptive Statistics of the Six Themes.

Variable	**Type of Variable**	**Obs.**	**Mean**	**Std. Dev.**	**Min**	**Max**
Control	Ordinal	1,506	1.878	0.909	1	3
Transparency	Ordinal	1,506	2.756	1.127	1	5
Monetary Incentives	Continuous	1,506	0.278	0.266	0.001	1.000
Non-Monetary Incentives	Continuous	1,506	0.280	0.243	0.000	1.000
Framing: Gain/Loss	Continuous	1,506	0.332	0.300	0.000	1.000
Framing: Time Orientation	Continuous	1,506	0.131	0.037	0.028	0.385

As it is possible to see, we have two types of variables: ordinal and continuous. Therefore, we could not use the standard methods to perform factor analysis (e.g., Pearson’s correlations) since they require the variables to be continuous. Instead, we implemented the factor analysis by using the matrix of the polychoric correlations among the variables (see Table E.2), which allows to get a measure of association when variables are either continuous, binary, or ordinal, and to extract the uncorrelated latent factors.

Table E.2 - Polychoric correlation matrix

		(1)	(2)	(3)	(4)	(5)	(6)
(1)	Control	1					
(2)	Transparency	0.7849	1				
(3)	Monetary Incentives	-0.3397	-0.3560	1			
(4)	Non-Monetary Incentives	-0.3445	-0.3411	0.6231	1		
(5)	Framing: Gain/Loss	-0.3233	-0.3388	0.5396	0.6752	1	
(6)	Framing: Time Orientation	-0.2087	-0.2384	0.2042	0.2378	0.2030	1

In order to detect the number of latent constructs starting from the six themes identified, there are different methods of factor retention, and some consideration can be made on the results of the unrotated factor analysis performed. Firstly, as it is possible to see from Table E.3, the analysis highlights that there are at maximum two latent factors since only Factor1 and Factor2 have eigenvalues greater than zero; moreover, the two factors account for most of the variability present in our data. Secondly, Figure E.1 plots the number of factors against their eigenvalues. According to the Cattell's scree test, the number of factors to be kept is the number of factors whose eigenvalues lie above the point in which the plot displays an "elbow". As it is possible to see, also in this case, the plot indicates that a solution with two factors is the one to be preferred. Lastly, Figure E.2 shows the Horn's parallel analysis done on our data which is the most consistent method used to determine the number of factors to be retained. It is defined as a "*procedure that compares the measured eigenvalues from the data matrix against a Monte-Carlo simulated matrix of random data of the equivalent size*" (Allen 2017, p. 518). The analysis shows that we can only keep factors that have eigenvalues greater than the equivalent factors in the "random" matrix, that is, the number of factors that lie above the point in which the dashed line ("random" data) crosses the solid line (real data). In our case, we can retain, once again, two main factors.

Table E.3 - Factor Analysis (Unrotated)

Factor analysis/correlation	Number of obs. = 1,506
Method: principal factors	Retained factors = 2
Rotation: (unrotated)	Number of params = 11

Factor	Eigenvalue	Difference	Proportion	Cumulative
Factor1	**2.514**	**1.791**	**0.892**	**0.892**
Factor2	**0.723**	**0.732**	**0.257**	**1.148**
Factor3	-0.008	0.073	-0.003	1.145
Factor4	-0.081	0.069	-0.029	1.117
Factor5	-0.150	0.028	-0.053	1.063
Factor6	-0.178	.	-0.063	1.000

LR test: independent vs. saturated: chi2(15) = 3589.41 Prob>chi2 = 0.0000

Factor loadings (pattern matrix) and unique variances

Variable	Factor1	Factor2	Uniqueness
Control	**-0.696**	**0.464**	0.300
Transparency	**-0.709**	**0.459**	0.287
Monetary Incentives	**0.660**	0.258	0.498
Non-Monetary Incentives	**0.729**	0.360	0.339
Framing: Gain/Loss	**0.679**	0.320	0.436
Framing: Time Orientation	0.313	-0.005	0.902

Figure E.1 – Scree Plot of the Eigenvalues (Unrotated Factor Analysis)

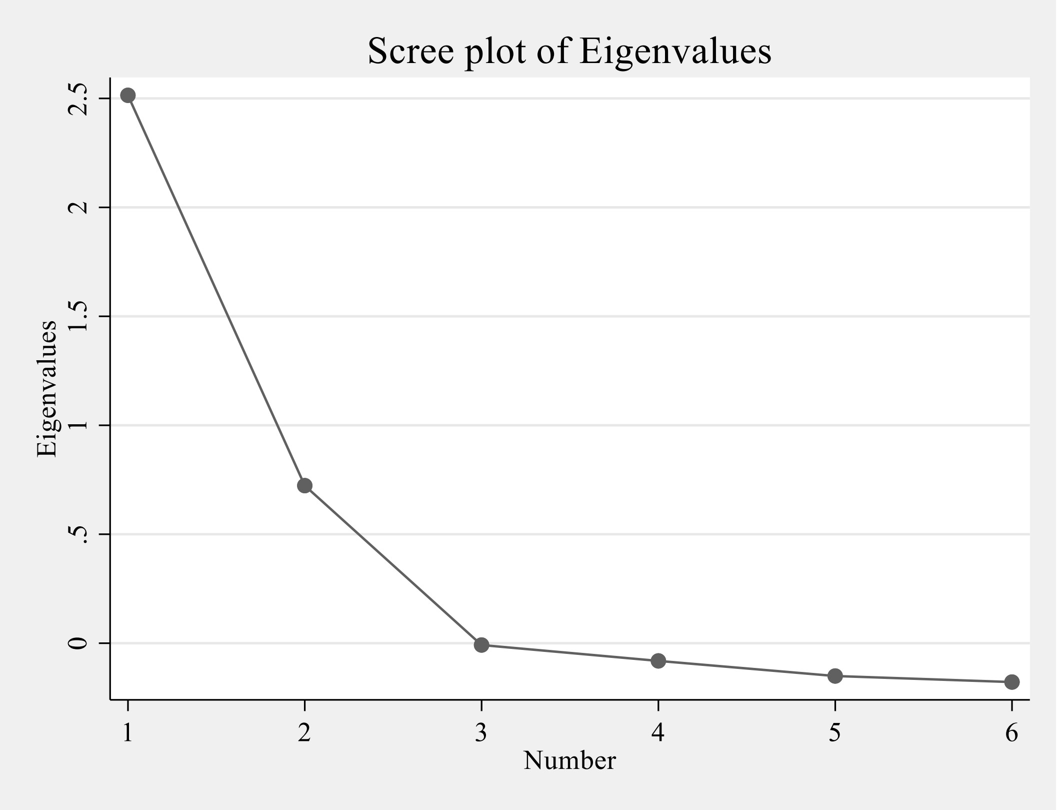
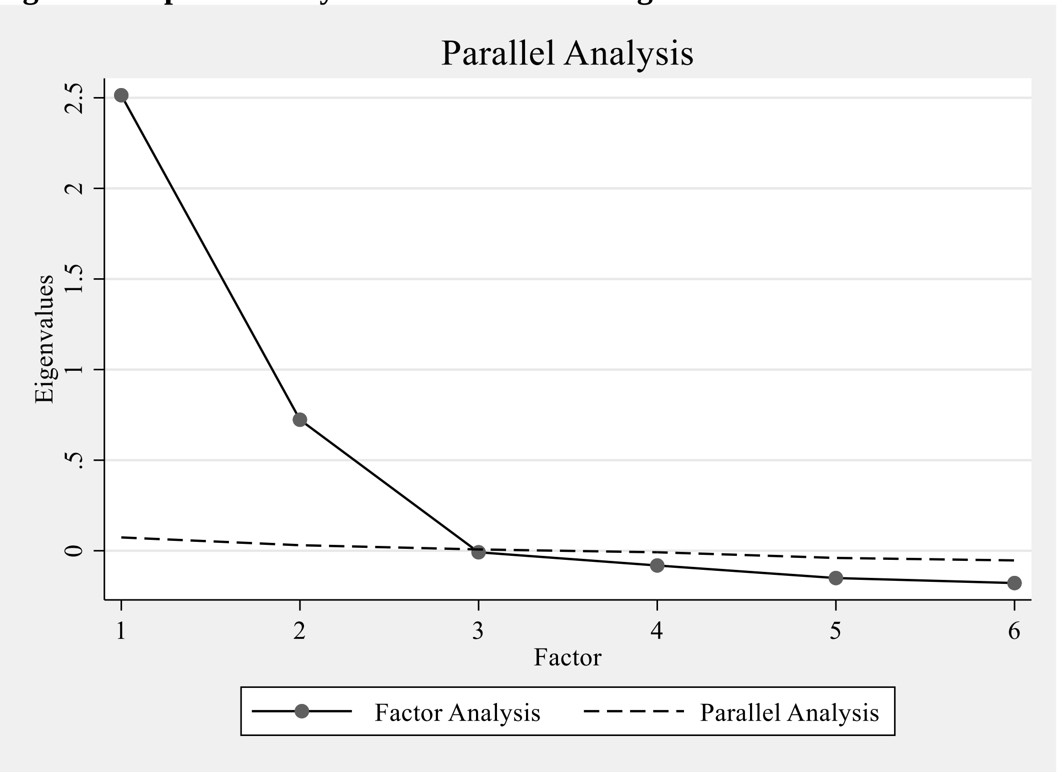


Figure E.2 – Parallel Analysis to Identify the Number of Latent Factors (it compares the eigenvalues produced by random data to the eigenvalues obtained from the real data).



In order to be better able to interpret and identify the latent factors and their loadings on the six themes, we used a Varimax rotation on the data we have. This allows to keep the axes orthogonal and to position them such that each variable loads predominantly on one of the factors, allowing to better detect how much each of the variables is weighted for each of the factors. Indeed, as it is possible to see from Table E.4, the rotated factor loadings allow for better discrimination among which factors load on Factor1 and which one load on Factor2 (compared to Table E.3). According to our previous analysis and expectations, we found that Control and Transparency present a high positive weight on Factor2 and negative weights on Factor1; in contrast, monetary incentives, non-monetary incentives, and framing load positively on Factor1 and negatively on Factor2. Therefore, we can name Factor1 as "*Informative Factor*" and Factor2 as "*Persuasive Factor*".

Table E.4 - Factor Analysis (Varimax Rotation)

```
Factor analysis/correlation                    Number of obs. = 1,506
    Method: principal factors                  Retained factors =   2
    Rotation: orthogonal varimax (Kaiser on)   Number of params =  11
```

Factor	Variance	Difference	Proportion	Cumulative
Factor1	1.752	0.266	0.621	0.621
Factor2	1.486		0.527	1.148

LR test: independent vs. saturated: chi2(15) = 3589.41 Prob>chi2 = 0.0000

Rotated factor loadings (pattern matrix) and unique variances

Variable	Factor1	Factor2	Uniqueness
Control	-0.225	**0.806**	0.300
Transparency	-0.238	**0.810**	0.287
Monetary Incentives	**0.668**	-0.235	0.498
Non-Monetary Incentives	**0.787**	-0.203	0.339
Framing: Gain/Loss	**0.724**	-0.201	0.436
Framing: Time Orientation	0.234	-0.208	0.902

Lastly, we predicted the two factors identified (e.g., informative and persuasive), and we checked if and how these two factors relate to the three LDA topics previously identified and named. Table E.5 presents the correlation matrix between the variables. As it is possible to see, once again, we have consistency between the factors and the LDA: factor 1 (persuasive factor) is positively correlated to LDA topic 2 (persuasive topic) and negatively correlated to LDA topic 1 (informative topic). The vice versa holds true for factor 2.

Table E.5 - Correlation Matrix between the Factors (Factor Analysis) and the Topics (LDA)

				Correlation Matrix				
		Variable	**Label**	**(1)**	**(2)**	**(3)**	**(4)**	**(5)**
Factor Analysis	(1)	Factor 1	Persuasive	1				
	(2)	Factor 2	Informative	0.22	1			
LDA	(3)	Topic 1	Informative	**-0.45**	**0.32**	1		
	(4)	Topic 2	Persuasive	**0.41**	**-0.50**	-0.83	1	
	(5)	Topic 3	Neither H. Informative, nor H. Persuasive	0.05	0.33	-0.27	-0.31	1

APPENDIX F – ROBUSTNESS CHECKS FOR THE LDA PERSUASIVE TOPIC MODELS

This section aims to present the main robustness checks we have done to corroborate the results achieved and described in Section 7.3.

We present results concerning correlations and model estimates achieved by using the different measurements of the “Number of Marketing Cookies” variable produced by the first and the second extraction that Cookiebot provided us with. We also double-checked the results by using the average of the measurements between the two extractions.

Additionally, we also compare the results of the models estimated on the whole sample of emails with the findings of those estimated on the sub-sample of emails that were sent by one unique company – we removed the 6.6% of companies that sent out multiple re-permission emails in the observational period.

As it is possible to notice, all the results are consistent and in line with the one presented in Section 7.3.

F.1. Robustness Check: Choice of the “Number of Marketing Cookies” Measurement

Table F.1.1 presents the correlations between the three measurements of the number of marketing cookies recorded on the firms’ websites. As it is possible to see, there is a high correlation between these variables, meaning that they are actually representing the same type of information.

Table F.1.2 and Table F.1.3 show, respectively, the results of the fractional logit models and the delta in the parameters’ estimates achieved. As stated in Chapter 5, Cookiebot allowed us to collect information about the number of cookies present on the companies’ websites at two different points in time. Therefore, we tested whether the choice of the measurement for the “Number of Marketing Cookies” variable was affecting and driving our results. Also in this case, we can clearly see that all the independent variables used (risks, benefits, and controls) present the same degree of significance, the same direction, and magnitude, independently from the measurement chosen for the "Number of Marketing Cookies" variable.

Table F.1.1 – Correlation Table between the three Measurements of the Number of Marketing Cookies.

	1° Extraction	2° Extraction	Average
1° Extraction	1		
2° Extraction	0.91	1	
Average	0.91	0.91	1

Table F.1.2 – Robustness Checks for Fractional Logit Models with DV = LDA Persuasive Topic. The three columns present the results achieved by using, as an independent variable, the measurement of, respectively, the first extraction, the second extraction, and the average between the two extractions provided by Cookiebot.

		DV = LDA Persuasive Topic					
		1° Extraction [1]		2° Extraction		Average Between 1° & 2° Extraction	
Benefits	# Marketing Cookies (1)	0.000	(0.001)	0.000	(0.001)	0.000	(0.001)
	ln(Expected Monthly Online Ad Revenue) [2]	0.023**	(0.012)	0.023*	(0.012)	0.023*	(0.012)
Risks	Website Popularity (3 Months pre-GDPR)	-0.390**	(0.154)	-0.388**	(0.154)	-0.389**	(0.154)
	# Data-Breaches (pre-GDPR)	-0.000	(0.000)	-0.000	(0.000)	-0.000	(0.000)
Email's Content	LDA Informative Topic	-4.559***	(0.117)	-4.558***	(0.117)	-4.559***	(0.117)
Controls	EU	-0.266***	(0.052)	-0.265***	(0.052)	-0.265***	(0.052)
	Firm Size	-0.010	(0.016)	-0.010	(0.016)	-0.010	(0.016)
	Firm Age	-0.000	(0.002)	-0.000	(0.002)	-0.000	(0.002)
	Sectors:						
	Media and Entertainment	-0.361***	(0.131)	-0.362***	(0.131)	-0.362***	(0.131)
	Professional Services	0.071	(0.090)	0.070	(0.090)	0.070	(0.090)
	Retail Trade	0.093	(0.095)	0.092	(0.095)	0.092	(0.095)
	Software and IT Services	0.071	(0.107)	0.071	(0.107)	0.071	(0.107)
	Travel, Tourism and Hospitality	0.212*	(0.110)	0.211*	(0.110)	0.212*	(0.110)
	Advertiser (0/1)=1 [3]	0.011	(0.107)	0.017	(0.106)	0.015	(0.107)
	Country Missing	-0.435*	(0.244)	-0.435*	(0.244)	-0.435*	(0.244)
	Firm Size Missing	-0.394**	(0.172)	-0.393**	(0.172)	-0.394**	(0.172)
	Firm Age Missing	0.318**	(0.135)	0.319**	(0.135)	0.319**	(0.135)
	Advertiser Missing	0.072	(0.130)	0.078	(0.129)	0.076	(0.130)
	Constant	1.609***		1.599***		1.603***	
	Observations	1364		1364		1364	
	Log-Pseudolikelihood	-498.86		-498.84		-498.85	
	AIC	1035.72		1035.69		1035.71	
	BIC	1134.86		1134.83		1134.85	

Notes:
Standard errors in parentheses
* p<0.10, ** p<0.05, *** p<0.01
The sample size has been reduced because there are companies for which the expected online ad revenue figure was not available.

[1] Results presented in this column are the ones presented in column (6) of the model described in Chapter 7.3.

[2] We add a small constant term to get around instances of zeroes in this variable as we take logs (Pattabhiramaiah, Sriram and Manchanda 2019).

[3] The "Advertiser" dummy variable is equal to 1 if the company is only an advertiser (N = 962), and 0 if the company is (i) a publisher and advertiser (N = 105); (ii) only a publisher (N = 12); or (iii) missing (N=427). We controlled for the missing values by adding a dummy variable which assume value equal to 1 if the information is missing and 0 otherwise.

Table F.1.3 – Differences in the Estimates of the Fractional Logit Models presented in Table F.1.2.

		Delta 1° & 2° Extraction	Delta 1° & Average	Delta 2° & Average
			Coef.	
Benefits	# Marketing Cookies (1)	0.000	0.000	0.000
	ln(Expected Monthly Online Ad Revenue)	0.000	0.000	0.000
Risks	Website Popularity (3 Months pre-GDPR)	-0.002	-0.001	0.001
	# Data-Breaches (pre-GDPR)	0.000	0.000	0.000
Email's Content	LDA Informative Topic	-0.001	0.000	0.001
Controls	EU	-0.001	-0.001	0.000
	Firm Size	0.000	0.000	0.000
	Firm Age	0.000	0.000	0.000
	Sectors:			
	Media and Entertainment	0.001	0.001	0.000
	Professional Services	0.001	0.001	0.000
	Retail Trade	0.001	0.001	0.000
	Software and IT Services	0.000	0.000	0.000
	Travel, Tourism and Hospitality	0.001	0.000	-0.001
	Advertiser (0/1)=1	-0.006	-0.004	0.002
	Country Missing	0.000	0.000	0.000
	Firm Size Missing	-0.001	0.000	0.001
	Firm Age Missing	-0.001	-0.001	0.000
	Advertiser Missing	-0.006	-0.004	0.002

F.2. Robustness Check: Subset of Companies with Only One Re-Permission Email

Table F.2.1 presents the results of the robustness checks done regarding the sample of companies sending the re-permission emails collected. As highlighted in Chapter 5.3, we have 6.6% of the companies in our sample that sent out multiple re-permission emails. Therefore, we tested whether there are relevant differences between the model estimated on the whole sample of companies – and of emails – and the model estimated on the subset of companies that sent out only one re-permission email. As it is possible to see, the results from these models are rather consistent, suggesting that the emails provided by the additional 6.6% of companies do not deviate our analysis.

Table F.2.1 – Robustness Checks for Fractional Logit Models with DV = LDA Persuasive Topic.

The two columns present the results achieved by using the measurement of, respectively, the whole sample of available emails and the subset of emails sent by companies recorded one time into the database (we removed 6.6% of the companies which sent out more different re-permission emails).

		DV = LDA Persuasive Topic			
		Whole Sample of Companies [1]		Sample of Companies without Duplicates [2]	
Benefits	# Marketing Cookies (1)	0.000	(0.001)	0.000	(0.001)
	ln(Expected Monthly Online Ad Revenue) [3]	0.023*	(0.012)	0.021	(0.013)
Risks	Website Popularity (3 Months pre-GDPR)	-0.388**	(0.154)	-0.900***	(0.297)
	# Data-Breaches (pre-GDPR)	-0.000	(0.000)	-0.000	(0.000)
Email's Content	LDA Informative Topic	-4.558***	(0.117)	-4.476***	(0.128)
Controls	EU	-0.265***	(0.052)	-0.293***	(0.057)
	Firm Size	-0.010	(0.016)	-0.008	(0.018)
	Firm Age	-0.000	(0.002)	0.002	(0.001)
	Sectors:				
	Media and Entertainment	-0.362***	(0.131)	-0.341**	(0.147)
	Professional Services	0.070	(0.090)	0.109	(0.096)
	Retail Trade	0.092	(0.095)	0.121	(0.104)
	Software and IT Services	0.071	(0.107)	0.111	(0.116)
	Travel, Tourism and Hospitality	0.211*	(0.110)	0.247**	(0.123)
	Advertiser (0/1)=1 [4]	0.017	(0.106)	-0.120	(0.122)
	Country Missing	-0.435*	(0.244)	-0.508**	(0.243)
	Firm Size Missing	-0.393**	(0.172)	-0.413**	(0.173)
	Firm Age Missing	0.319**	(0.135)	0.440***	(0.125)
	Advertiser Missing	0.078	(0.129)	-0.031	(0.141)
	Constant	1.599***		1.649***	
	Observations	1364		1164	
	Log-Pseudolikelihood	-498.84		-432.12	
	AIC	1035.69		902.25	
	BIC	1134.83		998.38	

Notes:

Standard errors in parentheses

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

The sample size has been reduced because there are companies for which the expected online ad revenue figure was not

[1] Results presented in this column are the ones presented in column (6) of the model described in Chapter 7.3.

[2] Results presented in this column are based on 1304 re-permission emails that were sent out by 1304 companies.

[3] We add a small constant term to get around instances of zeroes in this variable as we take logs (Pattabhiramaiah, Sriram and Manchanda 2019).

[4] The "Advertiser" dummy variable is equal to 1 if the company is only an advertiser (N = 962), and 0 if the company is (i) a publisher and advertiser (N = 105); (ii) only a publisher (N = 12); or (iii) missing (N=427). We controlled for the missing values by adding a dummy variable which assume value equal to 1 if the information is missing and 0 otherwise.

APPENDIX G – ROBUSTNESS CHECKS FOR THE PERSUASIVE THEMES MODELS

This section aims to present the main robustness checks we have done to corroborate the results achieved and described in Section 7.4.

Consequently, we present results concerning the correlation and model estimates comparison estimates achieved by using the different measurements of the "Number of Marketing Cookies" variable produced by the first and the second extraction that Cookiebot provided us with. We also double-checked the results by using the average of the measurements between the two extractions.

Additionally, we also compare the results of the models estimated on the whole sample of emails with the findings of those estimated on the sub-sample of emails that were sent by one unique company – we removed the 6.6% of companies that sent out multiple re-permission emails in the observational period.

As it is possible to notice, all the results are consistent and in line with the one presented in Section 7.4.

G.1. Robustness Check: Choice of the "Number of Marketing Cookies" Measurement

Table G.1.1 presents the correlations between the three measurements of the number of marketing cookies recorded on the firms' websites. As it is possible to see, there is a high correlation between these variables, meaning that they are actually representing the same type of information.

Table from G.1.2 to Table G.1.9 show the results of the fractional logit models and the delta in the parameters' estimates achieved in estimating the presence of monetary incentives, the presence of non-monetary incentives, the use of a particular gain/loss frame, and the use of time-oriented words in firms' re-permission emails. As stated in Chapter 5, Cookiebot allowed us to collect information about the number of cookies present on the companies' websites at two different points in time. Therefore, we tested whether the choice of the measurement for the "Number of Marketing Cookies" variable was affecting and driving our results. Also in this case, we can clearly see that all the independent variables used (risks, benefits, and controls) present the same degree of significance, the same direction, and magnitude, independently from the measurement chosen for the "Number of Marketing Cookies" variable.

Table F.1.1 – Correlation Table between the three Measurements of the Number of Marketing Cookies.

	1° Extraction	2° Extraction	Average
1° Extraction	1		
2° Extraction	0.91	1	
Average	0.91	0.91	1

Table G.1.2 – Robustness Checks for Fractional Logit Models with DV = "Monetary Incentive" Theme.

The three columns present the results achieved by using, as an independent variable, the measurement of, respectively, the first extraction, the second extraction, and the average between the two extractions provided by Cookiebot.

		DV = Monetary Incentives					
		1° Extraction [1]		2° Extraction		Average Between 1° & 2° Extraction	
Benefits	# Marketing Cookies (1)	0.000	(0.001)	-0.000	(0.001)	0.000	(0.001)
	ln(Expected Monthly Online Ad Revenue) [2]	0.042***	(0.013)	0.042***	(0.013)	0.042***	(0.013)
Risks	Website Popularity (3 Months pre-GDPR)	-0.000	(0.000)	-0.000	(0.000)	-0.000	(0.000)
	# Data-Breaches (pre-GDPR)	0.012	(0.147)	0.008	(0.146)	0.010	(0.147)
Email's Content	Non-Monetary Incentives	2.349***	(0.192)	2.353***	(0.192)	2.350***	(0.192)
	Framing: Gain/Loss	0.960***	(0.154)	0.961***	(0.154)	0.961***	(0.154)
	Framing: Time Orientation	0.387	(1.180)	0.396	(1.178)	0.390	(1.179)
	Control	-0.045	(0.048)	-0.043	(0.048)	-0.044	(0.048)
	Transparency	-0.112***	(0.041)	-0.113***	(0.041)	-0.113***	(0.041)
Controls	EU	0.196***	(0.075)	0.195***	(0.075)	0.196***	(0.075)
	Firm Size	0.018	(0.018)	0.018	(0.018)	0.018	(0.018)
	Firm Age	-0.001	(0.001)	-0.001	(0.001)	-0.001	(0.001)
	Sectors:						
	Media and Entertainment	-0.223*	(0.119)	-0.222*	(0.119)	-0.223*	(0.119)
	Professional Services	0.088	(0.097)	0.090	(0.097)	0.089	(0.097)
	Retail Trade	0.338***	(0.104)	0.340***	(0.104)	0.339***	(0.104)
	Software and IT Services	0.070	(0.123)	0.071	(0.123)	0.070	(0.123)
	Travel, Tourism and Hospitality	0.103	(0.127)	0.105	(0.127)	0.104	(0.127)
	Advertiser (0/1)=1 [3]	0.224*	(0.126)	0.207	(0.127)	0.216*	(0.127)
	Country Missing	-0.330	(0.266)	-0.333	(0.266)	-0.332	(0.266)
	Firm Size Missing	0.340*	(0.182)	0.339*	(0.182)	0.340*	(0.182)
	Firm Age Missing	-0.002	(0.192)	-0.003	(0.193)	-0.003	(0.192)
	Advertiser Missing	0.185	(0.146)	0.164	(0.147)	0.174	(0.147)
	Constant	-2.599***	(0.290)	-2.575***	(0.290)	-2.588***	(0.290)
	Observations	1364		1364		1364	
	Log-Pseudolikelihood	-528.53		-528.55		-528.55	
	AIC	1103.06		1103.10		1103.10	
	BIC	1223.08		1223.12		1223.12	

Notes:
Standard errors in parentheses
* p<0.10, ** p<0.05, *** p<0.01
The sample size has been reduced because there are companies for which the expected online ad revenue figure was not available.

[1] Results presented in this column are the ones presented in column (1) of the model described in Chapter 7.4.

[2] We add a small constant term to get around instances of zeroes in this variable as we take logs (Pattabhiramaiah, Sriram and Manchanda 2019).

3 The "Advertiser" dummy variable is equal to 1 if the company is only an advertiser (N = 962), and 0 if the company is (i) a publisher and advertiser (N = 105); (ii) only a publisher (N = 12); or (iii) missing (N=427). We controlled for the missing values by adding a dummy variable which assume value equal to 1 if the information is missing and 0 otherwise.

Table G.1.3 – Differences in the Estimates of the Fractional Logit Models presented in Table G.1.2.

		Delta 1° & 2° Extraction	**Delta 1° & Average**	**Delta 2° & Average**
			Coef.	
Benefits	# Marketing Cookies (1)	0.000	0.000	0.000
	ln(Expected Monthly Online Ad Revenue)	0.000	0.000	0.000
Risks	Website Popularity (3 Months pre-GDPR)	0.000	0.000	0.000
	# Data-Breaches (pre-GDPR)	0.004	0.002	-0.002
Email's Content	Non-Monetary Incentives	-0.004	-0.001	0.003
	Framing: Gain/Loss	-0.001	-0.001	0.000
	Framing: Time Orientation	-0.009	-0.003	0.006
	Control	-0.002	-0.001	0.001
	Transparency	0.001	0.001	0.000
Controls	EU	0.001	0.000	-0.001
	Firm Size	0.000	0.000	0.000
	Firm Age	0.000	0.000	0.000
	Sectors:			
	Media and Entertainment	-0.001	0.000	0.001
	Professional Services	-0.002	-0.001	0.001
	Retail Trade	-0.002	-0.001	0.001
	Software and IT Services	-0.001	0.000	0.001
	Travel, Tourism and Hospitality	-0.002	-0.001	0.001
	Advertiser (0/1)=1	0.017	0.008	-0.009
	Country Missing	0.003	0.002	-0.001
	Firm Size Missing	0.001	0.000	-0.001
	Firm Age Missing	0.001	0.001	0.000
	Advertiser Missing	0.021	0.011	-0.010

Table G.1.4 – Robustness Checks for Fractional Logit Models with DV = "Non-Monetary Incentive" Theme.

The three columns present the results achieved by using, as an independent variable, the measurement of, respectively, the first extraction, the second extraction, and the average between the two extractions provided by Cookiebot.

		DV = Non-Monetary Incentives					
		1° Extraction [1]		2° Extraction		Average Between 1° & 2° Extraction	
Benefits	# Marketing Cookies (1)	0.001	(0.001)	0.001*	(0.001)	0.001	(0.001)
	ln(Expected Monthly Online Ad Revenue) [2]	-0.038***	(0.009)	-0.038***	(0.009)	-0.038***	(0.009)
Risks	Website Popularity (3 Months pre-GDPR)	-0.000	(0.000)	-0.000	(0.000)	-0.000	(0.000)
	# Data-Breaches (pre-GDPR)	0.268**	(0.114)	0.273**	(0.114)	0.271**	(0.114)
Email's Content	Monetary Incentives	1.579***	(0.123)	1.580***	(0.123)	1.579***	(0.123)
	Framing: Gain/Loss	1.849***	(0.109)	1.848***	(0.109)	1.848***	(0.109)
	Framing: Time Orientation	2.392***	(0.842)	2.369***	(0.839)	2.381***	(0.840)
	Control	-0.064	(0.041)	-0.065	(0.041)	-0.065	(0.041)
	Transparency	0.014	(0.034)	0.015	(0.034)	0.014	(0.034)
Controls	EU	0.033	(0.057)	0.035	(0.057)	0.034	(0.057)
	Firm Size	0.008	(0.015)	0.008	(0.015)	0.008	(0.015)
	Firm Age	0.001	(0.001)	0.001	(0.001)	0.001	(0.001)
	Sectors:						
	Media and Entertainment	-0.032	(0.095)	-0.031	(0.095)	-0.032	(0.095)
	Professional Services	0.100	(0.080)	0.099	(0.080)	0.099	(0.080)
	Retail Trade	0.001	(0.093)	0.002	(0.093)	0.002	(0.093)
	Software and IT Services	0.152	(0.098)	0.151	(0.097)	0.151	(0.098)
	Travel, Tourism and Hospitality	0.096	(0.095)	0.096	(0.095)	0.096	(0.095)
	Advertiser (0/1)=1 [3]	-0.059	(0.085)	-0.047	(0.085)	-0.051	(0.085)
	Country Missing	-0.335	(0.208)	-0.333	(0.208)	-0.334	(0.208)
	Firm Size Missing	-0.110	(0.129)	-0.110	(0.129)	-0.110	(0.129)
	Firm Age Missing	0.082	(0.132)	0.082	(0.133)	0.082	(0.133)
	Advertiser Missing	0.076	(0.104)	0.087	(0.104)	0.084	(0.104)
	Constant	-2.125***	(0.216)	-2.140***	(0.216)	-2.135***	(0.216)
	Observations	1364		1364		1364	
	Log-Pseudolikelihood	-503.28		-503.17		-503.22	
	AIC	1052.56		1052.35		1052.45	
	BIC	1172.58		1172.37		1172.46	

Notes:
Standard errors in parentheses
* p<0.10, ** p<0.05, *** p<0.01
The sample size has been reduced because there are companies for which the expected online ad revenue figure was not available.

[1] Results presented in this column are the ones presented in column (2) of the model described in Chapter 7.4.

[2] We add a small constant term to get around instances of zeroes in this variable as we take logs (Pattabhiramaiah, Sriram and Manchanda 2019).

[3] The "Advertiser" dummy variable is equal to 1 if the company is only an advertiser (N = 962), and 0 if the company is (i) a publisher and advertiser (N = 105); (ii) only a publisher (N = 12); or (iii) missing (N=427). We controlled for the missing values by adding a dummy variable which assume value equal to 1 if the information is missing and 0 otherwise.

Table G.1.5 – Differences in the Estimates of the Fractional Logit Models presented in Table G.1.4.

		Delta 1° & 2° Extraction	**Delta 1° & Average**	**Delta 2° & Average**
			Coef.	
Benefits	# Marketing Cookies (1)	0.000	0.000	0.000
	ln(Expected Monthly Online Ad Revenue)	0.000	0.000	0.000
Risks	Website Popularity (3 Months pre-GDPR)	0.000	0.000	0.000
	# Data-Breaches (pre-GDPR)	-0.005	-0.003	0.002
Email's Content	Monetary Incentives	-0.001	0.000	0.001
	Framing: Gain/Loss	0.001	0.001	0.000
	Framing: Time Orientation	0.023	0.011	-0.012
	Control	0.001	0.001	0.000
	Transparency	-0.001	0.000	0.001
Controls	EU	-0.002	-0.001	0.001
	Firm Size	0.000	0.000	0.000
	Firm Age	0.000	0.000	0.000
	Sectors:			
	Media and Entertainment	-0.001	0.000	0.001
	Professional Services	0.001	0.001	0.000
	Retail Trade	-0.001	-0.001	0.000
	Software and IT Services	0.001	0.001	0.000
	Travel, Tourism and Hospitality	0.000	0.000	0.000
	Advertiser (0/1)=1	-0.012	-0.008	0.004
	Country Missing	-0.002	-0.001	0.001
	Firm Size Missing	0.000	0.000	0.000
	Firm Age Missing	0.000	0.000	0.000
	Advertiser Missing	-0.011	-0.008	0.003

Table G.1.6 – Robustness Checks for Fractional Logit Models with DV = "Gain/Loss Frame" Theme.

The three columns present the results achieved by using, as an independent variable, the measurement of, respectively, the first extraction, the second extraction, and the average between the two extractions provided by Cookiebot.

		DV = Framing: Gain/Loss					
		1° Extraction [1]		2° Extraction		Average Between 1° & 2° Extraction	
Benefits	# Marketing Cookies (1)	0.000	(0.001)	0.000	(0.001)	0.000	(0.001)
	ln(Expected Monthly Online Ad Revenue) [2]	-0.004	(0.013)	-0.004	(0.013)	-0.004	(0.013)
Risks	Website Popularity (3 Months pre-GDPR)	0.000*	(0.000)	0.000*	(0.000)	0.000*	(0.000)
	# Data-Breaches (pre-GDPR)	-0.458***	(0.117)	-0.459***	(0.117)	-0.458***	(0.117)
Email's Content	Monetary Incentives	1.036***	(0.172)	1.037***	(0.172)	1.036***	(0.172)
	Non-Monetary Incentives	3.035***	(0.190)	3.036***	(0.190)	3.035***	(0.190)
	Framing: Time Orientation	1.748*	(0.992)	1.747*	(0.992)	1.748*	(0.992)
	Control	-0.036	(0.048)	-0.036	(0.048)	-0.036	(0.048)
	Transparency	-0.069*	(0.041)	-0.070*	(0.041)	-0.069*	(0.041)
Controls	EU	0.206***	(0.072)	0.205***	(0.072)	0.206***	(0.072)
	Firm Size	-0.023	(0.018)	-0.022	(0.018)	-0.023	(0.018)
	Firm Age	0.001	(0.001)	0.001	(0.001)	0.001	(0.001)
	Sectors:						
	Media and Entertainment	0.028	(0.131)	0.028	(0.131)	0.028	(0.131)
	Professional Services	-0.246**	(0.101)	-0.245**	(0.101)	-0.245**	(0.101)
	Retail Trade	-0.233**	(0.115)	-0.233**	(0.115)	-0.233**	(0.115)
	Software and IT Services	-0.397***	(0.120)	-0.397***	(0.120)	-0.397***	(0.120)
	Travel, Tourism and Hospitality	0.006	(0.119)	0.007	(0.119)	0.006	(0.119)
	Advertiser (0/1)=1 [3]	-0.229*	(0.134)	-0.236*	(0.135)	-0.232*	(0.135)
	Country Missing	0.366	(0.315)	0.365	(0.315)	0.366	(0.315)
	Firm Size Missing	-0.252	(0.187)	-0.252	(0.187)	-0.252	(0.187)
	Firm Age Missing	0.231	(0.187)	0.231	(0.187)	0.231	(0.187)
	Advertiser Missing	-0.309**	(0.157)	-0.318**	(0.157)	-0.313**	(0.157)
	Constant	-1.625***	(0.270)	-1.615***	(0.272)	-1.620***	(0.271)
	Observations	1364		1364		1364	
	Log-Pseudolikelihood	-546.54		-546.55		-546.55	
	AIC	1139.07		1139.10		1139.09	
	BIC	1259.09		1259.12		1259.11	

Notes:
Standard errors in parentheses
* p<0.10, ** p<0.05, *** p<0.01
The sample size has been reduced because there are companies for which the expected online ad revenue figure was not available.

[1] Results presented in this column are the ones presented in column (3) of the model described in Chapter 7.4.

[2] We add a small constant term to get around instances of zeroes in this variable as we take logs (Pattabhiramaiah, Sriram and Manchanda 2019).

[3] The "Advertiser" dummy variable is equal to 1 if the company is only an advertiser (N = 962), and 0 if the company is (i) a publisher and advertiser (N = 105); (ii) only a publisher (N = 12); or (iii) missing (N=427). We controlled for the missing values by adding a dummy variable which assume value equal to 1 if the information is missing and 0 otherwise.

Table G.1.7 – Differences in the Estimates of the Fractional Logit Models presented in Table G.1.6.

		Delta 1° & 2° Extraction	Delta 1° & Average	Delta 2° & Average
			Coef.	
Benefits	# Marketing Cookies (1)	0.000	0.000	0.000
	ln(Expected Monthly Online Ad Revenue)	0.000	0.000	0.000
Risks	Website Popularity (3 Months pre-GDPR)	0.000	0.000	0.000
	# Data-Breaches (pre-GDPR)	0.001	0.000	-0.001
Email's Content	Monetary Incentives	-0.001	0.000	0.001
	Non-Monetary Incentives	-0.001	0.000	0.001
	Framing: Time Orientation	0.001	0.000	-0.001
	Control	0.000	0.000	0.000
	Transparency	0.001	0.000	-0.001
Controls	EU	0.001	0.000	-0.001
	Firm Size	-0.001	0.000	0.001
	Firm Age	0.000	0.000	0.000
	Sectors:			
	Media and Entertainment	0.000	0.000	0.000
	Professional Services	-0.001	-0.001	0.000
	Retail Trade	0.000	0.000	0.000
	Software and IT Services	0.000	0.000	0.000
	Travel, Tourism and Hospitality	-0.001	0.000	0.001
	Advertiser (0/1)=1	0.007	0.003	-0.004
	Country Missing	0.001	0.000	-0.001
	Firm Size Missing	0.000	0.000	0.000
	Firm Age Missing	0.000	0.000	0.000
	Advertiser Missing	0.009	0.004	-0.005

Table G.1.8 – Robustness Checks for Fractional Logit Models with DV = "Time Orientation" Theme.

The three columns present the results achieved by using, as an independent variable, the measurement of, respectively, the first extraction, the second extraction, and the average between the two extractions provided by Cookiebot.

		DV = Framing: Time Orientation					
		1° Extraction [1]		2° Extraction		Average Between 1° & 2° Extraction	
Benefits	# Marketing Cookies (1)	0.000	(0.000)	0.000	(0.000)	0.000	(0.000)
	ln(Expected Monthly Online Ad Revenue) [2]	0.013***	(0.004)	0.013***	(0.004)	0.013***	(0.004)
Risks	Website Popularity (3 Months pre-GDPR)	0.000	(0.000)	0.000	(0.000)	0.000	(0.000)
	# Data-Breaches (pre-GDPR)	-0.034	(0.033)	-0.033	(0.033)	-0.034	(0.033)
Email's Content	Monetary Incentives	0.021	(0.055)	0.021	(0.055)	0.021	(0.055)
	Non-Monetary Incentives	0.184***	(0.060)	0.183***	(0.060)	0.183***	(0.060)
	Framing: Gain/Loss	0.071	(0.043)	0.071	(0.043)	0.071	(0.043)
	Control	-0.013	(0.013)	-0.013	(0.013)	-0.013	(0.013)
	Transparency	-0.038***	(0.011)	-0.038***	(0.011)	-0.038***	(0.011)
Controls	EU	-0.013	(0.020)	-0.013	(0.020)	-0.013	(0.020)
	Firm Size	-0.012**	(0.005)	-0.012**	(0.005)	-0.012**	(0.005)
	Firm Age	0.000	(0.000)	0.000	(0.000)	0.000	(0.000)
	Sectors:						
	Media and Entertainment	-0.025	(0.034)	-0.025	(0.034)	-0.025	(0.034)
	Professional Services	0.013	(0.026)	0.012	(0.026)	0.013	(0.026)
	Retail Trade	0.039	(0.030)	0.039	(0.030)	0.039	(0.030)
	Software and IT Services	-0.011	(0.030)	-0.011	(0.030)	-0.011	(0.030)
	Travel, Tourism and Hospitality	-0.019	(0.031)	-0.019	(0.031)	-0.019	(0.031)
	Advertiser (0/1)=1 [3]	-0.023	(0.033)	-0.021	(0.033)	-0.022	(0.033)
	Country Missing	-0.055	(0.065)	-0.054	(0.065)	-0.054	(0.065)
	Firm Size Missing	-0.073*	(0.043)	-0.072*	(0.043)	-0.072*	(0.043)
	Firm Age Missing	0.076*	(0.044)	0.076*	(0.044)	0.076*	(0.044)
	Advertiser Missing	0.002	(0.039)	0.005	(0.039)	0.004	(0.039)
	Constant	-1.926***	(0.072)	-1.929***	(0.072)	-1.927***	(0.072)
	Observations	1364		1364		1364	
	Log-Pseudolikelihood	-378.68		-378.68		-378.68	
	AIC	803.36		803.35		803.35	
	BIC	923.37		923.37		923.37	

Notes:
Standard errors in parentheses
* p<0.10, ** p<0.05, *** p<0.01
The sample size has been reduced because there are companies for which the expected online ad revenue figure was not available.

[1] Results presented in this column are the ones presented in column (4) of the model described in Chapter 7.4.

[2] We add a small constant term to get around instances of zeroes in this variable as we take logs (Pattabhiramaiah, Sriram and Manchanda 2019).

[3] The "Advertiser" dummy variable is equal to 1 if the company is only an advertiser (N = 962), and 0 if the company is (i) a publisher and advertiser (N = 105); (ii) only a publisher (N = 12); or (iii) missing (N=427). We controlled for the missing values by adding a dummy variable which assume value equal to 1 if the information is missing and 0 otherwise.

Table G.1.9 – Differences in the Estimates of the Fractional Logit Models presented in Table G.1.8.

		Delta 1° & 2° Extraction	Delta 1° & Average	Delta 2° & Average
			Coef.	
Benefits	# Marketing Cookies (1)	0.000	0.000	0.000
	ln(Expected Monthly Online Ad Revenue)	0.000	0.000	0.000
Risks	Website Popularity (3 Months pre-GDPR)	0.000	0.000	0.000
	# Data-Breaches (pre-GDPR)	-0.001	0.000	0.001
Email's Content	Monetary Incentives	0.000	0.000	0.000
	Non-Monetary Incentives	0.001	0.001	0.000
	Framing: Gain/Loss	0.000	0.000	0.000
	Control	0.000	0.000	0.000
	Transparency	0.000	0.000	0.000
Controls	EU	0.000	0.000	0.000
	Firm Size	0.000	0.000	0.000
	Firm Age	0.000	0.000	0.000
	Sectors:			
	Media and Entertainment	0.000	0.000	0.000
	Professional Services	0.001	0.000	-0.001
	Retail Trade	0.000	0.000	0.000
	Software and IT Services	0.000	0.000	0.000
	Travel, Tourism and Hospitality	0.000	0.000	0.000
	Advertiser (0/1)=1	-0.002	-0.001	0.001
	Country Missing	-0.001	-0.001	0.000
	Firm Size Missing	-0.001	-0.001	0.000
	Firm Age Missing	0.000	0.000	0.000
	Advertiser Missing	-0.003	-0.002	0.001

G.2. Robustness Check: Subset of Companies with Only One Re-Permission Email

Results from Table G.2.1 to Table G.2.4 present the robustness checks done on the models estimating the four theory-based themes regarding the selection of the sample of companies that have sent the re-permission emails collected. As highlighted in Chapter 5.3, we have 6.6% of the companies in our sample that sent out multiple re-permission emails. Therefore, we tested whether there are relevant differences between the models estimated on the whole sample of companies – and of emails – and the models estimated on the subset of companies that sent out only one re-permission email. As it is possible to see, the results from these models are rather consistent, suggesting that the emails provided by the additional 6.6% of companies do not deviate our analysis.

Table G.2.1 – Robustness Checks for Fractional Logit Models with DV = “Monetary Incentive” Theme.

The two columns present the results achieved by using the measurement of, respectively, the whole sample of available emails and the subset of emails sent by companies recorded one time into the database (we removed 6.6% of the companies which sent out more different re-permission emails).

		DV = Monetary Incentives			
		Whole Sample of Companies [1]		Sample of Companies without Duplicates [2]	
Benefits	# Marketing Cookies (1)	0.000	(0.001)	0.000	(0.001)
	ln(Expected Monthly Online Ad Revenue) [3]	0.042***	(0.013)	0.032**	(0.015)
Risks	Website Popularity (3 Months pre-GDPR)	-0.000	(0.000)	-0.000	(0.000)
	# Data-Breaches (pre-GDPR)	0.012	(0.147)	0.047	(0.239)
Email's Content	Non-Monetary Incentives	2.349***	(0.192)	2.298***	(0.206)
	Framing: Gain/Loss	0.960***	(0.154)	1.075***	(0.169)
	Framing: Time Orientation	0.387	(1.180)	1.845	(1.318)
	Control	-0.045	(0.048)	-0.077	(0.053)
	Transparency	-0.112***	(0.041)	-0.050	(0.044)
Controls	EU	0.196***	(0.075)	0.218***	(0.081)
	Firm Size	0.018	(0.018)	0.018	(0.020)
	Firm Age	-0.001	(0.001)	-0.002	(0.002)
	Sectors:				
	Media and Entertainment	-0.223*	(0.119)	-0.214*	(0.128)
	Professional Services	0.088	(0.097)	0.037	(0.107)
	Retail Trade	0.338***	(0.104)	0.321***	(0.114)
	Software and IT Services	0.070	(0.123)	0.048	(0.126)
	Travel, Tourism and Hospitality	0.103	(0.127)	0.122	(0.140)
	Advertiser (0/1)=1 [4]	0.224*	(0.126)	0.195	(0.140)
	Country Missing	-0.330	(0.266)	-0.169	(0.245)
	Firm Size Missing	0.340*	(0.182)	0.231	(0.188)
	Firm Age Missing	-0.002	(0.192)	-0.107	(0.196)
	Advertiser Missing	0.185	(0.146)	0.090	(0.159)
	Constant	-2.599***	(0.290)	-2.773***	(0.314)
	Observations	1364		1164	
	Log-Pseudolikelihood	-528.53		-445.48	
	AIC	1103.06		936.96	
	BIC	1223.08		1053.33	

Notes:
Standard errors in parentheses
* p<0.10, ** p<0.05, *** p<0.01
The sample size has been reduced because there are companies for which the expected online ad revenue figure was not available.

[1] Results presented in this column are the ones presented in column (1) of the model described in Chapter 7.4.

[2] Results presented in this column are based on 1304 re-permission emails that were sent out by 1304 companies.

[3] We add a small constant term to get around instances of zeroes in this variable as we take logs (Pattabhiramaiah, Sriram and Manchanda 2019).

[4] The "Advertiser" dummy variable is equal to 1 if the company is only an advertiser (N = 962), and 0 if the company is (i) a publisher and advertiser (N = 105); (ii) only a publisher (N = 12); or (iii) missing (N=427). We controlled for the missing values by adding a dummy variable which assume value equal to 1 if the information is missing and 0 otherwise.

Table G.2.2 – Robustness Checks for Fractional Logit Models with DV = “Non-Monetary Incentive” Theme.

The two columns present the results achieved by using the measurement of, respectively, the whole sample of available emails and the subset of emails sent by companies recorded one time into the database (we removed 6.6% of the companies which sent out more different re-permission emails).

		DV = Non-Monetary Incentives			
		Whole Sample of Companies [1]		Sample of Companies without Duplicates [2]	
Benefits	# Marketing Cookies (1)	0.001	(0.001)	0.000	(0.001)
	ln(Expected Monthly Online Ad Revenue) [3]	-0.038***	(0.009)	-0.037***	(0.009)
Risks	Website Popularity (3 Months pre-GDPR)	-0.000	(0.000)	-0.000	(0.000)
	# Data-Breaches (pre-GDPR)	0.268**	(0.114)	0.432**	(0.210)
Email's Content	Monetary Incentives	1.579***	(0.123)	1.588***	(0.138)
	Framing: Gain/Loss	1.849***	(0.109)	1.895***	(0.120)
	Framing: Time Orientation	2.392***	(0.842)	1.272	(0.927)
	Control	-0.064	(0.041)	-0.036	(0.044)
	Transparency	0.014	(0.034)	-0.002	(0.037)
Controls	EU	0.033	(0.057)	0.016	(0.063)
	Firm Size	0.008	(0.015)	0.001	(0.017)
	Firm Age	0.001	(0.001)	0.001	(0.001)
	Sectors:				
	Media and Entertainment	-0.032	(0.095)	-0.040	(0.104)
	Professional Services	0.100	(0.080)	0.123	(0.088)
	Retail Trade	0.001	(0.093)	-0.000	(0.100)
	Software and IT Services	0.152	(0.098)	0.155	(0.104)
	Travel, Tourism and Hospitality	0.096	(0.095)	0.109	(0.104)
	Advertiser (0/1)=1 [4]	-0.059	(0.085)	-0.063	(0.100)
	Country Missing	-0.335	(0.208)	-0.467**	(0.205)
	Firm Size Missing	-0.110	(0.129)	-0.097	(0.127)
	Firm Age Missing	0.082	(0.132)	0.183	(0.134)
	Advertiser Missing	0.076	(0.104)	0.083	(0.117)
	Constant	-2.125***	(0.216)	-1.979***	(0.238)
	Observations	1364		1164	
	Log-Pseudolikelihood	-503.28		-430.16	
	AIC	1052.56		906.33	
	BIC	1172.58		1022.70	

Notes:
Standard errors in parentheses
* p<0.10, ** p<0.05, *** p<0.01
The sample size has been reduced because there are companies for which the expected online ad revenue figure was not available.

[1] Results presented in this column are the ones presented in column (2) of the model described in Chapter 7.4.

[2] Results presented in this column are based on 1304 re-permission emails that were sent out by 1304 companies.

[3] We add a small constant term to get around instances of zeroes in this variable as we take logs (Pattabhiramaiah, Sriram and Manchanda 2019).

[4] The "Advertiser" dummy variable is equal to 1 if the company is only an advertiser (N = 962), and 0 if the company is (i) a publisher and advertiser (N = 105); (ii) only a publisher (N = 12); or (iii) missing (N=427). We controlled for the missing values by adding a dummy variable which assume value equal to 1 if the information is missing and 0 otherwise.

Table G.2.3 – Robustness Checks for Fractional Logit Models with DV = "Gain/Loss Frame" Theme.
The two columns present the results achieved by using the measurement of, respectively, the whole sample of available emails and the subset of emails sent by companies recorded one time into the database (we removed 6.6% of the companies which sent out more different re-permission emails).

		DV = Framing: Gain/Loss			
		Whole Sample of Companies [1]		Sample of Companies without Duplicates [2]	
Benefits	# Marketing Cookies (1)	0.000	(0.001)	0.000	(0.001)
	ln(Expected Monthly Online Ad Revenue) [3]	-0.004	(0.013)	-0.002	(0.014)
Risks	Website Popularity (3 Months pre-GDPR)	0.000*	(0.000)	0.000**	(0.000)
	# Data-Breaches (pre-GDPR)	-0.458***	(0.117)	-0.514**	(0.217)
Email's Content	Monetary Incentives	1.036***	(0.172)	1.153***	(0.187)
	Non-Monetary Incentives	3.035***	(0.190)	2.975***	(0.202)
	Framing: Time Orientation	1.748*	(0.992)	1.999*	(1.077)
	Control	-0.036	(0.048)	-0.051	(0.048)
	Transparency	-0.069*	(0.041)	-0.063	(0.042)
Controls	EU	0.206***	(0.072)	0.204***	(0.079)
	Firm Size	-0.023	(0.018)	-0.017	(0.020)
	Firm Age	0.001	(0.001)	0.001	(0.002)
	Sectors:				
	Media and Entertainment	0.028	(0.131)	0.048	(0.137)
	Professional Services	-0.246**	(0.101)	-0.216**	(0.108)
	Retail Trade	-0.233**	(0.115)	-0.169	(0.123)
	Software and IT Services	-0.397***	(0.120)	-0.360***	(0.130)
	Travel, Tourism and Hospitality	0.006	(0.119)	-0.048	(0.128)
	Advertiser (0/1)=1 [4]	-0.229*	(0.134)	-0.238	(0.154)
	Country Missing	0.366	(0.315)	0.428	(0.314)
	Firm Size Missing	-0.252	(0.187)	-0.216	(0.191)
	Firm Age Missing	0.231	(0.187)	0.152	(0.191)
	Advertiser Missing	-0.309**	(0.157)	-0.271	(0.174)
	Constant	-1.625***	(0.270)	-1.754***	(0.300)
	Observations	1364		1164	
	Log-Pseudolikelihood	-546.54		-463.25	
	AIC	1139.07		972.50	
	BIC	1259.09		1088.87	

Notes:
Standard errors in parentheses
* p<0.10, ** p<0.05, *** p<0.01
The sample size has been reduced because there are companies for which the expected online ad revenue figure was not available.

[1] Results presented in this column are the ones presented in column (3) of the model described in Chapter 7.4.

[2] Results presented in this column are based on 1304 re-permission emails that were sent out by 1304 companies.

[3] We add a small constant term to get around instances of zeroes in this variable as we take logs (Pattabhiramaiah, Sriram and Manchanda 2019).

[4] The "Advertiser" dummy variable is equal to 1 if the company is only an advertiser (N = 962), and 0 if the company is (i) a publisher and advertiser (N = 105); (ii) only a publisher (N = 12); or (iii) missing (N=427). We controlled for the missing values by adding a dummy variable which assume value equal to 1 if the information is missing and 0 otherwise.

Table G.2.4 – Robustness Checks for Fractional Logit Models with DV = "Time Orientation" Theme.

The two columns present the results achieved by using the measurement of, respectively, the whole sample of available emails and the subset of emails sent by companies recorded one time into the database (we removed 6.6% of the companies which sent out more different re-permission emails).

		DV = Framing: Time Orientation			
		Whole Sample of Companies [1]		Sample of Companies without Duplicates [2]	
Benefits	# Marketing Cookies (1)	0.000	(0.000)	-0.000	(0.000)
	ln(Expected Monthly Online Ad Revenue) [3]	0.013***	(0.004)	0.011***	(0.004)
Risks	Website Popularity (3 Months pre-GDPR)	0.000	(0.000)	0.000	(0.000)
	# Data-Breaches (pre-GDPR)	-0.034	(0.033)	-0.091	(0.075)
Email's Content	Monetary Incentives	0.021	(0.055)	0.082	(0.062)
	Non-Monetary Incentives	0.184***	(0.060)	0.096	(0.063)
	Framing: Gain/Loss	0.071	(0.043)	0.081*	(0.047)
	Control	-0.013	(0.013)	-0.017	(0.014)
	Transparency	-0.038***	(0.011)	-0.035***	(0.012)
Controls	EU	-0.013	(0.020)	-0.019	(0.020)
	Firm Size	-0.012**	(0.005)	-0.008	(0.006)
	Firm Age	0.000	(0.000)	0.001**	(0.000)
	Sectors:				
	Media and Entertainment	-0.025	(0.034)	-0.041	(0.036)
	Professional Services	0.013	(0.026)	0.015	(0.029)
	Retail Trade	0.039	(0.030)	0.027	(0.032)
	Software and IT Services	-0.011	(0.030)	0.004	(0.032)
	Travel, Tourism and Hospitality	-0.019	(0.031)	-0.015	(0.034)
	Advertiser (0/1)=1 [4]	-0.023	(0.033)	-0.058	(0.036)
	Country Missing	-0.055	(0.065)	-0.090	(0.067)
	Firm Size Missing	-0.073*	(0.043)	-0.073*	(0.043)
	Firm Age Missing	0.076*	(0.044)	0.119***	(0.041)
	Advertiser Missing	0.002	(0.039)	-0.017	(0.042)
	Constant	-1.926***	(0.072)	-1.890***	(0.077)
	Observations	1364		1164	
	Log-Pseudolikelihood	-378.68		-322.89	
	AIC	803.36		691.79	
	BIC	923.37		808.16	

Notes:
Standard errors in parentheses
* p<0.10, ** p<0.05, *** p<0.01
The sample size has been reduced because there are companies for which the expected online ad revenue figure was not available.

[1] Results presented in this column are the ones presented in column (4) of the model described in Chapter 7.4.

[2] Results presented in this column are based on 1304 re-permission emails that were sent out by 1304 companies.

[3] We add a small constant term to get around instances of zeroes in this variable as we take logs (Pattabhiramaiah, Sriram and Manchanda 2019).

[4] The "Advertiser" dummy variable is equal to 1 if the company is only an advertiser (N = 962), and 0 if the company is (i) a publisher and advertiser (N = 105); (ii) only a publisher (N = 12); or (iii) missing (N=427). We controlled for the missing values by adding a dummy variable which assume value equal to 1 if the information is missing and 0 otherwise.

APPENDIX H – PILOT STUDY: DO DIFFERENT DATA-REQUESTS DRIVE USERS' OPT-IN BEHAVIOR?

This thesis's main focus is the study of firms' behaviors in terms of data-related communication when they are forced to disclose their data-related strategies. Nonetheless, as stated in the limitation section of this thesis (Chapter 8), it could have been interesting to also investigate whether the different communication strategies implemented by companies were effective in prompting consumers' behaviors. Unfortunately, we missed individual-level data about the opt-in rates that each of the re-permission emails collected was able to generate. Therefore, we were not able to make any inference about which of the communication strategies identified – e.g., informative and/or persuasive – was more effective in prompting consumers' disclosure behaviors. In order to support our contention that there exists a relationship between the firm's data request design (e.g., informative and persuasive themes) and the consumer's data disclosure behavior, we decided to conduct two post-tests whereby we manipulate the themes used in privacy communication campaigns and evaluate the customers' intention to disclose.

As stated in the previous chapters of the thesis, academic literature suggests that either the use of persuasive or informative cues in data requests presents its advantages, triggering consumers' personal information disclosure in different ways. Various studies demonstrated that in order to get access to a higher volume of data, it is crucial to carefully evaluate and develop an appropriate firm's communication strategy. For example, it is essential to select the stimuli which actually encourage customers to disclose without triggering reactance and data denial. Understanding which are the levers that a company can exploit in order to alter individuals' privacy behavior is fundamental in a world of empowered customers and may provide valuable insights for managers who can learn how to design their data requests to maximize the probability of gaining data access while preserving firms' reputation. Therefore, we believe that the provision of further empirical results about the extant relation between communication strategies and customers' data access decisions can make the argumentations of this thesis more convincing and robust.

In this appendix, we present some preliminary results that allow understanding whether the various communication stimuli (e.g., experimental conditions) found in our re-permission emails' sample were (i) differently perceived by final recipients and (ii) lead to divergent consumers disclosure behaviors. In the next sections, we describe the experimental conditions and present the results obtained.

H.1. Experimental Conditions

In order to be able to analyze the link between the type of data-related communication content and user's behavior in terms of data access, we decided to create four test conditions in a 2 x 2 between-subject design (see Figure H.1.1). The factors identified regard the text of the data request in terms of presence/absence of persuasive cues and presence/absence of informative content. Therefore, customers will be randomly assigned to one of the following four groups:

- *Control Group*: individuals in this condition will receive a neutral type of data request. This means that the communication contains neither informative nor persuasive elements, and it only asks consumers to provide personal information (Panel A).
- *Treatment 1 – Persuasive Only Group*: individuals in this condition will receive communication which mainly presents persuasive elements in the form of incentives – e.g., the company is willing to provide a discount/and or gift in exchange for data – and of framing – e.g., the company highlights the expected gains of data disclosure – (Panel B).
- *Treatment 2 – Informative Only Group*: individuals in this condition will receive communication that mainly presents informative cues. The message will be designed to be perceived as highly transparent, advising users about the firms' data collection, usage, and sharing. Additionally, in line with the results of our predictive models (see Appendix D, Table

D.1), we also crafted these communications to be wordier since we found that transparent re-permission emails present are indeed extremely long (Panel C).

- *Treatment 3 – Both Persuasive and Informative Group*: individuals in this condition will receive communication that presents both persuasive and informative elements, representing what we can address as the "*combined*" data-related communication (Panel D). The order in which persuasive and informative elements are presented in the message was also randomized.

As previously said, we conducted two distinct post-tests. The main difference between them regards the degree of persuasion of the communications designed, affecting the experimental stimuli seen by individuals in treatment groups 1 and 3. Indeed, messages in Post-Test (2) were constructed to be perceived as more persuasive than the ones in Post-Test (1). This was achieved by making both persuasive elements more visible and relevant in the communication and the length of the persuasive text closer to the one designed for the "*informative*" type of data request – in order to have conditions that do not excessively differ in terms of cognitive effort required. Consequently, also the "*combined*" experimental condition changed from Post-Test (1) to Post-Test (2) since it accommodates the different degrees of persuasion of the communications designed.

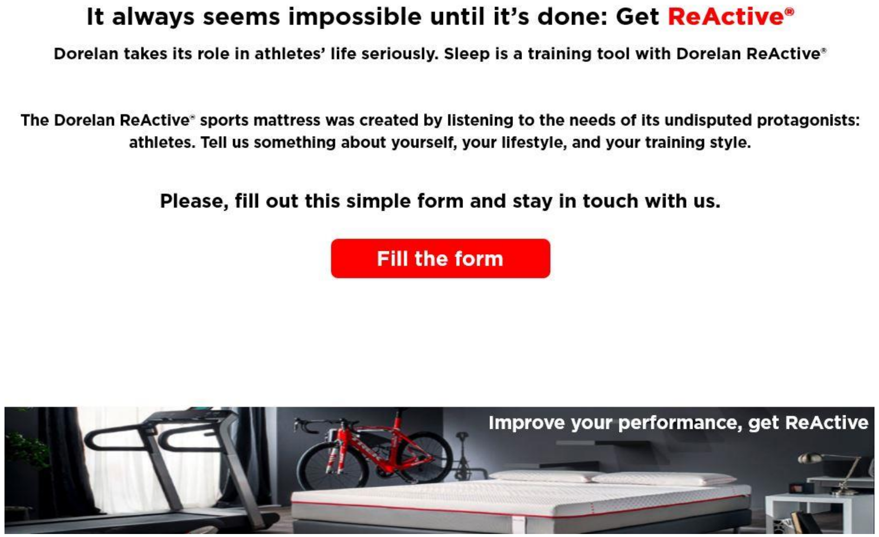
Figures H.1.2 and H.1.3 show the different experimental conditions used in the two post-tests. We crafted the messages' content – in terms of information and persuasion – by inspecting the re-permission emails making up our sample that we classified as persuasive and informative. Therefore, we were inspired by real data requests sent by existing companies on the occasion of the GDPR enforcement. This makes our experimental conditions more plausible and credible.

Figure H.1.1. Experimental Conditions.

		Persuasive Content	
		Absent	Present
Informative Content	Absent	**Control** (Generic message)	**Treatment 1** (Only Persuasive)
	Present	**Treatment 2** (Only Transparent)	**Treatment 3** (Persuasive + Transparent)

Figure H.1.2. Data Requests Designed: Post-Test (1)

Panel A – "Control" Group



It always seems impossible until it's done: Get ReActive®

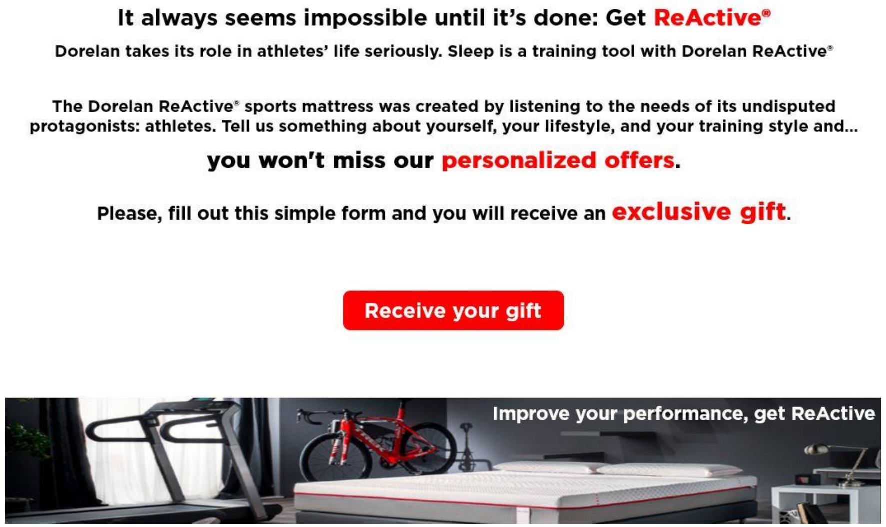
Dorelan takes its role in athletes' life seriously. Sleep is a training tool with Dorelan ReActive®

The Dorelan ReActive® sports mattress was created by listening to the needs of its undisputed protagonists: athletes. Tell us something about yourself, your lifestyle, and your training style.

Please, fill out this simple form and stay in touch with us.

Fill the form

Panel B – "Persuasive Only" Group



It always seems impossible until it's done: Get ReActive®

Dorelan takes its role in athletes' life seriously. Sleep is a training tool with Dorelan ReActive®

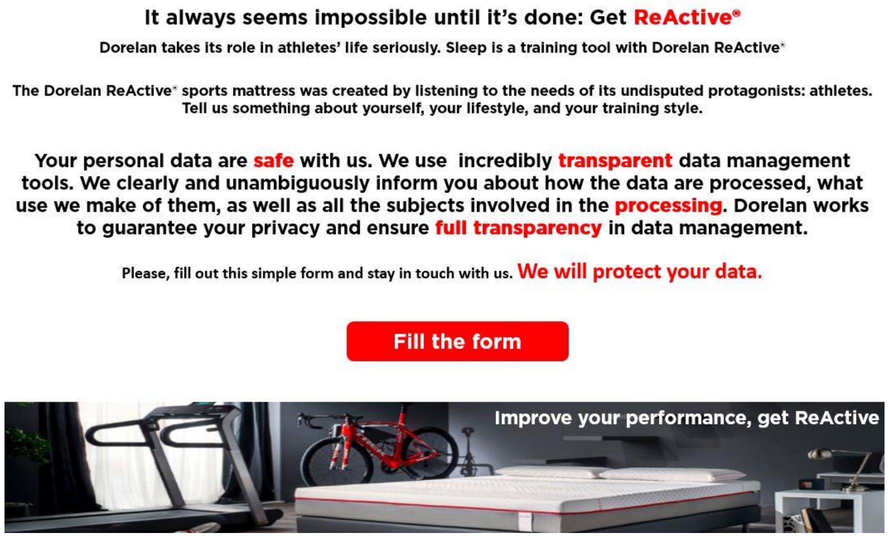
The Dorelan ReActive® sports mattress was created by listening to the needs of its undisputed protagonists: athletes. Tell us something about yourself, your lifestyle, and your training style and...

you won't miss our personalized offers.

Please, fill out this simple form and you will receive an exclusive gift.

Receive your gift

Panel C – "Informative Only" Group



It always seems impossible until it's done: Get ReActive®

Dorelan takes its role in athletes' life seriously. Sleep is a training tool with Dorelan ReActive®

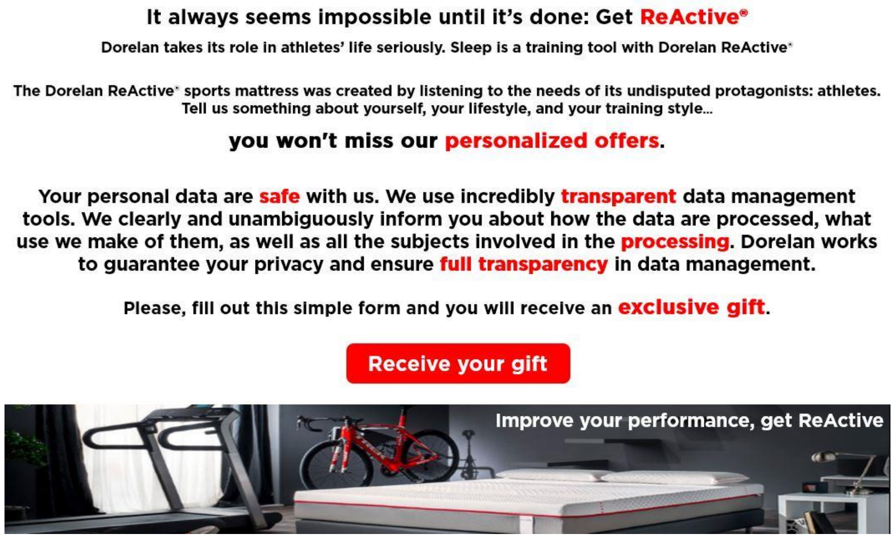
The Dorelan ReActive® sports mattress was created by listening to the needs of its undisputed protagonists: athletes. Tell us something about yourself, your lifestyle, and your training style.

Your personal data are safe with us. We use incredibly transparent data management tools. We clearly and unambiguously inform you about how the data are processed, what use we make of them, as well as all the subjects involved in the processing. Dorelan works to guarantee your privacy and ensure full transparency in data management.

Please, fill out this simple form and stay in touch with us. We will protect your data.

Fill the form

Panel D – "Both Persuasive and Informative" Group



It always seems impossible until it's done: Get ReActive®

Dorelan takes its role in athletes' life seriously. Sleep is a training tool with Dorelan ReActive®

The Dorelan ReActive® sports mattress was created by listening to the needs of its undisputed protagonists: athletes. Tell us something about yourself, your lifestyle, and your training style...

you won't miss our personalized offers.

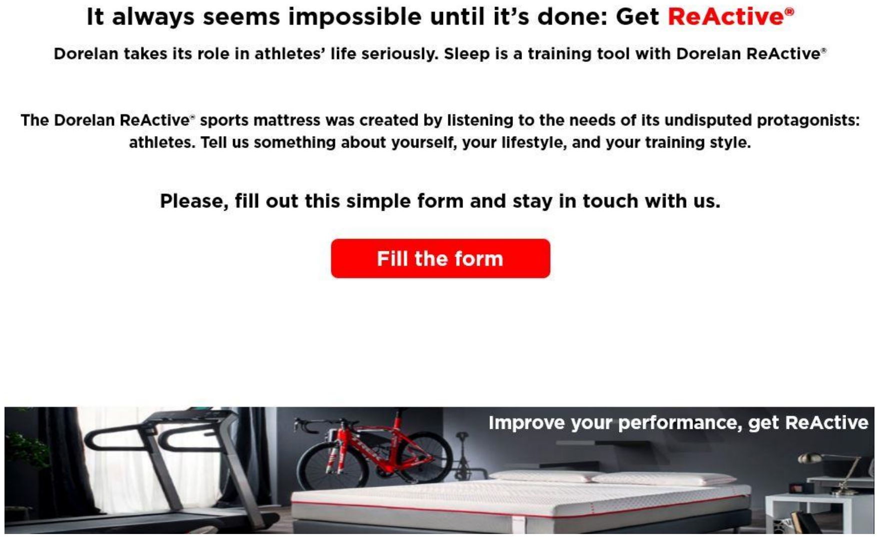
Your personal data are safe with us. We use incredibly transparent data management tools. We clearly and unambiguously inform you about how the data are processed, what use we make of them, as well as all the subjects involved in the processing. Dorelan works to guarantee your privacy and ensure full transparency in data management.

Please, fill out this simple form and you will receive an exclusive gift.

Receive your gift

Figure H.1.3. Data Requests Designed: Post-Test (2)

Panel A – "Control" Group



It always seems impossible until it's done: Get ReActive®

Dorelan takes its role in athletes' life seriously. Sleep is a training tool with Dorelan ReActive®

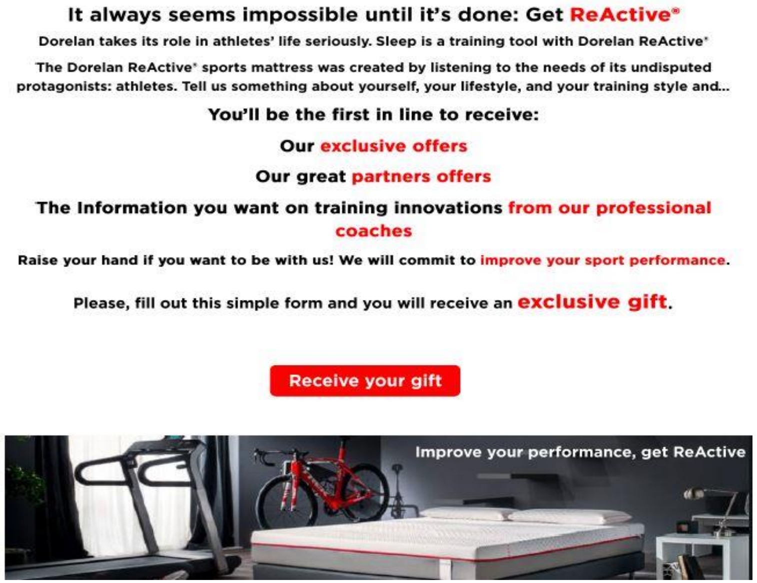
The Dorelan ReActive® sports mattress was created by listening to the needs of its undisputed protagonists: athletes. Tell us something about yourself, your lifestyle, and your training style.

Please, fill out this simple form and stay in touch with us.

Fill the form

Improve your performance, get ReActive

Panel B – "Persuasive Only" Group



It always seems impossible until it's done: Get ReActive®

Dorelan takes its role in athletes' life seriously. Sleep is a training tool with Dorelan ReActive®

The Dorelan ReActive® sports mattress was created by listening to the needs of its undisputed protagonists: athletes. Tell us something about yourself, your lifestyle, and your training style and...

You'll be the first in line to receive:

Our exclusive offers

Our great partners offers

The Information you want on training innovations from our professional coaches

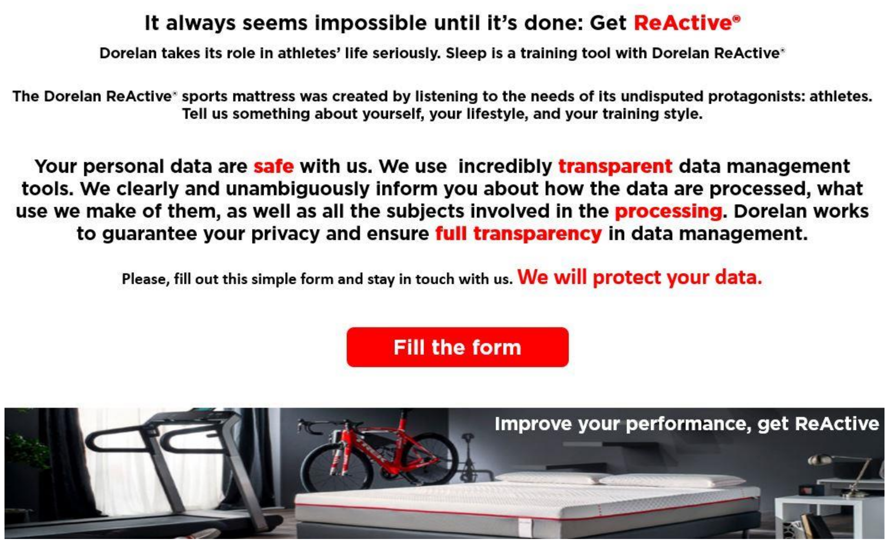
Raise your hand if you want to be with us! We will commit to improve your sport performance.

Please, fill out this simple form and you will receive an exclusive gift.

Receive your gift

Improve your performance, get ReActive

Panel C – "Informative Only" Group



It always seems impossible until it's done: Get ReActive®

Dorelan takes its role in athletes' life seriously. Sleep is a training tool with Dorelan ReActive®

The Dorelan ReActive® sports mattress was created by listening to the needs of its undisputed protagonists: athletes. Tell us something about yourself, your lifestyle, and your training style.

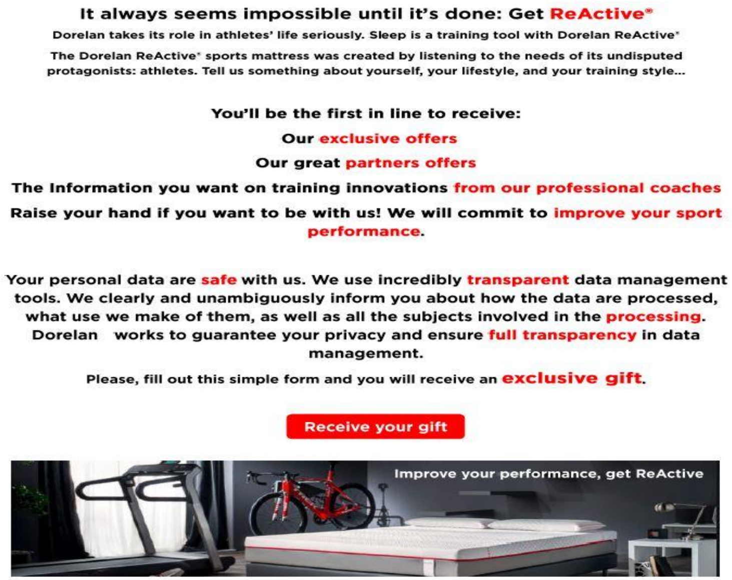
Your personal data are safe with us. We use incredibly transparent data management tools. We clearly and unambiguously inform you about how the data are processed, what use we make of them, as well as all the subjects involved in the processing. Dorelan works to guarantee your privacy and ensure full transparency in data management.

Please, fill out this simple form and stay in touch with us. We will protect your data.

Fill the form

Improve your performance, get ReActive

Panel D – "Both Persuasive and Informative" Group



It always seems impossible until it's done: Get ReActive®

Dorelan takes its role in athletes' life seriously. Sleep is a training tool with Dorelan ReActive®

The Dorelan ReActive® sports mattress was created by listening to the needs of its undisputed protagonists: athletes. Tell us something about yourself, your lifestyle, and your training style...

You'll be the first in line to receive:

Our exclusive offers

Our great partners offers

The Information you want on training innovations from our professional coaches

Raise your hand if you want to be with us! We will commit to improve your sport performance.

Your personal data are safe with us. We use incredibly transparent data management tools. We clearly and unambiguously inform you about how the data are processed, what use we make of them, as well as all the subjects involved in the processing. Dorelan   works to guarantee your privacy and ensure full transparency in data management.

Please, fill out this simple form and you will receive an exclusive gift.

Receive your gift

Improve your performance, get ReActive

H.2. Manipulation Checks

We evaluate the recipients' perceived degree of persuasion and information of the communication received as well as their subsequent intention to disclose personal information. This will allow us to assess whether the communications were differently perceived (informative vs. persuasive) and to provide evidence that the use of different communication stimuli drives divergent consumers' disclosure behavior. We asked our respondents to answer different sets of questions (aimed at measuring our constructs of interest) on a 5-point Likert-type scale ranging from (1) "Strongly Disagree" to (5) "Strongly Agree". Table H.2.1 shows the list of items used to measure our constructs.

Table H.2.1 – Measurement of our Constructs of Interest.

Construct	**Item**
Informative: Transparency (adapted from Wu et al. 2012)	This company… • explains how the collected personal information will be used. • informs whether personal information will be disclosed to a third party. • protects my personal information. • is transparent about the use of my personal information. • gives me clear information about how my personal data are treated.
Persuasive: Incentives and Framing (adapted from Sen, Gürhan-Canli, and Morwitz 2001; Martin, Borah, and Palmatier 2017)	This company… • provides a clear incentive for sharing personal information. • rewards me for sharing my personal information. • stresses the positive implications of sharing personal information. • stresses the negative implications of not sharing personal information. • explains to me the benefits of filling the form.
Behavioral Intention	How likely are you to fill this form and share your personal information with this company?

H.3. Post-Test Results

The questionnaires were developed on Qualtrics and distributed to a Prolific Panel of respondents at two points in time: on the 18th of March 2021 and on the 2nd of April 2021. 188 individuals completed the first survey, whereas 191 completed the second one. All the respondents get rewarded with a small monetary incentive to complete the task. Table H.3.1 shows the distribution of the respondents in our experimental groups for the two post-tests.

In order to get an overall degree of perceived persuasion and information for each communication, we performed factor analyses on the items measuring the different constructs. Table H.3.2 presents descriptive statistics for the items measuring our constructs of interests as well as the scree plots from the factor analyses, which indicate the selection of just one factor (Kaiser's Criterion) for each set of items. As it is possible to see by inspecting the tables of the summary statistics, there is one item measuring persuasion (Pers4), which has been rated low in terms of persuasion. Notably, this item's goal was to measure the individuals' perception of the presence of a loss-type of framing, a condition that was not inserted in any of the communications shown. Therefore, this result is in line with our expectation and suggests that users carefully read the messages proposed to them and interpret the communications in the correct way.

Table H.3.3 and Table H.3.4 show the results of the ANOVA analyses performed on the informative and persuasive constructs (previously described) as well as the contrasts between the means of the different treatment groups and the Control Group (we performed pairwise comparisons using Dunnett's Multiple Comparison Test). As desired, our results suggest that respondents were able to discriminate between a persuasive and informative type of message. Persuasive messages were perceived as less informative and more persuasive, while informative messages were perceived as less persuasive and more informative. Moreover, our combined version of the data request – the one including both informative and persuasive elements) – reached high values both in terms of persuasion and information. Lastly, the "*generic*" version of the message was recognized as neither highly informative nor highly persuasive, in line with our intention. Therefore, from this first type of analysis, we can conclude that consumers correctly perceived the stimuli of the messages sent to them, providing initial support for the contention that companies may have intentionally designed their data requests as they wanted them to be perceived by their recipients.

Moreover, we also wanted to provide corroboration to the argument there are differences in the way in which the different messages prompted users' disclosure behavior. Therefore, we run ordered logit models to test how the different communications strategies relate to the individuals' behavioral intention. Results are presented in Table H.3.5.

From the first model, we can see that providing a message with only informative content decreases the likelihood that consumers provide data ($Informative_1 = -0.842$), whereas the use of persuasion does not play a significant role in terms of data provision – compared to a generic type of data request. This result held even when we increased the degree of persuasion in the messages (Post-Test (2)), since the coefficient for "*Informative Only*" is still significant, even if marginally ($Informative_2 = -0.635$). Therefore, it seems that users are incredibly warned about their data and companies' security standards and tend to react adversely to highly transparent data-related communications. This suggests that data-related communications are likely to make privacy more salient, evocating feelings of concerns and reactance in users, who are less willing to disclose personal information. This provides initial support for our contention that the GDPR requirement of designing merely informative re-permission emails could have led companies to experience severe data losses, inducing them to try to find other *workarounds* to get data.

Table H.3.1 – Distribution of the Respondents into the Experimental Conditions.

Post-Test (1) – N = 189

		Persuasive Content	
		Absent	Present
Informative Content	Absent	**Control** (Generic message) **N = 48 (25.4%)**	**Treatment 1** (Only Persuasive) **N = 48 (25.4%)**
	Present	**Treatment 2** (Only Transparent) **N = 44 (23.3%)**	**Treatment 3** (Persuasive + Transparent) **N = 49 (25.9%)**

Post-Test (2) – N = 191

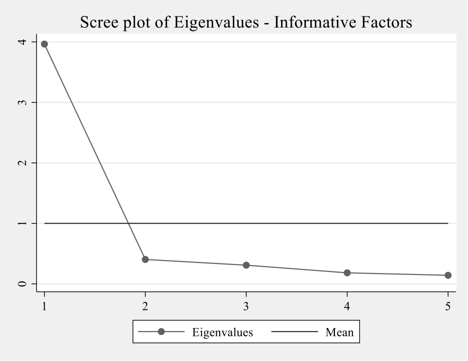
		Persuasive Content	
		Absent	Present
Informative Content	Absent	**Control** (Generic message) **N = 47 (24.6%)**	**Treatment 1** (Only Persuasive) **N = 49 (25.7%)**
	Present	**Treatment 2** (Only Transparent) **N = 46 (24.0%)**	**Treatment 3** (Persuasive + Transparent) **N = 49 (25.7%)**

Table H.3.2 – Descriptive Statistics for the Items Evaluating Informative and Persuasive Messages and Scree Plot from Factor Analysis (continue on the next page).

Post-Test (1)

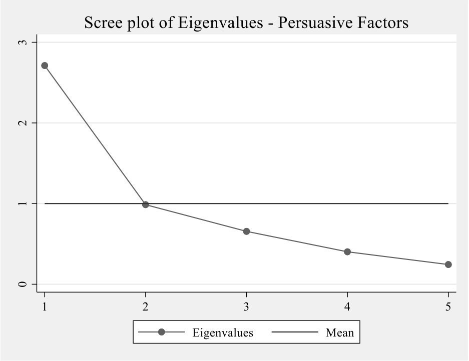
Informative Items

The Company…	**Variable**	**Obs**	**Mean**	**Std. Dev.**	**Min**	**Max**
Explains how the collected personal information will be used.	Trasp1	189	2.69	1.30	1	5
Informs whether personal information will be disclosed to a third party.	Trasp2	189	2.44	1.27	1	5
Protects my personal information.	Trasp3	189	2.85	1.09	1	5
Is transparent about the use of my personal information.	Trasp4	189	2.82	1.31	1	5
Gives me clear information about how my personal data are treated.	Trasp5	189	2.71	1.32	1	5



Persuasive Items

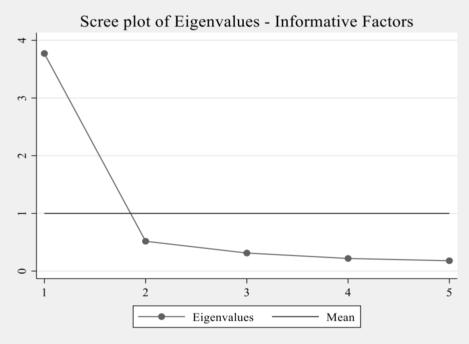
The Company…	**Variable**	**Obs**	**Mean**	**Std. Dev.**	**Min**	**Max**
Provides a clear incentive for sharing personal information.	Pers1	189	3.17	1.26	1	5
Rewards me for sharing my personal information.	Pers2	189	3.21	1.35	1	5
Stresses the positive implications of sharing personal information.	Pers3	189	3.05	1.19	1	5
Stresses the negative implications of not sharing personal information.	Pers4	189	2.01	0.99	1	5
Explains to me the benefits of filling the form.	Pers5	189	3.33	1.17	1	5



Post-Test (2)

Informative Items

The Company…	Variable	Obs	Mean	Std. Dev.	Min	Max
Explains how the collected personal information will be used.	Trasp1	191	2.80	1.32	1	5
Informs whether personal information will be disclosed to a third party.	Trasp2	191	2.52	1.32	1	5
Protects my personal information.	Trasp3	191	2.86	1.10	1	5
Is transparent about the use of my personal information.	Trasp4	191	2.80	1.29	1	5
Gives me clear information about how my personal data are treated.	Trasp5	191	2.64	1.29	1	5



Persuasive Items

The Company…	Variable	Obs	Mean	Std. Dev.	Min	Max
Provides a clear incentive for sharing personal information.	Pers1	191	3.14	1.28	1	5
Rewards me for sharing my personal information.	Pers2	191	3.19	1.31	1	5
Stresses the positive implications of sharing personal information.	Pers3	191	3.15	1.20	1	5
Stresses the negative implications of not sharing personal information.	Pers4	191	2.08	0.95	1	4
Explains to me the benefits of filling the form.	Pers5	191	3.52	1.14	1	5

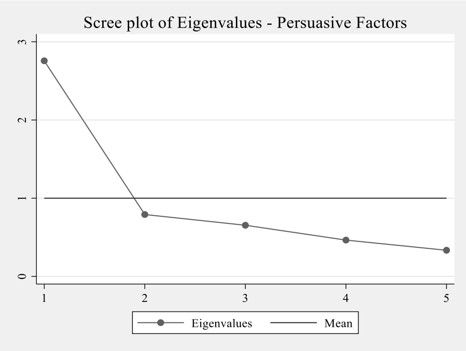
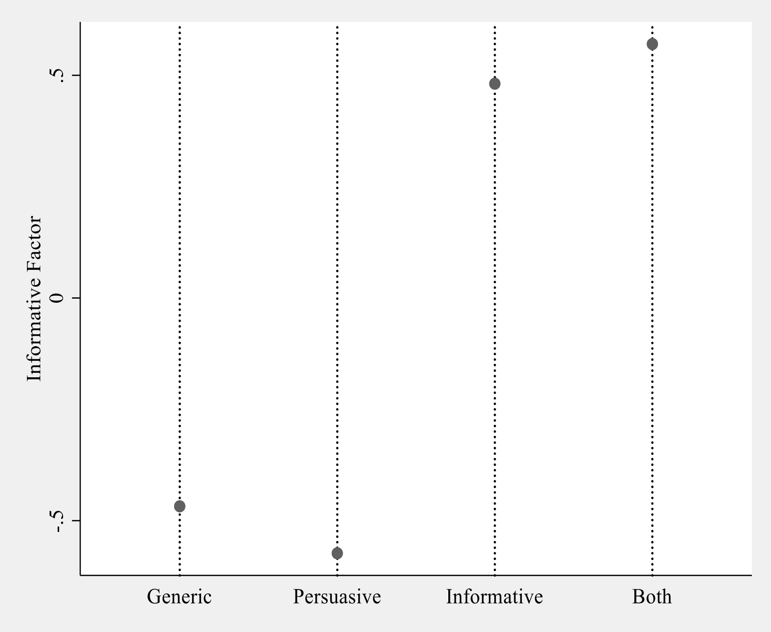


Table H.3.3 - ANOVA on Informative Factor and Pairwise Comparisons (continue on the next page).

Post-Test (1)

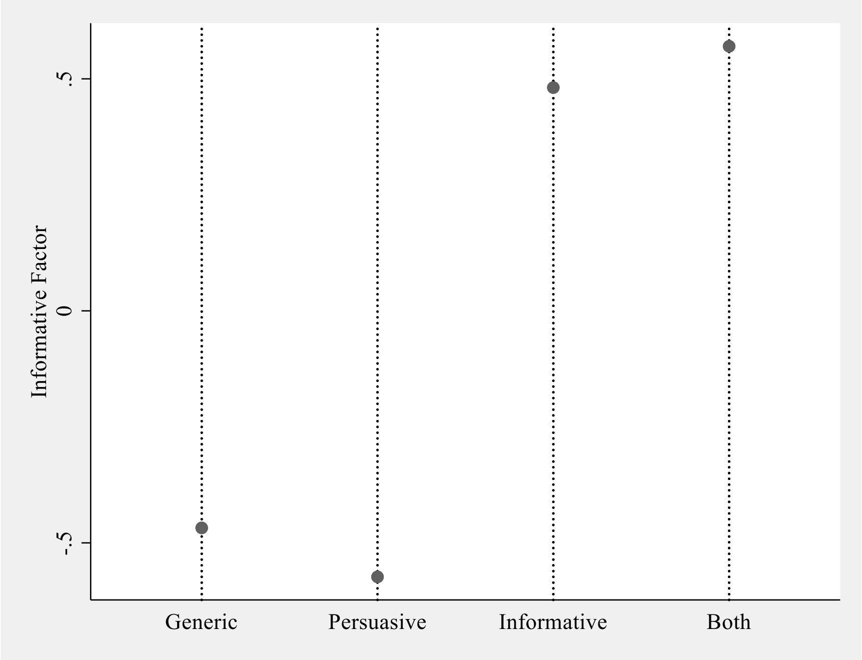


Groups	**Mean**	**Std. Dev.**	**Freq.**
Generic	-0.625	0.749	48
Persuasive	-0.625	0.717	48
Informative	0.629	0.725	44
Combined	0.660	0.898	49
Total	**0.000**	**1.000**	**189**

			Analysis of Variance		
Source	**SS**	**df**	**MS**	**F**	**Prob > F**
Between groups	76.238	3.000	25.413	42.070	0.000
Within groups	111.762	185.000	0.604		
Total	**188.000**	**188.000**	**1.000**		

				Dunnett		
Informative Factor	**Contrast**	**Std. Err.**	**t**	**P>t**	**[95% Conf.**	**Interval]**
Groups:						
Persuasive vs Generic	0.000	0.159	0.000	1.000	-0.376	0.376
Informative vs Generic	**1.254**	0.162	7.730	0.000	0.870	1.638
Combined vs Generic	**1.285**	0.158	8.140	0.000	0.911	1.659

Post-Test (2)



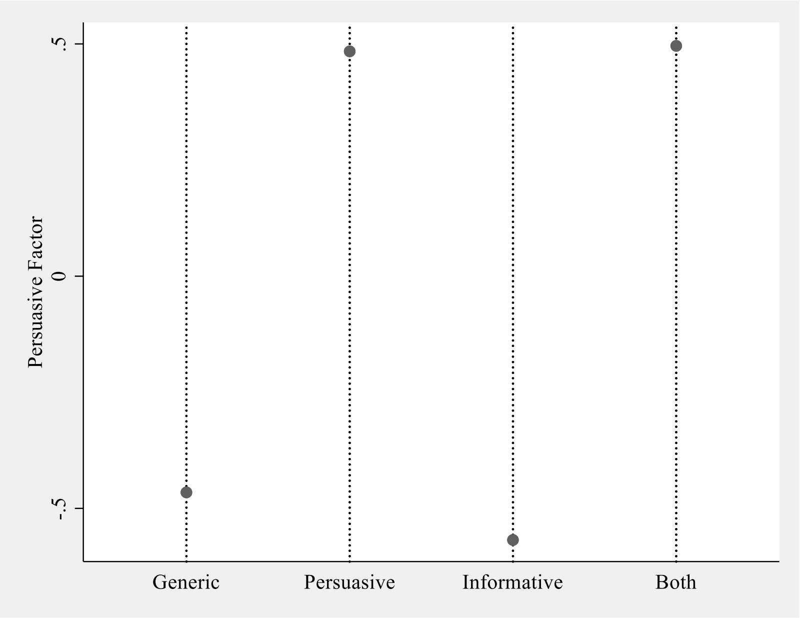
Groups	**Mean**	**Std. Dev.**	**Freq.**
Generic	-0.468	0.780	47
Persuasive	-0.573	0.775	49
Informative	0.481	1.018	46
Combined	0.570	0.836	49
Total	**0.000**	**1.000**	**191**

			Analysis of Variance		
Source	**SS**	**df**	**MS**	**F**	**Prob > F**
Between groups	52.962	3.000	17.654	24.090	0.000
Within groups	137.038	187.000	0.733		
Total	**190.000**	**190.000**	**1.000**		

			Dunnett			
Informative Factor	**Contrast**	**Std. Err.**	**t**	**P>t**	**[95% Conf.**	**Interval]**
Groups:						
Persuasive vs Generic	-0.105	0.175	-0.600	0.931	-0.559	0.348
Informative vs Generic	**0.949**	0.178	5.340	0.000	0.489	1.409
Combined vs Generic	**1.038**	0.175	5.940	0.000	0.585	1.491

Table H.3.4 - ANOVA on Persuasive Factors and Pairwise Comparisons (continue on the next page).

Post-Test (1)

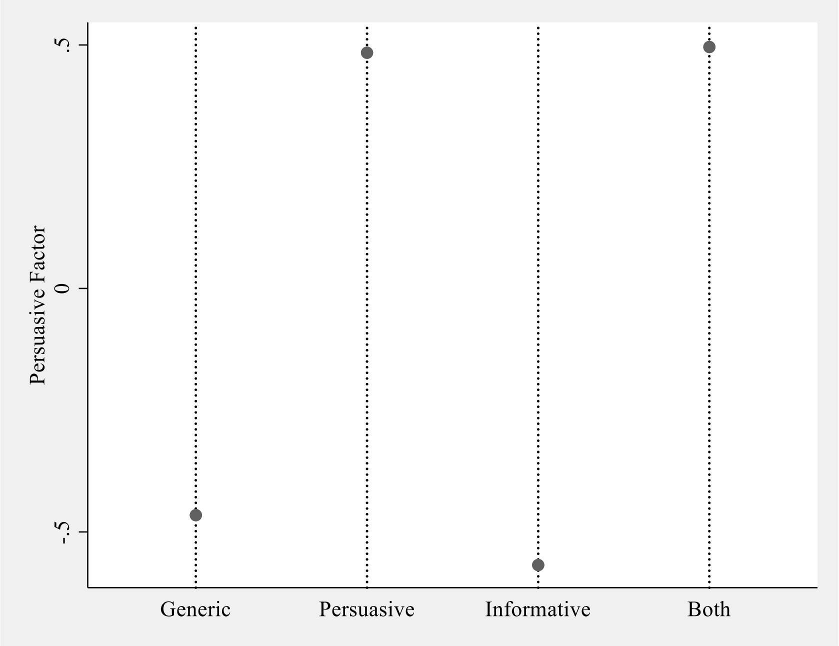


Groups	**Mean**	**Std. Dev.**	**Freq.**
Generic	-0.570	0.966	48
Persuasive	0.613	0.636	48
Informative	-0.467	0.821	44
Combined	0.377	0.979	49
Total	**0.000**	**1.000**	**189**

			Analysis of Variance		
Source	**SS**	**df**	**MS**	**F**	**Prob > F**
Between groups	50.205	3.000	16.735	22.470	0.000
Within groups	137.795	185.000	0.745		
Total	**188.000**	**188.000**	**1.000**		

			Dunnett			
Persuasive Factor	**Contrast**	**Std. Err.**	**t**	**P>t**	**[95%**	**Conf. Interval]**
Groups:						
Persuasive vs Generic	**1.183**	0.176	6.720	0.000	0.766	1.600
Informative vs Generic	0.103	0.180	0.570	0.894	-0.324	0.530
Combined vs Generic	**0.947**	0.175	5.400	0.000	0.532	1.362

Post-Test (2)



Groups	**Mean**	**Std. Dev.**	**Freq.**
Generic	-0.466	0.916	47
Persuasive	0.484	0.879	49
Informative	-0.568	0.886	46
Combined	0.496	0.796	49
Total	**0.000**	**1.000**	**191**

			Analysis of Variance		
Source	**SS**	**df**	**MS**	**F**	**Prob > F**
Between groups	48.555	3.000	16.185	21.400	0.000
Within groups	141.445	187.000	0.756		
Total	**190.000**	**190.000**	**1.000**		

					Dunnett	
Persuasive Factor	**Contrast**	**Std. Err.**	**t**	**P>t**	**[95%**	**Conf. Interval]**
Groups:						
Persuasive vs Generic	**0.950**	0.178	5.350	0.000	0.489	1.410
Informative vs Generic	-0.102	0.180	-0.570	0.942	-0.570	0.365
Combined vs Generic	**0.961**	0.178	5.410	0.000	0.501	1.422

Table H.3.5 – Ordered Logit Models – DV = Intention to Disclose (5 points Likert scale)

	Intention to Disclose					
	Post-Test (1)			Post-Test (2)		
Informative Only	-0.842	**	(0.398)	-0.635	*	(0.384)
Persuasive Only	0.012		(0.378)	0.229		(0.363)
Informative & Persuasive	0.754		(0.537)	0.161		(0.529)
cutoff1	-1.169	***	(0.298)	-0.711	***	(0.27)
cutoff2	0.542	***	(0.288)	0.631	***	(0.267)
cutoff3	0.912	***	(0.292)	1.147	***	(0.278)
cutoff4	2.945	***	(0.428)	3.503	***	(0.504)
Observations	188			191		
Log lik.	-255.76			-259.77		
Wald Chi-squared	6.61			5.54		
Model	Ordered Logit			Ordered Logit		