The Race for Data: Utilizing Informative or Persuasive Cues to Gain Opt-in?

Published Version:
D'Assergio, C., Manchanda, P., Montaguti, E., Valentini, S. (2025). The Race for Data: Utilizing Informative or Persuasive Cues to Gain Opt-in?. JOURNAL OF MARKETING, 89(3), 36-59 [10.1177/00222429241288456].

Availability:
This version is available at: https://hdl.handle.net/11585/998128 since: 2024-12-05

Published:
DOI: http://doi.org/10.1177/00222429241288456

DOI: 10.1177/00222429241288456

JOURNAL of Marketing

The Race for Data: Utilizing Informative or Persuasive Cues to Gain Opt-in?

Journal:	*Journal of Marketing*
Manuscript ID	JM-22-0596.R3
Manuscript Type:	Revised Submission
Research Topics:	Privacy, Public Policy and Marketing
Methods:	Field Experiments, Text Analysis, Natural Language Processing, Regression Models

The Race for Data: Utilizing Informative or Persuasive Cues to Gain Opt-in?

Caterina D'Assergio

Puneet Manchanda

Elisa Montaguti

Sara Valentini

Caterina D'Assergio, Consultant, Information Technology Department, Marazzi Group Srl, Viale Regina Pacis, 39, 41049 Sassuolo (MO), Italy, email: caterina.dassergio@gmail.com

Puneet Manchanda, Isadore and Leon Winkelman Professor of Marketing, Ross School of Business, University of Michigan, R5490, 701 Tappan Street, Ann Arbor, MI 48109 Phone: 734–936–2445 Fax: 734–936–8716, email: pmanchan@umich.edu

Elisa Montaguti, Full Professor of Marketing, Management Department, University of Bologna, Italy, Via Capo di Lucca, 34, 40126 Bologna, Italy, Phone (+39) 051 209 8089, Fax: (+39) 051 246 411, email: elisa.montaguti@unibo.it

Sara Valentini, Associate Professor, Marketing Department, Bocconi University, Italy Via Röntgen, 1, 20136 Milano, Italy Phone: (+39) 02 58362487 Email: s.valentini@unibocconi.it

The authors thank Varad Deolankar, David Faro, Aishwarya Joshi, Garrett Johnson, Xingyi Li, Catherine Tucker, Qiaoni Shi, and Kai Zhu for their feedback and advice on this project. The authors also thank the participants at the 2022 "Customer Journeys in a Digital World" Conference at Bocconi Univesity, the special session on privacy at EMAC 2021, 2021 EMAC Digital Marketing SIG, and the 2020 AIML Conference. Finally, the authors also thank the company they cooperated with in Study 2.

The Race for Data: Utilizing Informative or Persuasive Cues to Gain Opt-in?

Abstract

The EU's General Data Protection Regulation (GDPR) mandates explicit user opt-in consent for data access. It recommends transparency in opt-in requests about data collection, storage, and use, without specifying the format of these requests. Consequently, the GDPR gives firms flexibility in designing opt-in messages. This research uses theory, multiple datasets, and methods to investigate firms' communication formats for opt-in requests, addressing three questions: 1) how do firms design their opt-in requests? 2) does the chosen format affect consumer response? 3) what drives firms' choices of formats? The analysis of 1,506 re-permission emails from 1,396 firms post-GDPR shows that 26% use only persuasive cues to request data, while 24% blend persuasive and informative cues. Notably, businesses with an offline presence use more persuasive cues compared to purely digital entities. A field experiment rationalizes this behavior showing that informative cues alone did not improve opt-in; a mix of persuasive and informative cues proved more successful. Additionally, firms dependent on personal data utilize persuasive cues more often than firms concerned with reputational risks of GDPR non-compliance. This study offers pivotal insights for regulators, firms, and consumers, revealing variations in how different firms acquire consent and the impact of their strategies on user behavior.

Keywords: privacy, opt-in, GDPR, persuasion, regulatory compliance, user consent, text analysis, field experiment, cookies, online ad revenues, transparency.

In an increasingly data-intensive business landscape, relationships between firms and consumers show a mixture of firm- and consumer-interested behaviors. Notably, some firms strive to persuade users to grant them access to personal information, as it can support targeting and personalization, leading to higher revenues and profits. For example, Under Armour offered a 25% discount on a subsequent order as a reward to consumers that granted the company access to their personal data, whereas LinkedIn highlighted the benefits that users could derive from relinquishing their data and the costs for denying that access. In contrast, other firms provide clear and transparent information in managing data to establish or strengthen their relationship with consumers (Bleier, Goldfarb, and Tucker 2020). For example, companies like the AA, exemplify the approach of providing clear and transparent information, as demonstrated by the AA Privacy Notice, which offers a comprehensive and user-friendly explanation of their data processing methods (Chaffey 2018). For these firms, their straightforward messages typically communicate how they intend to use consumer data, while clarifying how users can retain control over their personal information. This anecdotal evidence confirms what industry watchers believe: that firms use very different strategies to craft their requests to gain access to personal data (Davis 2018). In this paper, we focus on firms' alternative strategies to gain access to data, as well as to monitor their effectiveness and identify their drivers, with a view to revealing the implications for consumers, firms, and regulators.

The European General Data Protection Regulation (hereafter GDPR) is the context for our study. Central to the GDPR is "opt-in," a data consent mechanism for which individuals must actively agree to have their personal data collected, used, or shared. GDPR, therefore, represents an ideal context because it *requires all firms* processing data of EU individuals to ask users for data consent (opt-in), regardless of a firm's size or location. While GDPR mandates that firms ask for users to opt-in, it doesn't prescribe a specific format for these data consent requests. This presents firms with a unique challenge and opportunity: while they would like to secure user consent, they also need to adhere to the regulatory mandate.

One of the fundamental guiding principles of GDPR is to ensure fairness and transparency in communications; hence, information that firms give individuals must be clear, honest, and easily understandable. More specifically, regulators require firms to "…use a concise, transparent, intelligible and easily accessible form…" (Article 12, GDPR), However, forcing firms to ask for user consent, as the GDPR does, might not solve privacy challenges entirely, as the act of asking for permission can hide opportunities to encourage customers to opt-in, particularly when "how" to ask is not regulated in terms of format and language. Although firms are expected to be transparent, they could interpret the GDPR's freedom of format to their advantage by including persuasive cues in their messages to encourage opt-in. Indeed, the business press highlights that expecting firms to change and comply with the regulator's mandate is near-impossible for some companies because tracking users is central to their respective business models (Thornhill 2022).

In this study, our primary focus is investigating the format of the GDPR opt-in requests. In May 2018, both European and non-European firms sent a large number of emails to their customers, asking or re-asking permission for data use. These are often referred to as re-permission emails. Note that although all firms seek to obtain data consent, their approaches reveal a variety of strategies. We have firms whose messages are primarily characterized by informative cues. These cues are elements within the communication that detail how personal data is collected, treated, and what actions users can take. We define a message as "informative" when it contains these cues and ensures clear and complete communication about personal data management, thereby adhering to the GDPR's core principles of transparency. We also have firms employing messages characterized by persuasive cues. These cues emphasize the desirability of data exchange by providing rewards or framing it as beneficial to gain advantages or avoid potential losses associated with the service or product. In this context, we define a message as "persuasive" if it encourages users to opt-in through the allure of rewards and/or strategic framing. Additionally, there is a vast middle ground where firms blend informative and persuasive elements in their requests for opt-in.

Figure 1: The Race for Data Spectrum: How do Firms ask for Data and Why do they do so?

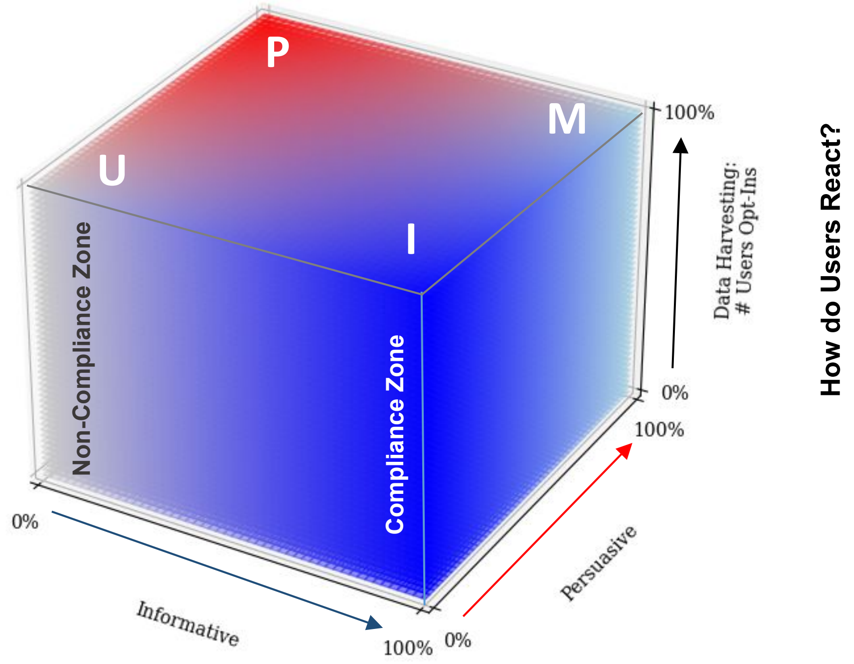


Figure 1 visually depicts the spectrum of data request strategies used by firms. The x-axis tracks the extent to which messages incorporate informative cues and adhere to the tenet of the GDPR, spanning from 0%, which indicates no informative cues, to 100%, which indicates full compliance with GDPR (i.e. providing complete and clear information on how data are collected, stored and used, and return control of data to users as per art. 12 of the GDPR). Similarly, the y-axis assesses how much messages rely on persuasive cues, ranging from 0% to 100%. Finally, the z-axis shows the user opt-in rates, gauging the effectiveness of these strategies, where the lower end indicates no users opted in and the upper end indicates all contacted users opted in.

The letters *I* (Informative), *P* (Persuasive), *M* (Mixed), and *U* (Uninformative & Unpersuasive) at the corners of Figure 1 indicate four different possible opt-in request strategies employed by firms. *I* represents opt-in request strategies that use only informative cues (100% on the x-axis) and do not rely on persuasion (0% on the y-axis). At the opposite corner, *P* represents messages that use only persuasive cues (0% on the x-axis, and 100% on the y-axis). Notably, including persuasive cues is not necessarily in violation of the regulation, but neglecting to use informative cues is.[1] Blended opt-in requests including both informative and persuasive cues are also possible and are represented by *M*, which can theoretically reach 100% on both axes, but depending on the combination of cues, they can be more or less central in Figure 1. Finally, *U* is also a possible scenario representing opt-in requests that are neither informative nor persuasive. Each strategy can be more or less effective in terms of users' opt-ins (z-axis).

I opt-in requests, adhering strictly to GDPR requirements, fall in the "compliance zone", by using 100% informative cues. By contrast, both *P* and *U* fall into the non-compliance zone and might incur the risk of penalties reaching up to 4% of global revenues, as per Article 83, as they use informative cues insufficiently. For mixed strategies *M*, being compliant depends on the number of informative cues included.

In our study we show that: (1) 26% of firms in our sample employed exclusively persuasive tactics (*P*), while 24% blended persuasive with informative cues (*M*), leveraging the freedom of format provided by the GDPR. (2) We show that there is firm heterogeneity in how these businesses use these tactics. Specifically, firms with an offline presence or those selling physical products are more likely to use persuasive tactics compared to their purely

[1] This work's focus is not on defining compliance (in a strictly legal sense) but rather on understanding which opt-in request formats have been adopted by firms.

digital counterparts. (3) Our field experiment reveals that message format matters: while a moderate but exclusive presence of persuasive cues can effectively increase opt-ins, the combination of persuasive and informative cues in the same message consistently shows enhanced opt-ins across different settings.[2] Consumers particularly respond well to persuasive cues involving monetary incentives, indicating a willingness to trade money for data. In contrast, messages that are solely informative do not yield additional opt-ins. (4) Finally, we find that firms with low reputational risks associated with non-compliance and high potential benefits from data use - such as expected online ad revenue and enhanced data harvesting capabilities - are more likely to adopt persuasive content.

Our research adds to the growing body of literature focusing on GDPR's impact. We show that the opt-in request format matters and plays a critical role in user decisions regarding opt-in. Our analysis is broad in nature, spanning multiple industries and product categories. This breadth is important as the efficacy of different opt-in messages varies across these units. We also add to the literature on the effect of data disclosure, showing that a mix of informative and persuasive cues is superior to purely informative elements in triggering user opt-in.

Finally, our research also contributes to the broader field of regulatory interventions by providing insights for all members of the ecosystem – firms, consumers, and policymakers. Firms get insights into how they should respond to regulatory changes, specifically how to design their data requests to maximize data access. This study highlights consumers' willingness to trade data for monetary rewards, enabling firms to fine-tune their data monetization approaches. Consumers should exercise caution and vigilance when responding to data requests, as firms often utilize persuasive cues to obtain data consent.

---

[2] The effectiveness of each messaging strategy was statistically tested against a generic message that does include neither informative nor persuasive cues. A lab experiment detailed in Web Appendix C corroborates the effectiveness of combining persuasive and informative cues.

Consumers might also advocate for fair compensation when their data is exchanged for monetary rewards.

For policymakers, the specifics of how firms respond to regulation provide valuable insights. This is more relevant in this setting, where the regulations provide a lot of discretion to firms. Our results suggest that giving firms latitude in terms of actions leads to them acting in their self-interest. Lastly, we equip policymakers with an accessible and scalable method to analyze firms' data requests. By identifying patterns and tactics utilized by firms in their data requests, regulators can develop more effective mechanisms to monitor compliance, ultimately promoting an environment where consumers' rights are more robustly protected.

Institutional Setting

The legislative scenario for personal data protection has evolved drastically in the last few years. The two most prominent pieces of regulation implemented in the Western world are the GDPR in Europe and the California Consumer Privacy Act (CCPA) in North America (which has been amended and updated in 2023 as the California Privacy Rights Act (CPRA)). Although the GDPR is designed for European firms, it also affects companies operating outside Europe that must comply with the European GDPR if they collect data from EU citizens and residents. Furthermore, many firms have realized that the cost of determining which website visitors are subject to GDPR costs more than applying it to all visitors irrespective of their origin. Multiple studies and business publications corroborate the notion that GDPR is the initial privacy regulation to directly impact US companies (e.g., Marthews and Tucker 2019; Ovcharenko 2021).

Although the GDPR and the CCPA differ in some notable ways, the GDPR is considered stricter than the CCPA. For instance, the GDPR requires firms to gain “explicit opt-in,” which means gaining users’ consent before accessing their personal data. By contrast, the CCPA only requires “opt-out.” In addition, while GDPR applies to all firms operating in the EU, CCPA/CPRA applies to only firms with more than 100,000 employees.

The GDPR, which went into effect on May 25, 2018, *applies to all organizations that process the personal data of individuals in the European Union*, regardless of an organization's size, which means that it is also relevant for small or micro businesses if they process personal data as a regular part of their business operations.

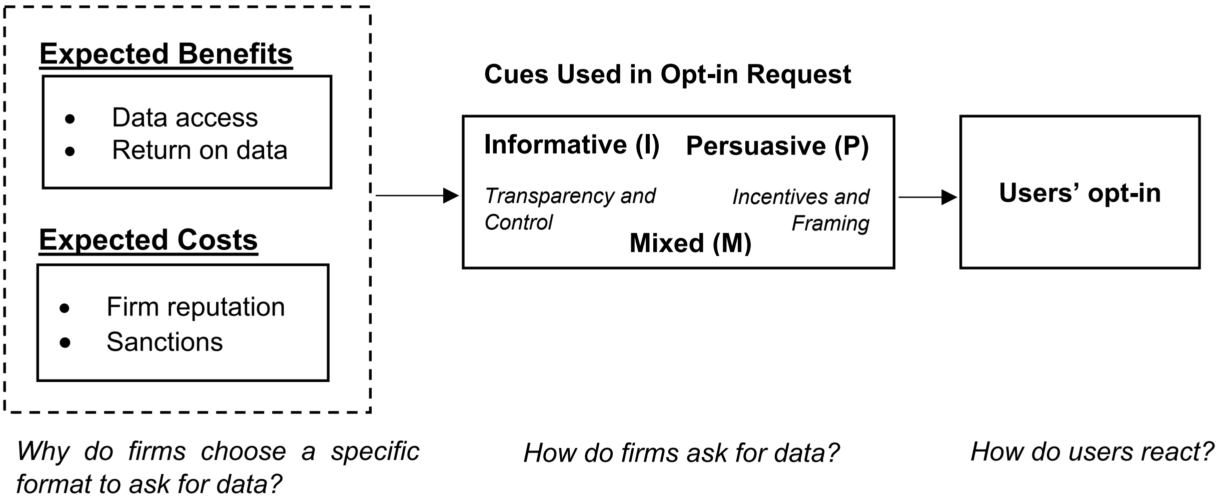
To comply with GDPR, firms sent a large number of emails around May 2018 to inform their customers (including prospective customers) about these fundamental changes and to ask or re-ask their permission to use their data. One of the objectives of this regulation is to ensure transparency, implying that communications and information provided to individuals must be complete, clear, easily understandable, and accessible. However, the regulation fell short of imposing a specific format for re-permission emails. Therefore, firms were free to devise formats for these emails. Although the phenomenon of opt-in emails "exploded" in 2018, asking permission to use personal data remains a continuing challenge. For example, firms must still collect opt-ins when dealing with prospects, acquiring new customers, offering subscriptions to newsletters, and so forth. As repeatedly noted in the business press (e.g., Harford 2022; Murgia 2022; Rawnsley 2022), the process of obtaining opt-ins and gaining the possibility of tracking individuals' behavior continues to be an ongoing important business and social phenomenon.

Theoretical Background and Conceptual Framework

Under the GDPR, firms must obtain users' explicit consent to collect their data (Art. 7 GDPR). Our study delves into the process of obtaining users' opt-in, which is closely related to the literature investigating the factors influencing consumers' data disclosure. Notably, consumer data disclosure refers to the voluntary act of sharing or providing personal information with a company. The opt-in represents a particular approach for eliciting data disclosure, wherein users must give explicit consent at the beginning of the data exchange to a set of rules that regulate how information is collected, used, and shared (Johnson, Bellman, Lohse 2002). To the best of our knowledge, Godinho de Matos and Adjerid's work (2021) is the only study to have examined how consumers and markets respond to enhanced consent (i.e., re-permission requests) by focusing on a single firm. They find that the GDPR's "enhanced consent" did not decrease the amount of data collected, at least for the least sensitive data. Our study explores how the opt-in request is formulated, how consumers respond to different request formats, and why firms make choices about formats, across a very large number of firms.

We ground our theoretical development on the conceptual framework depicted in Figure 2. This framework encompasses (a) the themes firms can use to craft their requests and the subsequent consumer reactions, and (b) the reasons behind firms' decisions on how to design an opt-in communication. We will begin by illustrating which alternative firms face when crafting these messages, then we will focus on how messages are expected to affect consumer likelihood of granting opt-ins. Lastly, we will discuss the factors affecting the decision-making process in this design.

Figure 2: Firms and consumer behavior vis a vis GDPR Opt-in Ecosystem



We refer to the framework in Figure 2 to address the following questions:

How Do Firms Ask for Data? How Do Consumers React to Data Request Formats?

Past literature shows that the format of the data request can play a role in affecting consumers' likelihood to opt-in (Johnson et al. 2002). The main factors that characterize such requests and impact the decision to opt-in can be grouped into four main categories: transparency, control, incentives, and framing. *Transparency* refers to the provision of information offered to users about how data are collected, used, and managed as per Art. 12 of the GDPR. This entails ensuring that customers understand the processes involved in data collection, storage, and utilization. *Control* denotes the ability of individuals to dictate how their personal information is collected, used, and shared (as per Art.13 of the GDPR). *Incentives* are the benefits offered to users (e.g., discounts, gifts) to encourage opt-in. Finally, *framing* relates to the presentation and structure of data request formats. Notably, transparency and enhanced control can make consumers' decisions sounder because they enrich the information content of the communication, whereas framing and incentives are often used to persuade users to agree and grant data consent.

Transparency

Past literature highlights that transparency increases users' likelihood to disclose personal data (Aguirre et al. 2015; Benson, Saridakis, and Tennakoon 2015; Mohan, Buell, and John 2020), as granting transparency leads to fewer concerns about privacy and fosters disclosure. For example, Athey, Catalini, and Tucker (2017) show that the provision of a clearly stated privacy policy leads to increased trust and reduces privacy concerns, which positively influences the probability of granting access to personal data. In a similar vein, Godinho de Matos and Adjerid (2021) found that opt-in rates for non-sensitive data increased once GDPR-compliant consent was obtained.

Control

Past research also reports that the missed provision of consumers' control over their disclosed data results in more privacy concerns and a decreasing propensity to purchase (Xu et al. 2012; Dinev and Hart 2004; Malhotra, Kim, and Agarwal 2004; Phelps, Nowak, and Ferrell 2000). Tucker (2014) found that providing control makes people perceive data collection as less intrusive and fosters data disclosure. Similarly, Martin, Borah, and Palmatier (2017) show that offering customers control can mitigate privacy concerns, especially during data breaches, potentially improving returns.

Taken together, these results suggest that transparency and control can boost trust, reduce privacy concerns, and lead to data disclosure. Therefore, we expect that crafting a transparent and control-providing message can increase the likelihood of opt-ins while reducing the risk of non-compliance with regulatory authorities.

On the other side of the spectrum of the central box of Figure 2, we consider messages including persuasive cues, such as incentives and framing.

Incentives

Empirical research indicates that consumers are likely to grant firms access to their data when they are offered monetary or non-monetary incentives. For example, Chellappa and Sin (2005), Grossklags and Acquisti (2007) and Athey, Catalini, and Tucker (2017) found that consumers were more likely to opt-in when they were offered a modest discount. By contrast, Krafft, Arden, and Verhoef (2017) found that incentives have a limited impact on boosting consent because they provoke reactance.

Recent research has concentrated on how firms and consumers value data and exchange it for money or goods. Kummer and Schulte (2019) show that firms are ready to trade data with money. In particular, app developers require more personal details from customers subscribing to free apps than from customers subscribing to paid apps. On the other hand, consumers modify their assessment of privacy based on what they are offered to obtain it. Tomaino, Wertenbroch and Walters (2023) found that consumers place a greater value on their data when they exchange them for money rather than goods. Choi, Jerath and Sarvary (2023) demonstrate that when consumers choose to opt-in, they weigh the advantages of receiving more precisely targeted product information against the downside of increased advertising exposure. Further, Collis, Moehring, Sen and Acquisti (2022) reported that a wide disparity in data valuation among consumers exists. Our study contributes to this literature, as we seek to understand how the presence of incentives or other forms of persuasive arguments in data requests can affect opt-in. Although we are not interested in what affects users' data valuation per se, we are interested in the impact of factors that influence users' data valuation on opt-in decisions.

Framing

The framing of messages has also been shown to influence data disclosure (Johnson et al. 2002). John, Acquisti, and Loewenstein (2011) report that the disclosure of private information is responsive to environmental cues such as (i) how people are asked (direct vs indirect questions), (ii) how forms to be completed are designed (professional vs unprofessional), and (iii) the initial prompt of the request (evoking privacy concerns or not). Acquisti, John, and Loewenstein (2013) document that individuals tend to weigh losses more heavily than gains when making decisions regarding data disclosure to companies. This suggests that framing, or the way information is presented, plays a pivotal role in privacy-related decisions. Consequently, data requests that emphasize the negative consequences of denying consent may be more effective than those highlighting the positive outcomes of opt-in. Furthermore, Adjerid, Acquisti, and Loewenstein (2019) explored the role of framing in privacy choices within digital contexts, in which consumers frequently confront both upstream and downstream privacy decisions. Initially, consumers decide whether they wish to be part of a particular platform, then choose which data to upload. The study's findings indicate that subtle changes in the framing of information significantly influence consumers' upstream disclosure choices, but this does not appear to affect their downstream behavior. This means that once users have been persuaded to opt-in based on how the request was presented, their subsequent decisions do not seem to compensate for the framing effect. This insight underscores the enduring impact of initial framing on user behavior, highlighting its importance in shaping privacy decisions.

Past research has also examined how the use of time-framed messages can influence consumer decision-making (Thaler and Sunstein 2008). Privacy decisions involve an inter-temporal tradeoff as users contrast the immediate value associated with personalized marketing with future negative consequences from sharing their data. Past studies have shown that, given this tradeoff, individuals tend to use hyperbolic discounting and overvalue the rewards that opt-in can offer in the short term, undervaluing future costs that granting access to data may provide (Acquisti et al. 2017; Jentzsch, Preibusch, Harasser 2012).

Overall, as illustrated in Figure 2, this literature suggests that firms can choose to design their data-request messages by pursuing two potentially contrasting strategies. They can either provide information about privacy and data security, incorporating informative cues such as transparency and control in their requests or create messages that offer rewards in exchange for data and/or use framing elements. By being transparent and assuring customers about their ability to control data, firms try to enhance trust, reduce privacy concerns, and comply with regulators' requirements; by using incentives or framing, firms try to maximize data collection running the risk of a regulator's retaliation. Importantly, firms can also employ a blend of these messages, potentially to achieve the benefits of both approaches.

Table 1 provides a more detailed summary of this rich literature.

Table 1: Literature Review Summary

Papers	How do firms ask for data?		How do consumers react to data requests?	Why do firms choose a specific format to ask for data?	
	Informative Cues	**Persuasive Cues**	**Disclosure / Opt-in**	**Costs and Risks**	**Benefits**
Acquisti, Adjerid, and Brandimarte, 2013	Control & Transparency				
Adjerid et al., 2013	Transparency	Framing	Transparency (+)		
Aguirre et al., 2015	Transparency		Framing (+)		
Benson, Saridakis, and Tennakoon, 2015	Transparency		Transparency (+)		
Dinev and Hart, 2004	Control		Purchase (-)		
Godinho de Matos, and Adjerid, 2021	Transparency		Transparency (+)		
Malhotra, Kim, and Agarwal, 2004	Control & Transparency		Purchase (-)		
Martin, Borah, and Palmatier, 2017	Control & Transparency		Retention (-)		
Milne and Culnan, 2004	Transparency	Framing	Transparency (+) Framing (+)		
Mohan, Buell, and John, 2019	Transparency		Transparency (+)		
Tucker, 2014	Control & Transparency		Transparency (+) Control (+)		
Phelps, Nowak, and Ferrell, 2000	Control		Purchase (-)		
Xu, Teo, Tan, and Agrawal, 2012	Control		Purchase (-)		
Acquisti, John, and Loewenstein, 2013		Framing			
Athey, Catalini, and Tucker, 2017	Transparency	Incentives	Transparency (+) Incentives (+)		
Chellappa and Sin, 2005		Incentives	Incentives (+)		
Acquisti et al., 2017		Framing	Framing (+)		
Jentzsch, Preibusch, and Harasser, 2012		Framing	Framing (+)		
Grossklags and Acquisti, 2007		Incentives & Framing	Incentives (+)		
John, Lowenstein, and Loewenstein, 2011		Framing			
Krafft, Arden, and Verhoef, 2017		Incentives	Incentives (-)		
Johnson et al., 2002		Framing			
Collis et al., 2021	Transparency		Transparency (+)		
Adjerid, Acquisti, and Loewenstein, 2019		Framing	No compensation for framed upward choices		
Kummer and Schulte, 2019		Incentives			Firms trade money for data
Tomaino, Werthenbroch, and Walters, 2023		Incentives	Data valued more with money than goods		
Acquisti, Friedman, and Telang, 2006				Privacy violations impact more popular firms	
Bleier, Goldfarb, and Tucker, 2020				GDPR reduces data-driven innovations	
Jia et al., 2021				GDPR reduced data-driven innovation	
Johnson, Shriver, and Goldberg, 2023				GDPR reduced competition	

Papers	How do firms ask for data?		How do consumers react to data requests?	Why do firms choose a specific format to ask for data?	
	Informative Cues	**Persuasive Cues**	**Disclosure / Opt-in**	**Costs and Risks**	**Benefits**
Campbell, Goldfarb, and Tucker, 2015				Opt-in affects small firms	
Aziz and Telang, 2015					Cookies boost ad revenues
Bleier and Eisenbeiss, 2015					Personalized ads boost effectiveness
Goldberg, Johnson, and Shriver, 2019				GDPR cut web visits and revenues	
Goldfarb and Tucker, 2011				Removing cookies reduces adv effectiveness	
Johnson, Shriver, and Du, 2020				Opt-in reduces ad revenues	
Sun, Zhu, Li, Zhang, and Xu, 2024				Data access ban reduces browsing and purchases	
Marotta, Abhishek, and Acquisti, 2019					Cookies boost publishers' revenues
Sharma, Sun, and Wagman, 2019				Regulation cuts small publishers' and advertisers' revenues	
This Paper	**YES Informative**	**YES Persuasive**	**YES Opt-in**	**YES Expected Costs**	**YES Expected Benefits**

Why do Firms Choose a Specific Format to Ask for Data?

The left side of Figure 2 focuses on firms' decision-making, illustrating that their choice of which mix of informative and persuasive cues to use is based on evaluating the trade-off between the costs and benefits of the chosen messaging types, following Fischhoff (2015). More specifically, we argue that firms trade off expected returns they derive from data gathered through opt-in with expected costs they might incur when users and regulators negatively react to formats they choose. To do so, firms must not only identify benefits and costs associated with utilizing persuasive versus informative cues but also assess the size of the risk associated with adopting a format, as well as its relevance and impact with respect to the cost-benefit tradeoff (Fischhoff 2015).

Past work has examined how the introduction of privacy protection policies can affect firms (See Table 1 for a summary of the main contribution in this area). This suggests that firms have a strong incentive to guard and use consumers' data and thus take advantage of the flexibility provided by regulators to create opt-in requests.

Research has also examined how firms might try to circumvent the consequences of a regulatory system. This research shows that firms can react when regulators introduce policies meant to modify their behavior by targeting measures that the policy introduced (e.g., re-permission requests) rather than changing the focal behavior (i.e., assure privacy protection) that the policy addressed (Drugov and Troya 2019; Reynaert and Sallee 2021). This work suggests that firms are motivated to mitigate the adverse effects of a regulated system when given the opportunity. Within the scope of this literature, we argue that firms endeavor to enhance their access to users' data by employing persuasive cues in their communications.

The benefits firms expect from using a persuasive message center around incremental data they can collect compared to their use of informative messages. The use of persuasive cues may, indeed, result in higher opt-in rates as supported by the decision-making literature, indicating people's susceptibility to persuasive appeals (Petty, Wegener, and Fabrigar 1997). However, with respect to privacy decisions, it remains uncertain whether or not pursuing this strategy outweighs the adoption of informative cues, as the empirical literature on this subject offers mixed results.

The magnitude of expected benefits that firms associate with persuasive opt-in requests varies depending on a firm's capacity to gather data per opt-in and effectively leverage the data it collects. Research has shown that targeting strategies and personalized marketing campaigns can increase firms' profitability (Aziz and Telang 2015; Bleier and Eisenbeiss 2015; Goldberg, Johnson, and Shriver 2019; Goldfarb and Tucker 2011; Marotta, Abhishek, and Acquisti 2019). Nevertheless, firms require consumer consent to use these strategies and access and commercially utilize their data. Consequently, securing permission is key for firms that heavily rely on customer data for product targeting and developing incremental revenue streams (e.g., advertising revenues).

Thus, firms proficient in collecting a larger volume of data per user, particularly using multiple cookies (Aziz and Telang 2015) and those with superior abilities to monetize data, are likely to experience more substantial returns from adopting persuasive cues for data acquisition; as a result, we expect them to use them more.

Incorporating persuasive elements into communications aimed at soliciting data consent may also entail costs. Both consumers and regulatory bodies can detect a company's intent to influence users to opt-in. This may cast an unfavorable light on firms, potentially resulting in consumer complaints (Acquisti, John, and Loewenstein 2012; Kraff, Arden, and Verhoef 2017; Tucker 2014) and attracting the attention of consumer associations, thereby provoking regulatory scrutiny.

As these messages can signal firms' data management "quality," firms that want to reinforce their relationship with their customer base or establish (or re-establish) trust may opt for information-based messages instead of persuasive ones. Of course, the size of this risk is more pronounced when firms have more to lose (i.e., popular, visible firms with time-tested trusting relationships with customers). This is in line with past work showing that privacy violations have a greater impact on popular firms (Acquisti, Friedman and Telang 2006), when reputational risks are higher, firms are less likely to encourage consumers and offer monetary rewards to access their data (Kummer and Schulte 2019). Similarly, the relevance of such reputational risk becomes more pronounced when firms have previously experienced privacy violations (e.g., data breaches), as they have already encountered adverse consequences of poor data management. In such instances, we contend that firms are more inclined to comply with regulations and opt for either fully informative or informative and persuasive messages.

Finally, we argue that firms must assess risks associated with utilizing persuasive cues using the data that they collect. Risk might be acceptable when brings large benefits and cannot be reduced, whereas small risks might be unacceptable if they bring small benefits and can be eliminated (Fischhoff 2015). In the context of privacy, the benefits of using data for targeting and personalization have been shown to be very high (Sun, Yuan, Li, Zhang, and Xu 2023), whereas the risk of reputational damage can be mitigated by implementing a mixed strategy (informative and persuasive) and/or offering valuable compensation for the use of data. Thus, when the ratio between expected benefits and expected costs is high, firms have a high incentive to use persuasive cues in their opt-in requests.

Overview of Our Research Strategy

To address our three research questions, we use the conceptual framework illustrated in Figure 2, and we conduct three studies. Table 2 provides a comprehensive overview of our research scope and empirical strategy.

Study 1

This study's objective is to analyze how firms craft re-permission emails in response to the enactment of GDPR. We first describe our data and then our analytic strategy.

Data

We collect GDPR re-permission emails by using two approaches. First, we used a snowball technique by obtaining the emails received by (a) students from a large European university, (b) this paper’s authors, and (c) their colleagues. Second, we asked a Prolific panel to send us the GDPR re-permission emails they received directly from firms. We provided a monetary incentive to respondents who correctly followed our instructions.[3]

Table 2: Paper Scope and Analyses Overview

Research questions	Data	Source	Analyses	Tables and figures
(1) How did firms ask for data?	Collection of a sample of 1,506 re-permission emails	Snowball approach and request through a Prolific panel	Unsupervised: BERT Topic Supervised: Theory-Based content analysis	Tables 3–5 Figures 3–5 Web Appendixes A, B
(2) How do consumers react to data request format?	Observed opt-in Intention to opt-in	Field test in collaboration with a company Lab experiment	Field experiment Lab experiment	Tables 6–7 Web Appendixes C, D
(3) Why did firms choose a data request format?	No. of persistent marketing and third-party cookies	Cookiebot	Fractional logit regression model	Table 8 Figure 7 Web Appendixes E, F, G
	Online ad revenues	SemRush Rank2Traffic		
	No. of data breaches	Prilock Have I Been Pwned Wikipedia		

We collected 676 emails by using the first approach and 934 by using the second approach, yielding a total of 1,610 emails. Upon analysis, we identified emails from 1,396 distinct businesses, for which 9% of these businesses had multiple email entries in our dataset. Of this 9%, we detected 104 identical emails (duplicates). After we remove these duplicates, our final dataset comprises 1,506 re-permission emails sent by 1,396 different firms. We note that 5.9% of these emails were sent by the same company but are not duplicates, so we retained them.[4] More specifically, 94% of the businesses were represented

[3] We detailed retrieval procedures from their email accounts by using specific keywords. To qualify for payment, respondents needed to upload a minimum of three re-permission email PDFs.

[4] Please see Web Appendix A (Figure WA2) for an example of different GDPR re-permission emails sent by the same company.

by a unique email, 5% were represented by two distinct re-permission emails, and the remaining by more than two distinct emails.

For each email, we recorded information about the language of the e-mail: 71% were in English and 29% were in other European languages (French, German, Italian, or Spanish). Additionally, we recorded the date the email was sent.[5] Interestingly, the regulator gave firms two years to comply with the GDPR, but most of the emails were sent around May 2018 (see Web Appendix A, Figure WA1).

Table 3: Descriptive Statistics for Firms Sending Re-permission Emails (N = 1396)

Description	**Examples**	**Number of Firms**	**Percentage**
Services	Accenture, Aruba, British Airways, Facebook, NH hotels, Airbnb, HSBC	889	63.68
Products			
Mainly Digital Products	Bitcoin, Canva, Photobox, Moo, Steam, Riot Games, Zynga	230	16.48
Mainly Physical Products	BMW, Clarks, Danone, IKEA, Maybelline, Nokia, Zara	277	19.84
Total		1396	100.00

We have coded businesses to distinguish firm type (service, digital or physical products). Table 3 summarizes this classification, providing examples of brands in each category. Although the majority of these businesses are based in Europe (65%), several other countries outside the EU are also represented (e.g., North America, 25%). For more descriptive statistics about the sample, please see the Web Appendix F.

Analysis of GDPR Re-permission Emails: Mapping the Themes

We combine a supervised, theory-driven approach with an unsupervised data-driven method, to map the content of the 1506 re-permission emails we collected. Both methods are detailed below.

[5] In 2018, 71% of emails were sent, and in 2019, 3% of emails were sent; for 26% of emails, the sending date is missing. See also Figure WA1, Web Appendix A. As some firms were slow to adopt GDPR (source: gdpr.eu), we retained all emails in our sample.

Theory-based Approach

Our theoretical framework is anchored in past literature which delineates four critical factors (see Table 1) that we use to analyze the content of re-permission emails. These factors are divided into two primary categories: 1) Informative cues that include a) transparency and b) control, and 2) Persuasive cues, which include a) incentives and b) message framing (gains/losses and time frame). Note that these factors are not mutually exclusive and may be integrated in diverse combinations within a single re-permission email request.

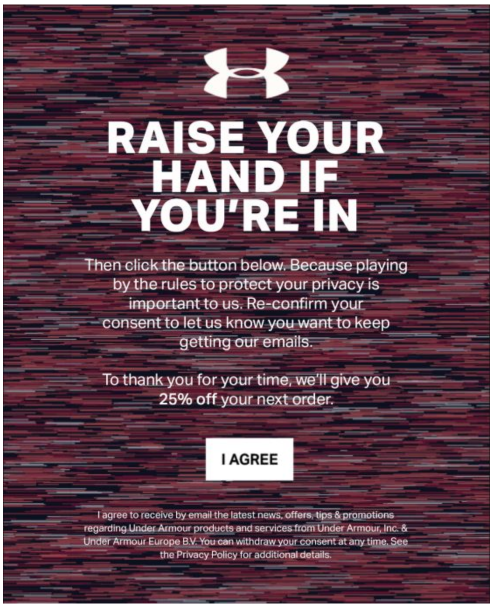
Panel A of Figure 3 illustrates examples of persuasive cues. The re-permission email from Under Armor exemplifies the use of incentives (particularly monetary ones) by offering a 25% discount on the next order. BMW employs both loss and time framing in its message, emphasizing that from May 25, 2018, inaction will result in users missing out on exclusive news, offers, and experiences. The direct and persuasive style is further accentuated by the use of an “I'm in” button for opt-ins.

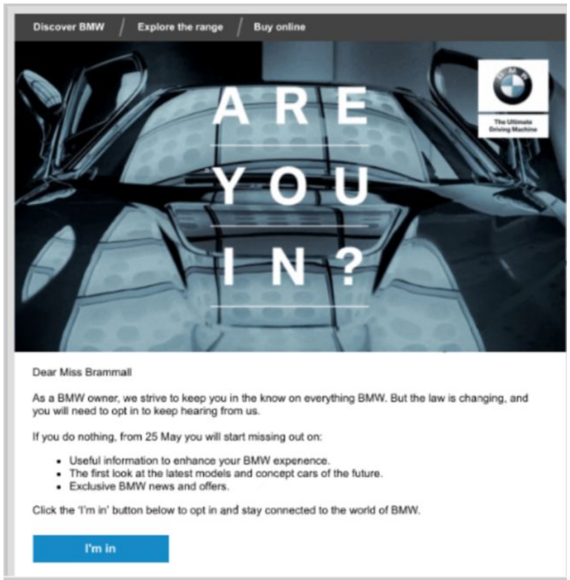
Panel B of Figure 3 reports examples of informative cues. National Geographic employs a bullet-point structure in its email, clearly highlighting key GDPR changes and thus exemplifying transparency in re-permission email design. Similarly, Riot Games conveys information about GDPR, underscoring the importance of transparency. Notably, within the ‘accessibility’ section, receivers are reminded that they can control and delete personal data. Meanwhile, Maybelline NYC distinctly mentions both transparency and control.

Overall, Figure 3 illustrates the varied styles employed in designing re-permission emails. Guided by these examples, the objective of our theory-based approach will be to identify both informative and persuasive cues present within the emails in our sample.

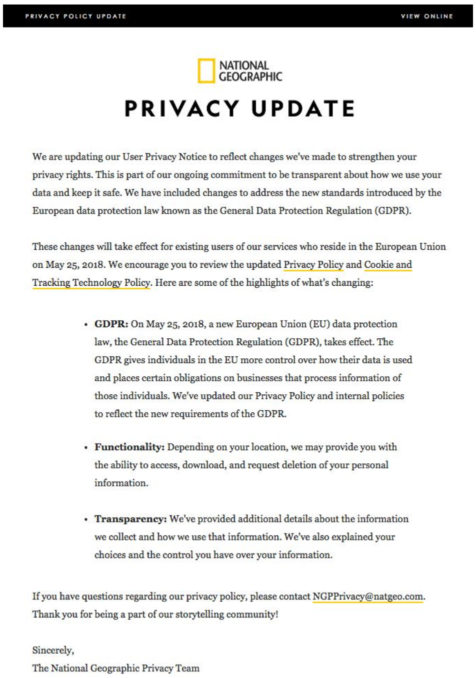
Figure 3: Examples of Re-Permission Emails

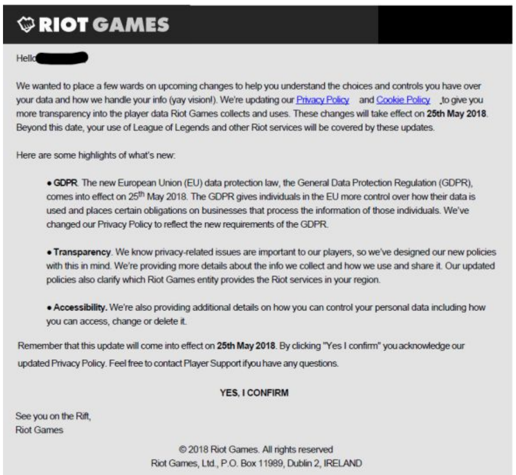
Panel A: Persuasive Cues – Incentives and Framing

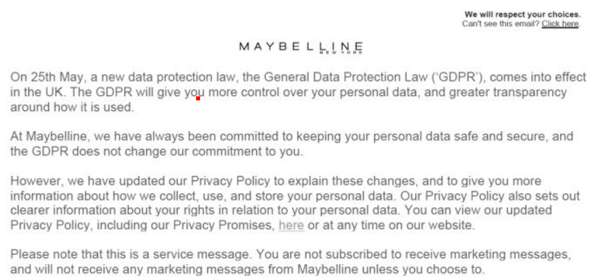




Panel B: Informative Cues - Control and Transparency







We coded all the re-permission emails into persuasive, informative, or both, along with corresponding dimensions (incentives, framing, transparency, and control). Initially, two independent judges were tasked with coding 20% of the sample. Following this initial annotation, two additional independent coders analyzed the remaining 80% of the sample. The initial content analysis of 20% helped refine the rest of 80%. For both, coders were supplied with a protocol to content-analyze the emails, transforming our theory-based themes (control, transparency, framing, and incentives) into distinct variables. This detailed protocol and examples of coded emails can be found in Web Appendix A (Table WA1 and Figure WA3). After finalizing all coding, we assessed interjudge reliability to confirm the degree of consistency between coders. The average Krippendorff's alpha is .90, and Cohen's kappa is .80 for control and transparency these metrics are above the commonly applied criteria for satisfactory reliability (Cohen 1960; Multon 2010).[6]

Figure 4 shows the results we obtained by using our theory-based content analysis. It outlines the main cues associated with our sample of emails. Panel A shows various persuasive elements used in re-permission emails: incentives, gain/loss, and time framing.

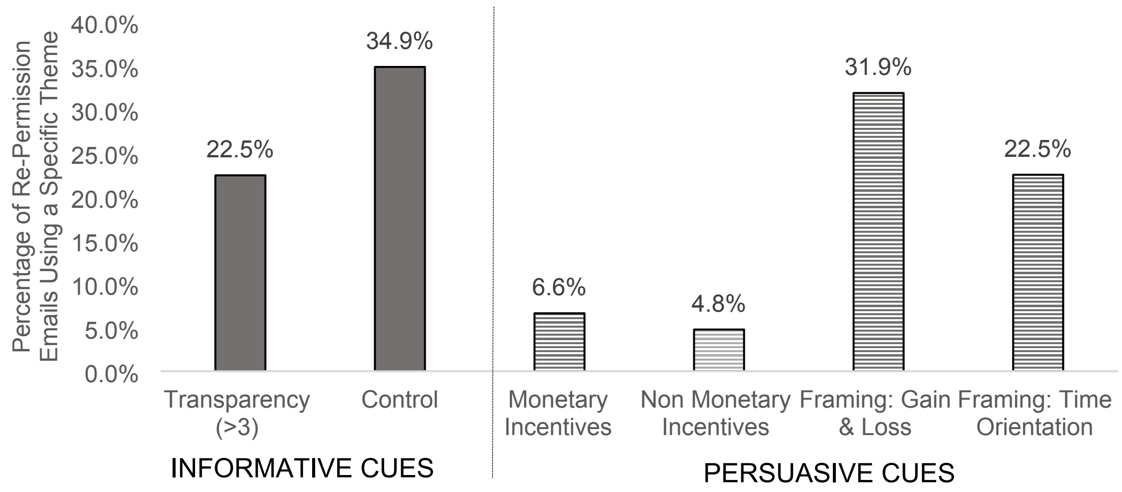
Panel B of Figure 4 illustrates the combination patterns of specific cues used in the same email. Notably, 26% of emails exclusively utilized persuasive cues (*P* area in Figure 1), while 33% relied solely on informative ones (*I* area in Figure 1). A combination of both cues was present in 24% of emails (*M* area of Figure 1). Finally, 18% of emails in our sample did not use either persuasive or informative cues (*U* area of Figure 1); for these emails, we use the label "Uninformative & Unpersuasive," which indicates that the email offers a snapshot of the GDPR but fails to meet the informative requirements outlined by the GDPR or do so in an unclear manner, particularly with respect to being consistent with Article 12. These messages also refrain from using persuasive cues.

---

[6] For the 20% sample, Krippendorff's alpha was .92 and Cohen's kappa was .82. For the 80% sample, Krippendorff's alpha was .88 and Cohen's kappa was .80.

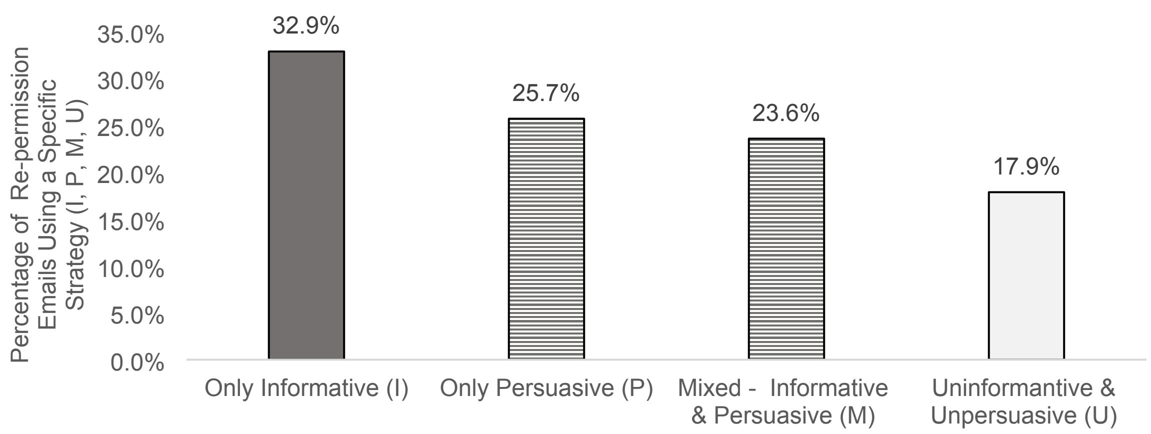
Figure 4: Main Cues Used in Re-permission Emails (N = 1,506)

Panel A: Percentage Distribution of Specific Theme Used in Re-permission Emails



Note: Transparency is measured using a five-level scale variable, with the bar representing the proportion of values above 3 (i.e., above-average transparency). The other bars, measured on a 0-1 scale, represent the proportion of emails with a specific theme.

Panel B: Percentage Distribution of Combinations of Cues Used in Re-permission Emails



Notes: We define "Only Informative" emails (*I*) with a combined score of [Control (0-1) + Transparency (1-5)] > 3, and do not employ any persuasive cues (i.e., scoring 0 on all persuasive cues identified in Panel A). "Only Persuasive" emails (*P*) are those that exclusively use persuasive cues and have a combined score of [Control (0-1) + Transparency (1-5)] ≤ 3. "Mixed Informative & Persuasive" emails (*M*) have a combined score of [Control (0-1) + Transparency (1-5)] > 3 and score more than 0 on at least one persuasive theme from Panel A. Finally, "Uninformative & Unpersuasive" emails (*U*) have a combined score of [Control (0-1) + Transparency (1-5)] ≤ 3 and score 0 on all persuasive cues from Panel A.

Finally, Table 4 outlines results that explore variations in theme utilization based on firm type, product type, and distribution strategy. It shows significant differences, especially when we classify firms by product type (digital vs. physical) and distribution channels (pure digital vs. those with an offline presence). Notably, firms selling physical products or with traditional brick-and-mortar stores tend to favor persuasive techniques compared to strictly digital firms. This tendency can be ascribed to the benefits firms can attain from effectively leveraging personal data across online and offline channels. Acquiring a comprehensive perspective of both existing customers and prospects can yield enhanced customer insights and more effective targeting (e.g., Angel 2018; Ucuzoglu and Hagel III 2020).

Table 4: Cues Used in Re-permission Emails Categorized by Firm Type and Offline Presence

Panel A: Firm Type

Combinations of Cues	Service (N=968)	Product (N=538)	z test	p-value	Total (N=1506)
Only Informative (I)	34%	32%	-0.820	0.41	33%
Only Persuasive (P)	26%	25%	-0.309	0.76	26%
Informative & Persuasive (M)	24%	23%	-0.640	0.52	24%
Uninformative & Unpersuasive (U)	16%	21%	2.066	0.04	18%

Panel B: Focus on Product Type (Digital versus Physical)

Combinations of Cues	Physical Product (N=301)	Digital Product (N=237)	z test	p-value	Total (N=538)
Only Informative (I)	25%	40%	-3.57	0.00	32%
Only Persuasive (P)	31%	18%	3.38	0.00	25%
Informative & Persuasive (M)	21%	24%	-0.77	0.44	23%
Uninformative & Unpersuasive (U)	23%	18%	1.27	0.21	21%

Panel C: Focus on Distribution (Only Digital versus Offline Presence)

Combinations of Cues	Offline Presence (N=509)	Only Digital Presence (N=997)	z test	p-value	Total (N=1506)
Only Informative (I)	30%	35%	-1.89	0.06	33%
Only Persuasive (P)	30%	24%	2.64	0.01	26%
Informative & Persuasive (M)	21%	25%	-1.67	0.10	24%
Uninformative & Unpersuasive (U)	19%	17%	1.15	0.25	18%

Data-Driven Unsupervised Approach

In this approach, we did not impose a structure on the data. Instead, we used BERT (Bidirectional Encoder Representations from Transformers), a transformer embedding-based approach developed by Devlin, Chang, Lee, and Toutanova (2019), to analyze email text and detect latent topics. As a deep contextual language model, BERT comprehends word context within a sentence by considering both preceding and following words, which enables a deeper understanding of word relationships and overall text meaning. This analysis involves investigating nuanced connections between words and concepts and leveraging transfer learning effectively (Hartmann and Netzer 2023). Compared to traditional bag-of-words models like Latent Dirichlet Allocation (LDA), BERT's transformer-based approach yields more accurate and context-aware word embeddings, as it captures a word's meaning in the context of an entire sentence.

In our analysis, we use hierarchical clustering with BERTopic to identify the optimal number of topics for our study (see Figure WB1 in the Web Appendix B). Hierarchical clustering effectively groups similar data points into clusters based on their semantic similarities, as captured by BERTopic (Catapang, Kyle, and Cleofas 2022). This approach strikes a balance between capturing relevant nuance within the data and avoiding excessive fragmentation that might hinder result interpretability. After carefully evaluating different levels of topic granularity, we settled on three distinct topics (see Web Appendix B for the discussion leading to this choice). Interestingly, our results indicate that the unsupervised analysis also identifies informative and persuasive cues as well as a category similar to "Uninformative & Unpersuasive," as identified through our theory-based approach. This category appears to center on legal nuances without adopting a persuasive or transparent tone.

We further explored the robustness of this classification by using LDA (Blei, Ng, and Jordan 2003), an alternative, widely used unsupervised approach. Remarkably, similar to the results that we obtained from hierarchical clustering with BERTopic, our LDA analysis also revealed three distinctive and coherent topics within the emails (see Table WB2, Web Appendix B). The convergence of outcomes from both BERTopic-based hierarchical clustering and traditional LDA analysis thus provides robust evidence for the presence of three salient themes within re-permission emails. Although it is well known that the topics inferred by unsupervised approaches are not always easily interpretable (Sievert and Shirley 2014, p. 64), we used several interactive visualization systems to facilitate a deep inspection of topic–term relationships in our BERTopic and LDA model. More importantly, the coders' classifications from our theory-based approach allowed us to cross-validate these findings.

In Table 5, we report the results of our cross-validation analysis and, in so doing, support the idea that the three themes provide consistent results.

Table 5: Cross-validation of Theory-based vs. Data-driven Approaches—Results Logit Model: DVs = Topics Identified by the BERTopic, IVs = Topics Identified by LDA, Theory-based Topics Coded Manually

		BERT Topic 1 Informative (Transparency)		BERT Topic 2 Persuasive		BERT Topic 3 Uninformative & Unpersuasive	
		Coef.	z	Coef.	z	Coef.	z
LDA Topics	LDA Topic 1 (Informative)	1.880	10.16	0.470	1.70	-2.403	-13.56
	LDA Topic 2 (Persuasive)	0.024	0.14	3.522	13.48	-1.802	-9.95
	Constant	-1.432	-8.88	-3.616	-15.09	0.383	2.42
Theory-Based Topics	Control	0.113	1.62	-0.319	-3.10	-0.015	-0.19
	Transparency	0.148	5.14	-0.478	-10.06	0.056	1.48
	Incentive [1]	-0.208	-2.37	0.410	3.33	0.002	0.02
	Framing [2]	-0.193	-3.20	0.551	6.26	-0.057	-0.80
	Constant	-0.771	-9.66	-0.801	-6.65	-1.576	-13.87

[1]We grouped the presence of any monetary and non-monetary incentives into the variable "Incentive."
[2]We grouped the presence of a gain/loss type and time of framing into the variable "Framing."

First, we found that the BERT and LDA topics are related and provide consistent results. Second, BERT topics are also related to the manually coded themes. The first and second topics complement each other in terms of significant themes. For example, BERT Topic 1 ("Informative") is positively related to the transparency variable and negatively related to incentives and framing. On the other hand, BERT Topic 2 ("Persuasive") shows the opposite pattern, as it is negatively related to both control and transparency and positively related to incentives and framing. Finally, BERT Topic 3 lacks a significant relationship with theory-based topics, prompting consideration of the label "Uninformative & Unpersuasive."

In conclusion, this data-driven unsupervised approach offers a compelling alternative for vast datasets of privacy-related communication. Stakeholders can use these valuable techniques, which can discern patterns and themes without extensive human intervention, to navigate and interpret the vast landscape of data privacy communication.

Crucially, both supervised and unsupervised methodologies yield consistent results that help us determine how firms request data consent. While 33% of firms closely adhere to regulators' mandates of transparency in their re-permission campaigns, 26% do not comply with these mandates. Conversely, 51% choose to exploit the regulation's format and language flexibility guidelines to include persuasive elements into their communication. Interestingly, we also show that firms selling physical products or those with a brick-and-mortar presence use many more persuasive techniques, possibly driven by the challenges they face in collecting consumer data.

Study 2

The purpose of this study is to explore how consumers react to the different strategies used by firms to ask for opt-in and personal data. In particular, we investigate whether firms are more likely to provide opt-in and access to different sources of personal information when requests for data contain persuasive elements, informative elements, or a combination of the two.

Design and Data

We partnered with a leading European sleep-products (e.g., mattresses, pillows, duvets, bed frames) firm that sells through both online and offline channels. The firm provided us with a list of 14,078 prospective customers (identified via prior internal initiatives) for whom it wished to obtain data, so the firm could better target and manage these customers in its CRM system. To do so, the firm needed to obtain opt-in per the GDPR. These prospects were individuals living in the EU who had never purchased from the firm.

The field experiment ran from June 30, 2021, to July 9, 2021 (soft launch), and from July 10, 2021, to July 30, 2021 (full experiment). Prospects received emails from the company asking them to complete a form requesting some personal data. A reminder was sent after one week to prospects who did not reply to the first message. Before completing the form, prospects had to accept the privacy terms and opt-in (i.e., allow the company to (potentially) use the collected personal information). If prospects accepted the privacy terms, they were asked to complete a form with 18 non-mandatory fields. The prospects' choices corresponding to our two outcome variables of interest are (1) the prospect's decision to opt-in or not, and (2) given opt-in, the number of discretionary personal data items provided (on the form).

We designed the experimental manipulation as follows. First, we created different communications varying the level of informative elements (representing GDPR intent) and persuasive elements (representing firms' desire to maximize opt-in collection) in the messages soliciting opt-in (for personal data). Next, we created six treatments using a 3 × 2 design with persuasive elements at three levels and informative elements at two levels.[7] The three levels of the persuasive elements are (1) a "high" level including several persuasive cues: a gain/loss frame, time orientation, and an economic incentive provided as a benefit for sharing personal information,[8] (2) a "moderate" persuasive level with only an economic incentive, and (3) an "absent" or control level with no persuasive cues.

The "informative" communication levels are (1) a "high" level, emphasizing the

[7] In Study 1, we analyzed re-permission emails and observed the richness of persuasive cues (gain/loss, time orientation frames coupled with economic incentives). Based on this evidence and a theory-based investigation of the opt-in requests, we created six experimental conditions. Notably, we opted for a three-level experimental design to manipulate persuasive cues.

[8] Prospects were athletes (both amateur and professional) interested in specific sports (e.g., biking, running). The incentive is a 50% discount for purchasing the "race number (bib)" for an upcoming sport event. The regular price of this kind of bib (gold bib) is 200 €, which provides athletes access to better facilities at races. The specific group of individuals who received our messages might impact the generalizability of the results.

company's commitment to protect personal data; these messages provide information about how the data will be treated and stress transparency, clarity, and safety with respect to storage and management;[9] and (2) an "absent" or control level that does not mention the company's specific commitment to the protection of personal data.

Table 6: Field-test Groups

Informative	**Persuasive** *Low*	*Moderate*	*High*
Low	**G1 Generic Message** Total contacts = 2,339	**G3 Moderately Persuasive** Total contacts = 2,340	**G5 Highly Persuasive** Total contacts = 2,340
High	**G2 Informative** Total contacts = 2,340	**G4 Informative & Moderately Persuasive** Total contacts = 2,340	**G6 Informative & Highly Persuasive** Total contacts = 2,379

Table 6 shows our field-test campaigns. These campaigns are designed in terms of degrees of persuasiveness (high, moderate, absent) and informativeness (high, low). The sample of 14,078 prospects was randomly assigned to six cells. Figures WC1 and WC2 in the Web Appendix C provide the full list of messages used.

Pre-test

To ensure that our experimental manipulation worked as intended, we conducted a series of pre-tests. We pre-tested messages used in the field test by using two panels of respondents interested in sleep products. Specifically, we used two different Prolific panels of respondents who received monetary compensation to complete the survey. Participants were randomly exposed to one of the messages that we designed for the field experiment (see Table 6) by using a between-subjects design. We collected data in March and April 2021. In the first sample (n = 188), we tested messages G1, G2, G3, and G4, and we tested the high-

[9] Transparency, clarity, and safety perfectly reflect key principles of the new GDPR (e.g., Recital (39) of the GDPR). In addition, our empirical analysis (Study 1) shows that a significant portion of firms used these cues in their re-permission email campaigns in 2018.

persuasiveness condition in the second sample (n = 191), notably messages G1, G2, G5, and G6. We find that our manipulations work in the expected direction in both samples (Tables WD1–WD2 and Figures WD1–WD2 in the Web Appendix D). We report items used to measure perceived persuasiveness, informativeness, and transparency in Table WD3 (Web Appendix D).

Field Experiment Results

In Table 7, we report results of a logistic regression in which the dependent variable represents privacy acceptance (1=the prospect decided to opt-in, and 0=the prospect did not opt-in), and the independent variables are experimental groups with a baseline of G1. Results show that a moderately persuasive message leads to higher opt-in (1.04, $p < 0.001$). Furthermore, while the impact of an informative message does not differ significantly between test and control (0.35, $p > 0.100$), it works when combined with a very persuasive message (0.68, p < 0.05).[10] Overall, these results show that the intensity of persuasiveness matters in shaping responses to opt-in requests. Consumers tend to opt-in to messages with monetary incentives (moderately persuasive), but when economic incentives are coupled with additional persuasive cues (highly persuasive), this strategy does not work. We posit that the highly persuasive condition might come across as overly aggressive or manipulative (Brehm 1966).

Table 7: Field Test: Logistic Regression (n = 14,078) DV=Opt-in = 1 (Yes); 0 Otherwise

	Coef.	z
Highly Informative (G2)	0.35	1.48
Moderately persuasive (G3)	1.04	4.97
Highly persuasive (G5)	-0.17	-0.66
Moderately persuasive × Highly Informative (G4)	-0.19	-0.70
Highly persuasive × Highly Informative (G6)	0.68	2.09
Constant	-4.28	-24.03
N = 14,078, LR $\chi^2(5)$ = 83.32, p-value < 0.001		

[10] We replicate this finding in a pre-registered lab experiment, conducted in a different industry (see Web Appendix C, Tables WC2-WC6)

In addition to opt-ins, firms often seek to collect specific personal data about their prospects for targeting purposes. One could argue that the way communication is designed (e.g., informative vs. persuasive) could impact not only users' decision to opt-in but also the amount of discretionary personal data that users provide (on the form) (Adjerid, Acquisti and Lowenstein 2019). For example, although users exposed to informative communications are less inclined to opt-in, they might concede more non-mandatory personal data than users exposed to persuasive cues, because the transparency of the communication reassures them.

We investigated this possibility in the same field experiment where we ask prospects who accept privacy conditions to complete a form requesting some personal data. As noted earlier, they could fill out a maximum of 18 different (non-mandatory) fields. In other words, prospects who opted-in had complete freedom over how much personal information they wished to share with the company.

We collected data on the average number of non-mandatory fields of the form completed by prospects who opted-in. Interestingly, our findings indicate no significant differences across conditions in terms of the average total number of non-mandatory personal data items provided ($F(5, 368)=1.34$, $p=0.245$). Similarly, the number of non-mandatory sensitive personal data items provided did not differ significantly across conditions ($F(5, 368)=1.59$, $p=0.162$).[11] These results suggest that the type of communication (informative vs. persuasive) does not significantly influence the amount of personal data, including sensitive data, that users choose to share. Our findings are in line with previous studies on cascaded privacy choices that show users do not compensate for upward decisions when facing downward decisions (Adjerid, Acquisti, and Lowenstein 2019).

[11] We refer to the literature to classify an information item as more or less sensitive (e.g., John, Acquisti, and Loewenstein 2011; Lwin, Wirtz, and Williams 2007). We consider email addresses to be less sensitive items and the type of car owned as a more sensitive item. Additionally, we run a separate one-way ANOVA for each type of information. The key result—no difference across treatments—remains unchanged.

In Study 1, we showed that a sizable portion of firms (49%) used persuasive cues in their re-permission emails. Study 2 rationalizes these practices, confirming that consumers are indeed more susceptible to persuasive messages with opt-in propensities. Consequently, Studies 1 and 2 show that firms are likely to behave "proactively," representing an unforeseen, and potentially, negative outcome generated by the new GDPR.

Study 3

In Study 3, we explore the costs and benefits of firms using persuasive cues. In line with our Figure 2, the benefits of employing persuasive elements include data access and return on data. Here, we gauge a firm's data access capability and its ability to monetize data collected. The expected costs for a firm incorporating persuasive cues into their opt-in requests depend on firm reputation and potential sanctions. These costs arise from the ability of consumers and regulatory bodies to discern the company's intention to sway users towards opting in. This could result in escalated costs, especially for well-known or previously privacy-violating firms. In the next sections, we provide a detailed description of how the benefits and costs of utilizing persuasive cues to gain opt-ins are measured.

Measuring the Expected Benefits of Using Persuasive Cues

Data Access

To measure a firm's data access capability, we assess the number of cookies placed by firms (or their partners) on their website. The vast majority of online data that we collect via cookies, placed on a wide variety of websites, often aim to profile consumers (Neumann, Tucker, and Whitfield 2019). Firms place first-party cookies on websites that users visit, while partners of the firm place third-party cookies. For example, a website may partner with advertisers to deliver ads or with an analytics company to help understand how people use their site. We collect both first- and third-party cookies via a partnership with Cookiebot.com, a cloud service provided by Cybot that automatically detects all cookies and similar trackers on a firm's website and distinguishes between different types of cookies. Cookiebot crawls online pages that represent the company's domain and records the number and type of active cookies present on that website. We used Cookiebot to scan up to 1,000 pages randomly selected for each of the 1,396 firms' domains in two snapshots, yielding a total of 980,182 pages analyzed.[12] After its scans, Cookiebot gave us information about the number, kind, and nature of cookies used by sample firms. Cookiebot distinguishes between four different types of cookies: (i) necessary (i.e., used for a website's basic functions that cannot be turned off), (ii) statistical (i.e., allowing website owners to understand how visitors interact with the page), (iii) preferences (i.e., enabling a website to remember users' preferences, such as preferred language or country of residence), and (iv) marketing (i.e., used to track visitors across websites and display ads that are relevant to individual users). We focus our analyses on marketing cookies as their intent is to enable companies' data-harvesting strategies.[13]

Cookiebot also provided us information about the cookie duration, including the number of temporary and persistent cookies.[14] Cookies can be either session-specific (temporary and existing only when the browser is open) or persistent (created to last across sessions to collect more information about users' online behavior). Tracking functions are typically associated with persistent cookies (Rutz, Trusov, and Bucklin 2011).

Return on data

To measure a firm's ability to extract value from our data, we use a proxy measure represented by the expected monetary value of the website's traffic (expected online (ad)

---

[12] We captured two snapshots at different times to account for potential variations in the day and pages sampled. Using two measures for each firm, we find a robust correlation of r = .92, indicating consistent extractions irrespective of timing and page selection.

[13] In our study, 'data harvesting' refers to the willingness and actions of a firm to collect and utilize personal data information from visitors to its website. As a measurable proxy for this concept, we employ the number of marketing cookies placed on the company's website. These cookies, typically used for tracking and personalization purposes, serve as an indicator of the extent to which a firm engages in the systematic collection of user data for marketing or other analytical purposes

[14] Figure WE1 in the Web Appendix E provides the distribution and some descriptive statistics about total marketing cookies (panel A) and persistent marketing cookies (panel B) used by companies in our sample.

revenues). We obtained this proxy measure from SEMrush,[15] an online platform that uses a panel of 5 million users worldwide to provide digital marketers with a wide variety of information collected from 790 million domains. A top website for online marketing services, SEMrush has won numerous industry awards in the digital marketing domain. Information provided by SEMrush is based on Google Ad Network data, allowing us to extract information about the value of advertisements placed through Google AdSense.[16] In addition, SEMrush provides an estimate of expected online advertising revenues at the firm's official website *j* as follows:

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

$Ad\ Revenues_j$ represents the traffic of website *j*, the cost per click (CPC), and click-through rate (CTR) in a specific industry and country. Therefore, it represents the expected potential future revenue streams from advertising that website *j* can generate on Google Ad Network. Therefore, it should not be interpreted as the company's actual revenues from website *j* but rather as the potential revenues that the company could obtain through advertising. In other words, $Ad\ Revenues_j$ represents the potential revenue that website *j* could make if the company decided to monetize the website by publishing advertisements via Google AdSense. Additionally, a firm's website may act as either a publisher, an advertiser, or both. Our sample of firms includes companies that act mainly as publishers (e.g., the *New York Times*), advertisers (e.g., Under Armour), or both (e.g., Google). Again, we stress that advertisers could also estimate the worth of a firm's official website and estimate its "potential" advertising revenue. Figure WE3 in the Web Appendix E reports the distribution

[15] We also use data from a second source (Rank2Traffic) as a robustness check.

[16] We note that the company could generate ad revenue by going through other ad networks. However, Google Ad Network dominates the market for ad selling and has the highest market share across all ad networks worldwide, excluding China (Statista 2021, pp. 51–55; Statista 2022). We therefore believe that the potential ad revenue based on the Google Ad Network is a good proxy for ad revenue across all networks.

of *Ad Revenues*$_j$ for our sample of firms.

Measuring the Expected Costs of Using Persuasive Cues

Firm Reputation

We argue that well-known companies often place a higher value on preserving their established online reputation. Consequently, we gather data on a firm's website popularity. One of the most widespread measures used in academic papers and the business press to measure a website's popularity is the Alexa Traffic Ranking (Libert and Nielsen 2018; Peukert et al. 2022). The Alexa rank is an Amazon proprietary measure that combines website-traffic statistics with visitor-engagement data—estimated using a panel of global users—over a period of three months. It returns a metric that can be used to compare websites' popularity over time, with rank 1 being the most popular. We collect this information by using the official Amazon AWIS API. For each company in our database, we record the Alexa Rank metric from January 2018 to December 2018 on the first day of each month. We then compute the average Alexa rank at the firm level. Figure WE2 in the Web Appendix E shows a well-dispersed distribution of the average Alexa rank. We additionally use SEMRush to gather pre-GDPR website traffic data for firm j, serving as a robustness check for the company's online popularity.

Sanctions

Past data breaches have been considered to assess anticipated costs associated with employing persuasive cues and evaluating resulting reputational harm. Thus, we collected information about the number of data breaches (if any) experienced before the GDPR enforcement by companies in our sample. We used different data sources (Prilock, Have I Been Pwned, and Wikipedia) to compile a list of these breaches. Privacy data breaches are public information that can severely damage a company's reputation if they are not well managed and if they attract consumers' and regulators' attention.

Controls

We incorporate a range of controls to account for company characteristics. These encompass firm type (digital products, physical products, and services), distribution strategy (purely digital firms vs. those operating physical stores), and other variables like company size, foundation year, country, product category, and competition index. For a detailed list and descriptions of these controls, refer to the Web Appendix F.

Model Specification and Estimation

Our purpose is to model the proportion of persuasive elements used in GDPR re-permission email *i* from company *j*. From Study 1, we obtained $Persuasive_{ij}$ (a variable with values between 0 and 1) that shows the degree of persuasiveness of the text of email *i* (see Figure WE4 in the Web Appendix E for the variable distribution). Highly persuasive emails possibly contain more than one persuasive cue for high values of this variable. By contrast, if persuasive cues are not used at all, then that variable equals 0. Given the nature of our DV, which varies between 0 and 1, we opted for a fractional logit model (Papke and Wooldridge 1996) and explain our choice in the Web Appendix G. This model is an extension of the generalized linear model with a nonlinear functional form (e.g., the logistic link function). The model is as follows:

$$E(x_{ij}) = G\left(x_{ij}\beta\right) \; \forall i,j \tag{1}$$

where $G(\cdot)$ is a known function satisfying $0 < G(z) < 1$ for all $z \in R$, such as the logistic function of the form $G(z) \equiv \wedge (z) \equiv \frac{exp\,(z)}{1+exp\,(z)}$. The model is then estimated through a quasi-likelihood method by maximizing a Bernoulli log-likelihood function of the form

$$l_{ij}(b) \equiv y_{ij} loglog\left[G\left(x_{ij}\beta\right)\right] + \left(1 - y_{ij}\right) log\left[1 - G\left(x_{ij}\beta\right)\right] \tag{2}$$

Equation (3) describes our model:

$$E(x_{ij}) = \frac{expexp\left(\beta_0 + \beta_1 Informative_{ij} + \sum_{k=2}^{K} \beta_k Benefits_{kj} + \sum_{l=K}^{L} \beta_l Costs_{lj} + + \sum_{m=L}^{M} \beta_m Controls_{mj}\right)}{1 + expexp\left(\beta_0 + \beta_1 Informative_{ij} + \sum_{k=2}^{K} \beta_k Benefits_{kj} + \sum_{l=K}^{L} \beta_l Costs_{lj} + + \sum_{m=L}^{M} \beta_m Controls_{mj}\right)} \tag{3}$$

where $Informative_{ij}$ is a variable varying between 0 and 1 that can be interpreted as the degree of informativeness of the text of email i. This variable is obtained from the analyses conducted in Study 1. We note that a re-permission email i could be characterized by both persuasive and informative cues or neither of the two; therefore, the variable $Informative_{ij}$ should not be interpreted as a complement of $Persuasive_{ij}$. $Benefits_{kj}$ indicates the set of variables related to the benefits associated with using persuasive cues from firm j in privacy campaigns—namely data access and returns on data measured, respectively, as the total number of marketing cookies and the expected $Ad\ Revenues_j$ of the website of company j.[17] $Costs_{lj}$ represents the set of variables used to express the costs associated with the use of persuasive elements in re-permission campaigns, measured as the number of data breaches and website popularity before 2018. Finally, $Controls_{mj}$ represents control factors such as the country of the headquarters of the firm, the industry, and so forth.

Results

In Table 8, Panel A, we report our results of the fractional logit regression analysis described in Equation 3. Notably, there is a positive, significant relationship between marketing cookie usage and expected online revenues when using persuasive cues in opt-in requests. In line with our conceptual framework, firms exploiting personal data value benefits more than costs involved in using persuasive messages. However, Panel A also shows a negative and significant interaction of data breaches and website popularity. These findings suggest that popular firms, which faced more data breaches before GDPR enforcement, tend to rely less on persuasive cues, potentially to rebuild trust and maintain a positive user reputation[18].

---

[17] Equation 3 omits firm-level fixed effects. In our sample, 94% of firms have a unique observation, 5% appear twice, and just 1% have more than two observations. Hence, for 94% of cases, i equals j. As a robustness check, we re-estimate excluding firms with multiple emails; when we do so, our results remain materially consistent.
[18] Further exploratory analyses, available upon request, reveal that when informativeness is employed as the dependent variable, the associations examined in our main analysis largely yield non-significant results. This outcome reinforces the centrality and distinct impact of persuasive strategies in the contexts explored, aligning with and substantiating the primary findings of our study.

Table 8
Fractional Logit Models' Results – DV = Data-Driven BERT Persuasive Topics (Equation (3))

Panel A: Full Model[A,B]

Variables	Measures	(1) Coef	(1) z	(2) Coef	(2) z	(3) Coef	(3) z	(4) Coef	(4) z
Benefits	ln(# Marketing Cookies)					**0.017**	**1.920**	**0.020**	**2.140**
	ln(Expected Monthly *AdRevenues$_j$*)			**0.039**	**2.320**			**0.040**	**2.380**
Costs & Risks	Website Popularity (3 months pre-GDPR)	0.000	0.750	0.000	1.520	0.000	0.680	0.000	1.480
	# Data Breaches (pre-GDPR)	0.170	0.970	0.080	0.440	0.175	1.010	0.083	0.460
	Website Popularity * # Data Breaches (pre-GDPR) [D]	**-0.000**	**-3.710**	**-0.000**	**-3.410**	**-0.000**	**-3.720**	**-0.000**	**-3.420**
DV=*Persuasive$_{ij}$* DV is measured through BERTopic analysis in Study 1		N=1506 LL= -499.29 BIC= 1313.21		N=1506 LL= -497.66 BIC= 1324.59		N=1506 LL= -498.72 BIC= 1319.40		N=1506 LL=-496.91 BIC= 1330.407	

Panel B: Results Distinct by Firm Type[A,B,C]

Variables	Measures	Service Coef	Service z	Product Coef	Product z	Digital Product Coef	Digital Product z	Physical Product Coef	Physical Product z
Benefits	ln(# Marketing Cookies)	**0.024**	**1.980**	0.013	0.830	-0.026	-1.370	**0.059**	**2.580**
	ln(Expected Monthly *AdRevenues$_j$*)	**0.034**	**1.670**	**0.055**	**1.860**	0.003	0.080	**0.067**	**1.660**
Costs & Risks	Website Popularity (3 months pre-GDPR)	0.000	1.640	0.000	0.420	0.000	-0.270	0.000	0.330
	# Data Breaches (pre-GDPR)	0.086	0.350	-0.154	-0.520	-0.570	-1.460	1.039	1.650
	Website Popularity * # Data Breaches (pre-GDPR)	-0.000	0.260	**-0.000**	**-2.490**	**-0.000**	**-1.690**	**-0.000**	**-3.110**
DV=*Persuasive$_{ij}$* DV is measured through BERTopic analysis in Study 1		N=968 LL= -310.86 BIC= 862.36		N=538 LL= -182.833 BIC= 585.741		N= 237 LL= -71.74 BIC= 280.19		N=301 LL= -105.74 BIC= 354.16	

Notes:
[A] For a robustness check using theory-based content analysis on persuasive cues, see the Web Appendix G, Table WG2.
[B] We included control variables in our analysis. See the Web Appendix F for the list of controls and a description, and the Web Appendix G for complete results, Table WG1.
[C] We offer a separate analysis by distribution strategy (Digital-only sales vs. Presence of physical sales channels) as a robustness check in the Web Appendix G, Table WG4. Results distinct by product categories are available upon request.
[D] Website Popularity ranges from 1 to 1,016,361, therefore allows even small coefficients to be significant. The interaction term coefficient of -.0000169, though small, indicates a substantial decrease in the use of persuasive cues in emails with increasing website popularity. For instance, the predicted probability of employing persuasive cues diminishes from 14.5% at a popularity rank of 10,000 to 10.2% at a rank of 1,000, given three data breaches.

Panel B differentiates between services and products, further segmenting products into digital and physical categories. Services are less cost-sensitive compared to products. However, a pronounced distinction emerges between firms offering digital versus physical goods. Physical product firms are more driven by benefits derived from exploiting personal data, while digital firms by costs in deciding whether to employ persuasive elements. These findings echo our analysis in Web Appendix G, comparing purely digital firms to those with online and offline outlets.

Discussion and Implications

Our Study 1 results indicate that 26% of firms rely solely on the use of persuasive cues in their opt-in requests, while an additional 24% combine persuasive and informative elements. This suggests that approximately half of the firms in our sample capitalize on the flexibility provided by the GDPR, choosing to incorporate persuasive cues in their opt-in requests. The reason for their doing so is clear: as our Study 2 results show causally, the best method for obtaining opt-ins is that of blending persuasive and informative cues in re-permission emails. By contrast, solely informative re-permission emails perform worse in terms of opt-ins. Study 3 results document some contingencies around this finding. Notably, the intention to harvest data and the opportunity to obtain high online ad-revenue streams promoted the use of persuasive cues. On the other hand, firms with higher levels of online traffic and popularity that incurred reputational damage in the past (via data breaches) are less likely to use persuasive cues, possibly to mitigate the effect of such data vulnerability (Martin et al. 2017). Thus, as costs seeking to maximize opt-in collection increase, the likelihood of using persuasion cues decreases dramatically. Figure 6 summarizes our key findings and relates them to our conceptual development, as outlined in Figure 2.

To gauge the effects of persuasion's benefits and costs, we run a policy simulation, assessing the impacts of varying each benefit and cost variable in our model. We simulate three levels for both benefits and costs/risks (low, medium, and high). Control variables are set to their mean values using a baseline industry.

Figure 5: Summary of Findings

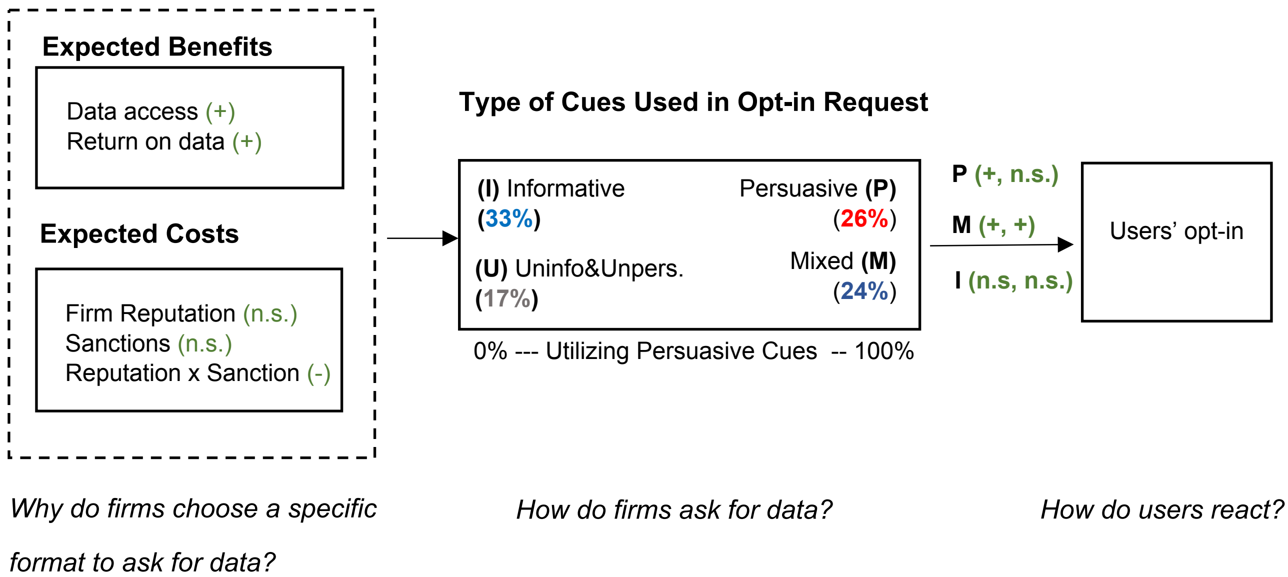
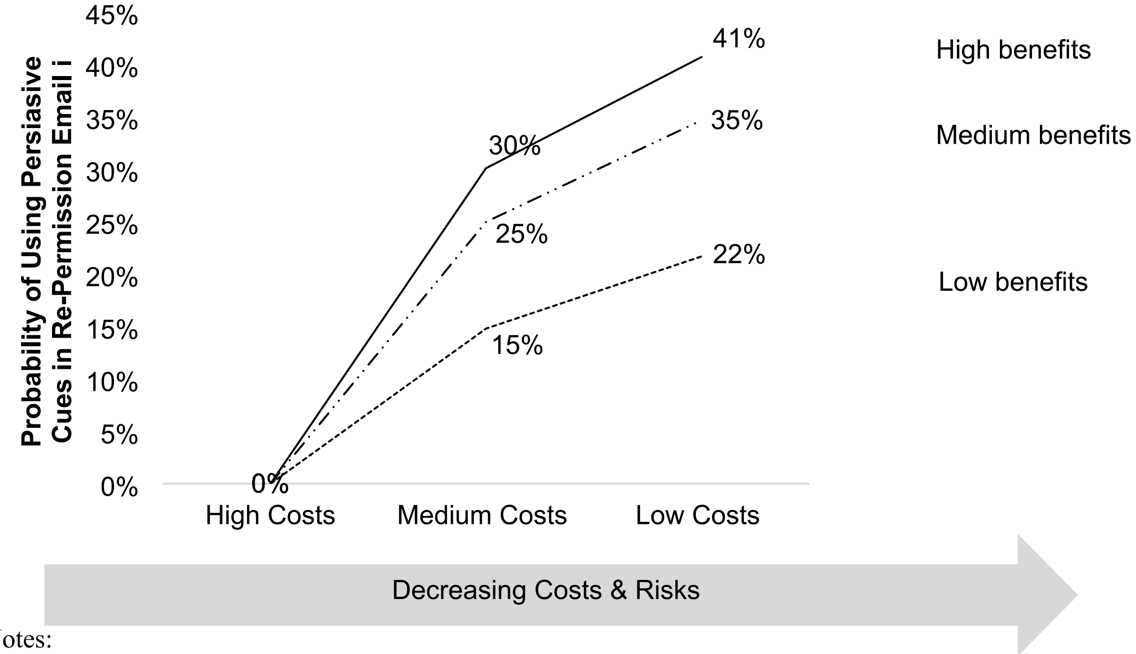


Figure 7, which shows the trade-off between benefits and costs, delineates the interplay between benefits and costs. As expected, when faced with high costs and low benefits, firms exhibit zero inclination to employ persuasive cues. Yet, as these costs wane, the propensity to use persuasive elements in re-permission campaigns grows. Even when there's a minimal strategic benefit from leveraging personal data, firms exhibit a 22% likelihood of employing persuasion given reduced risks (e.g., potential reputation damage, fines). This percentage is as high as 41% for firms with low costs and high benefits.

Further, these averages shift based on firm type. For example, in a high-benefit and low-cost landscape, 62% of firms specialized in home and furniture items with an offline store presence use persuasive cues, a percentage that drops to 17% in a low benefit-cost environment, suggesting that physical businesses discern greater value from data. Conversely, for purely digital firms that sell software items, the inclination ranges from 36% in the low scenario to 47% when benefits are high, and costs are low. Further insights can be found in Figure WG1 in the Web Appendix G.

Figure 6: Scenario Analysis: Variation in Costs and Benefits of Using Persuasion in Opt-in Requests



Notes:
For benefits, we used the variables *number of cookies* and *ad revenues*. For costs & risks, we used the *number of data breaches* and *website popularity*. To simulate the level for each variable, we took the value of the variables at the 5th, 50th, and 95th percentiles.

These results have implications for policymakers, as they indicate that mandatory opt-in requests can impact data access decisions inasmuch as firms may prioritize the use of persuasive over informative cues. Also, policymakers should note that purely digital businesses might not always actively pursue acquiring more data; their customer base might naturally grant data access, eliminating the need for further persuasion. On the other hand, businesses with a physical footprint often use persuasion techniques, aiming to harness the competitive advantage of monitoring users across both digital and offline touchpoints. Finally, we provide policymakers with a scalable and timely tool for probing opt-in requests.

Among our main findings is that using monetary incentives in exchange for data is an effective persuasive strategy to obtain opt-ins. This has implications for all relevant parties in online settings—regulators/policymakers, firms, and consumers.

For regulators, this is likely to be a call to re-examine how best to specify re-permission email protocols in ways that are fair to both firms and consumers (i.e., neither actor gains an advantage over the other). Notably, given our findings, the detection and possible action against monetary incentives is relatively easy for regulators.

For firms, our results show that using persuasive cues imposes a monetary cost (the incentive provided to opt-in). Moreover, using multiple persuasive cues simultaneously (e.g., tangible incentive and gain or loss frame) does not increase the likelihood of opt-in.

Finally, consumers should exercise caution when agreeing to opt-in for monetary incentives, especially from firms exploiting customer data more.

Conclusion, Limitations, and Directions for Future Research

This paper adds to the growing literature on the impact of privacy concerns on consumers, firms, and regulators. To the best of our knowledge, ours is the first attempt to investigate the effects of firms' freedom of choice in designing data consent requests. Our results documenting these effects have clear implications for regulators, consumers, and firms. While previous studies have explored specific contexts, our research is the first comprehensive exploration of the diverse strategies that different firms adopt to react to the enforcement of the GDPR. We assemble a rich set of data for a sizable sample of companies that, in 2018, had to gain and regain consent to meet the new higher standard for consent. We focus on the text of the opt-in email requests and analyze it using both unsupervised and supervised methods. Specifically, using a multi-method approach, we show how firms reacted to a new privacy policy aimed at user protection and control.

Our main finding is that, despite regulators' mandate for transparency and control in data request messages, 26% of firms relied solely on persuasive cues in their opt-in requests, while 24% of firms combined informative cues with persuasive elements. Thus, about half the firms in our sample reacted by managing the policy itself (i.e., need to obtain opt-in) instead of changing their behavior as the regulation intended. In particular, firms took advantage of the freedom left by the GDPR policy. This suggests that firms utilized persuasive cues in their messages because they rightly saw doing so as a rewarding strategy. Our field experiment findings are important, as they show that users respond favorably to the mixed use of persuasive and informative cues. In contrast, purely informative messages did not increase opt-in likelihood. This suggests that many firms (33% in our sample) that rely solely on informative cues might not achieve desired consumer opt-in, even if they adhere to regulators' directives for clear data request messages.

Overall, we see a pattern of unintended consequences as a result of the GDPR. Half of the firms try to obtain opt-ins via utilizing persuasive cues. Many firms stick to the informative cues— but may not gain or could even experience detrimental effects on data access. This outcome is likely to influence the privacy practices of these firms in the future, probably leading them away from using informative cues and toward using persuasive ones.

Our results also show that mixed strategies whereby informative and persuasive cues are merged are a rewarding strategy. This implies that firms that fail to gain access to data by using merely informative cues might instead add persuasive cues to their requests. Doing so could potentially provide them with more data, while limiting their risk of being sanctioned by regulators.

Interestingly, while our core finding underscores a penchant for firms leveraging persuasive cues to use personal data extensively, we also find pronounced differences in the strategies adopted by firms when we analyze product nature (digital vs. physical) and distribution channels (online-only vs. those with offline touchpoints). Perhaps counterintuitively, businesses grounded in the physical product realm or those maintaining physical stores are more inclined to use persuasive cues in opt-in requests. Their behavior could be driven not just by the allure of personal data but by the broader strategic paths it offers. One might argue that these businesses are chasing a holistic, 360-degree view of their customers, for they perceive a seamless, unified customer experience as a critical competitive edge. This vision, then, likely propels them to employ more direct and compelling strategies to solicit data consent, as they aim to maximize opt-in rates.

Our study of the post-GDPR environment can open future lines of inquiry. As firms that used data to generate revenue streams in our study were also more likely to use monetary incentives, a further avenue of research could investigate whether these money-for-data exchanges were fair and whether the value of this information was accurately priced. Researchers could examine the net effect of a regulation that requires opt-in (e.g., GDPR) versus privacy policies that do not include this explicit choice (e.g., CCPA) when users can be persuaded to grant data access.

Additionally, our results show that informative cues do not increase the likelihood that users will grant data access, indicating that strict opt-in privacy that imposes mandatory formats would limit possibilities of obtaining data consent. Namely, the treatment effect of the policy would be negative for the growth of a business. This finding could be further explored by examining the net effect on consumers who will not be exposed to personalized offers. In other words, further research could use our findings to enrich the growing theoretical literature examining the net effect of privacy regulation on consumer surplus (e.g., Hoffmann, Inderst, and Ottaviani 2020).

Our work has some limitations. In Study 1, we sourced a diverse re-permission email sample that showcased a breadth of company sizes and types, underscoring the GDPR's widespread influence. However, absent direct collaboration with these firms, our dataset lacks granularity with respect to the target audience of each email. For example, to the extent that it matters, our dataset doesn't distinguish between messages intended for prospective versus existing customers. Second, our Study 2 benefits from a unique collaboration with a prominent company selling physical goods, a domain where we show firms tend to employ more persuasion in data consent. This partnership provided a rich context to investigate opt-in requests. Nevertheless, the nature of field experiments inherently presents challenges in terms of generalizability. Securing collaborations is especially daunting when examining the delicate subject of privacy communications. To alleviate this, we conducted a lab experiment to explore the broader applicability of our field experiment findings, focusing on the digital domain (full details are in Web Appendix C). Interestingly, similar to the field experiment, the solely informative message did not significantly increase the likelihood of opting in. The main effect of utilizing exclusively persuasive cues was not replicated in the lab. However, messages combining persuasive and informative elements proved to be effective. Although the lab experiment cannot be viewed as a replication due to the variation in methodology, setting, and subjects involved, it is encouraging to observe consistent results in both the field and lab. Specifically, solely informative messages do not generate more opt-ins. In other words, to increase the chance of opt-in, firms need to use persuasive elements in their messages. Future research should further investigate the conditions under which messages using mixed (persuasive and informative) versus solely persuasive cues are most effective. Conducting heterogeneity analyses will help understand which subjects respond more or less effectively to mixed messages. Additionally, exploring a diverse range of industries would enhance our understanding of the generalizability of persuasive communication strategies in data consent requests.

Study 3 delves into potential motivations leading firms to use persuasive elements when seeking opt-ins. This study uses cross-sectional data and primarily offers descriptive insights. Despite our efforts with novel data coupled with several control variables to highlight possible mechanisms, we nevertheless cannot ascribe causal reasons for firm actions.

References

Acquisti, Alessandro, Alan Friedman, and Rahul Telang (2006), "Is There a Cost to Privacy Breaches? An Event Study," *ICIS Proceedings*.

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, Idris Adjerid, Rebecca Balebako, Laura Brandimarte, Lorrie Faith Cranor, Saranga Komanduri, Pedro. G. Leon, Norman Sadeh, Florian Schaub, Manya Sleeper, Yang Wang, Shomir Wilson (2017), "Nudges for Privacy and Security: Understanding and Assisting Users' Choices Online." *ACM Computing Surveys*, Vol. 50, No. 3.

Acquisti, Alessandro, Leslie K. John, and George Loewenstein (2012), "The Impact of Relative Standards on the Propensity to Disclose," *Journal of Marketing Research*, 49 (2), 160–74.

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

Adjerid, Idris, Alessandro Acquisti, and George Loewenstein (2019), "Choice architecture, Framing, and Cascaded Privacy Choices," *Management Science*, 65(5), 2267-2290.

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*, 1–11,

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.

Angel, Gary (2018), "The other 180°: Filling in the other half of the 360° customer view." *Applied Marketing Analytics,* 4 (2), 136-148.

Athey, Susan, Christian Catalini, and Catherine Tucker (2017), "The Digital Privacy Paradox: Small Money, Small Costs, Small Talk," Working Paper No. w23488, National Bureau of Economic Research.

Aziz, Arslan, and Rahul Telang (2015), "What Is a Cookie Worth?," workshop on the Economics of Information Security.

Benson, Vladlena, George Saridakis, and Hemamaali Tennakoon (2015), "Information Disclosure of Social Media Users: Does Control Over Personal Information, User Awareness and Security Notices Matter?," *Information Technology and People*, 28 (3), 426–41.

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

Bleier, Alexander, and Maik Eisenbeiss (2015), "Personalized Online Advertising Effectiveness: The Interplay of What, When, and Where." *Marketing Science* 34(5),669-688.

Bleier, Alexander, Avi Goldfarb, and Catherine Tucker (2020), "Consumer Privacy and the Future of Data-Based Innovation and Marketing," *International Journal of Research in Marketing*, 37 (3), 466–80.

Brehm, Jack W. (1966) "A Theory of Psychological Reactance" Academic Press, New York.

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

Catapang, Jasper Kyle, and Jerome V. Cleofas. "Topic modeling, clade-assisted sentiment analysis, and vaccine brand reputation analysis of COVID-19 vaccine-related Facebook comments in the Philippines." *2022 IEEE 16th International Conference on Semantic Computing (ICSC)*. IEEE, 2022.

Chaffey, Dave (2018). GDPR example privacy policies and consent forms. Smart Insights. Retrieved May 10, 2024, from [https://www.smartinsights.com/marketplace-analysis/digital-marketing-laws/gdpr-example-privacy-policies-and-consent-forms/]

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, 181–202.
Collis, Avinash, Alex Moehring, Ananya Sen, and Alessandro Acquisti (2021), "Information Frictions and Heterogeneity in Valuations of Personal Data," November 30. Available at SSRN: https://ssrn.com/abstract=3974826
Choi, W. Jason, Jerath, Kinshuk, and Miklos Sarvary (2023), "Consumer Privacy Choices and (Un)Targeted Advertising Along the Purchase Journey," *Journal of Marketing Research,* 60 (5), 889-907.
Cohen, Jacob (1960), "A Coefficient of Agreement for Nominal Scales*," Educational and Psychological Measurement,* 20, 27-46.
Davis, Ben (2018), "GDPR: 15 (Good & Bad) Examples of Repermissioning Emails & Campaigns," *Econsultancy* (April 23), https://econsultancy.com/gdpr-examples-repermissioning-emails-campaigns/.
Devlin, Jacob, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova (2018), "Bert: Pre-training of deep bidirectional transformers for language understanding." arXiv preprint arXiv:1810.04805.
Dinev, Tamara, and Paul Hart (2004), "Internet Privacy Concerns and Their Antecedents—Measurement Validity and a Regression Model," *Behaviour and Information Technology*, 23 (6), 413–22.
Drugov, Mikhail, and Marta Troya-Martinez (2019), "Vague Lies and Lax Standards of Proof: On the Law and Economics of Advice," *Journal of Economics and Management Strategy*, 28 (2), 298–315.
Fischhoff, Baruch (2015), "The Realities of Risk-cost-Benefit Analysis," *Science*, 350(6260).
Godinho de Matos, Miguel, and Idris Adjerid (2021), "Consumer Consent and Firm Targeting after GDPR: The Case of a Large Telecom Provider," *Management Science*, 68 (5), 3330-3378.
Goldberg, Samuel, Garrett Johnson, and Scott Shriver (2019), "Regulating Privacy Online: The Early Impact of the GDPR on European Web Traffic and E-Commerce Outcomes," SSRN Electronic Journal.
Goldfarb, Avi, and Catherine E. Tucker (2011), "Privacy Regulation and Online Advertising," *Management Science*, 57 (1), 57–71.
Grossklags, Jens, and Alessandro Acquisti (2007), "When 25 Cents Is Too Much: An Experiment on Willingness-to-Sell and Willingness-to-Protect Personal Information," *Workshop on the Economics of Information Security*.
Harford, Tim (2022), "How to Save More Lives and Avoid a Privacy Apocalypse," *Financial Times* (July 1), https://www.ft.com/content/fa2683ec-fca0-4c5f-805c-531f1412831e,
Hartmann, Jochen, and Oded Netzer. "Natural language processing in marketing." *Artificial Intelligence in Marketing*. Vol. 20. Emerald Publishing Limited, 2023. 191-215.
Hoffmann, Florian, Roman Inderst, and Marco Ottaviani. (2020), "Persuasion Through Selective Disclosure: Implications for Marketing, Campaigning, and Privacy Regulation," *Management Science*, 66 (11), 4958–79.
Jentzsch, Nicola, S. Preibusch, and A. Harasser (2012), "Study on Monetising Privacy. Heraklion: European Network and Information Security Agency (ENISA).
Jia, Jian, Ginger Zhe Jin, and Liad Wagman (2021). "The Short-run Effects of the General Data Protection Regulation on Technology Venture Investment." *Marketing Science*, 40.4: 661-684.
John, Leslie K., Alessandro Acquisti, and George Loewenstein (2011), "Strangers on a Plane: Context-Dependent Willingness to Divulge Sensitive Information," *Journal of Consumer Research*, 37 (5), 858–73.
Johnson, J. Erik, Bellman, Steven, and Gerald L. Lohse (2002), "Defaults, Framing, and Privacy: Why opting in-opting out," *Marketing Letters*, 13(1), 5-15.
Johnson, Garrett, Scott K. Shriver, and Samuel G. Goldberg, (2023), "Privacy & Market Concentration: Intended & Unintended Consequences of the GDPR," *Management Science*, 69(10), 5695-5721.
Johnson, Garrett, Scott K. Shriver, and Shooyin Du, (2020), "Consumer Privacy Choice in Online

Advertising: Who Opts Out and at What Cost to Industry?" *Marketing Science* 39(1), 33-51.

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 (August), 39–54.

Kummer, Michael, and Patrick Schulte (2019), "When Private Information Settles the Bill: Money and Privacy in Google's Market for Smartphone Applications," *Management Science*, 65 (8), 3470–94.

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

Lwin, May, Jochen Wirtz, and Jerome D. Williams (2007), "Consumer Online Privacy Concerns and Responses: A Power–Responsibility Equilibrium Perspective," *Journal of the Academy of Marketing Science*, 35 (4), 572–85.

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.

Marotta, Veronica, Vibhanshu Abhishek, and Alessandro Acquisti (2019), "Online Tracking and Publishers' Revenues: An Empirical Analysis," working paper, Workshop on the Economics of Information Security.

Marthews, Alex, and Catherine Tucker (2019), "Privacy Policy and Competition," Brookings paper.

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.

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.

Mohan, Bhavya, Ryan W. Buell, and Leslie K. John (2020), "Lifting the Veil: The Benefits of Cost Transparency," *Marketing Science*, 39 (6), 1105–21.

Multon, Karen D. (2010), "Interrater Reliability," Encyclopedia of Research Design, N. J. Salkind (ed.), Sage, Thousand Oaks, CA, 627–62.

Murgia, Madhumita (2022), "Google Closes Data Loophole Amid Privacy Fears Over Abortion Ruling," *Financial Times* (July 1), https://www.ft.com/content/d4c9b1d6-71a3-427c-acd9-75c3ffae8c18.

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.

Ovcharenko, Sergey (2021), "How Does GDPR Affect U.S. Companies in 2021," Alcor (September 27), https://alcor-bpo.com/legal-compliance-news/how-does-gdpr-affect-us-companies/

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.

Petty, Richard E., and John T. Cacioppo (1986), "The Elaboration Likelihood Model of Persuasion," in *Communication and Persuasion*. New York: Springer, 1–24.

Peukert, Christian, Stefan Bechtold, Michail Batikas, and Tobias Kretschmer (2022), "Regulatory Spillovers and Data Governance: Evidence from the GDPR" *Marketing Science* 41(4), 746-768.

Phelps, Joseph, Glen Nowak, and Elizabeth Ferrell (2000), "Privacy Concerns and Consumer Willingness to Provide Personal Information," *Journal of Public Policy and Marketing*, 19 (1), 27–41.

Rawnsley, Jessica (2022), "Reversing Roe Has Sharpened the data-Mining Threat," *Financial Times* (July 18), https://www.ft.com/content/915c6d2b-3035-4d0c-b018-32e793a176f1

Reynaert, Mathias, and James M. Sallee (2021), "Who Benefits When Firms Game Corrective Policies?," *American Economic Journal: Economic Policy*, 13(1),372–412.

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.
Sharma, Priyanka, Yidan Sun, and Liad Wagman, (2019), "The Differential Effects of Privacy Protections and Digital Ad Taxes on Publisher and Advertiser Profitability," Available at SSRN: https://ssrn.com/abstract=3503065
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, Jason Chuang, Spence Green, Marti Hearst, Jeffrey Heer, and Philipp Koehn, eds. Stroudsburg, PA: Association for Computational Linguistics, 63–70.
Statista (2021), "Digital Advertising Report 2021," Statista Digital Market Outlook, https://www.statista.com/study/42540/digital-advertising-report/.
Statista (2022), "Digital Advertising Market Share of Major Companies Worldwide 2016–2023," Statista Research Department (March 9), https://www.statista.com/statistics/290629/.
Sun, Tianshu, Zhe Yuan, Chunxiao Li, Kaifu Zhang, and Jun Xu (2024), "The Value of Personal Data in Internet Commerce: A High-stake Field Experiment on Data Regulation Policy," *Management Science*, 70(4), 2645-2660.
Thaler, Richard and Cass R. Sunstein (2008), *Nudge: Improving Decision Under Health*, Yale University Press.
Thornhill, John (2022), "The Threat of US Data Harvesting Grows as Abortion Ban Looms," *Financial Times* (May 26), https://www.ft.com/content/5f45a4e0-d9c0-4dbf-a1c7-bd86cfd0f605,
Tomaino, Geoff, Klaus Wertenbroch and Daniel J.Walters (2023), "Intransitivity of Consumer Preferences for Privacy," *Journal of Marketing Research*, (60), 489-507.
Tucker, Catherine (2014), "Social Networks, Personalized Advertising, and Privacy Controls," *Journal of Marketing Research*, 51 (5), 546–62.
Ucuzoglu, Joe, and John Hagel III (2020), "Use Your Customer Data to Actually Help Your Customers," *Harvard Business Review*, April 24.
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.

Web Appendices: The Race for Data: Informing or Persuading Users to Gain Opt-in?

Caterina D'Assergio

Puneet Manchanda

Elisa Montaguti

Sara Valentini

These materials have been supplied by the authors to aid in the understanding of their paper.

The AMA is sharing these materials at the request of the authors.

Caterina D'Assergio, Consultant, Information Technology Department, Marazzi Group Srl, Viale Regina Pacis, 39, 41049 Sassuolo (MO), Italy, email: caterina.dassergio@gmail.com

Puneet Manchanda, Isadore and Leon Winkelman Professor of Marketing, Ross School of Business, University of Michigan, R5490, 701 Tappan Street, Ann Arbor, MI 48109
Phone: 734–936–2445 Fax: 734–936–8716, email: pmanchan@umich.edu

Elisa Montaguti, Full Professor of Marketing, Management Department, University of Bologna, Italy, Via Capo di Lucca, 34, 40126 Bologna, Italy, Phone (+39) 051 209 8089, Fax: (+39) 051 246 411, email: elisa.montaguti@unibo.it

Sara Valentini, Associate Professor, Marketing Department, Bocconi University, Italy
Via Röntgen, 1, 20136 Milano, Italy Phone: (+39) 02 58362487 Email: s.valentini@unibocconi.it

Table of Contents

Web Appendix A: Analyzing Re-permission Emails - Theory-based Approach

Figure WA1: Distribution Plot of the Re-Permission Email's Sending Dates

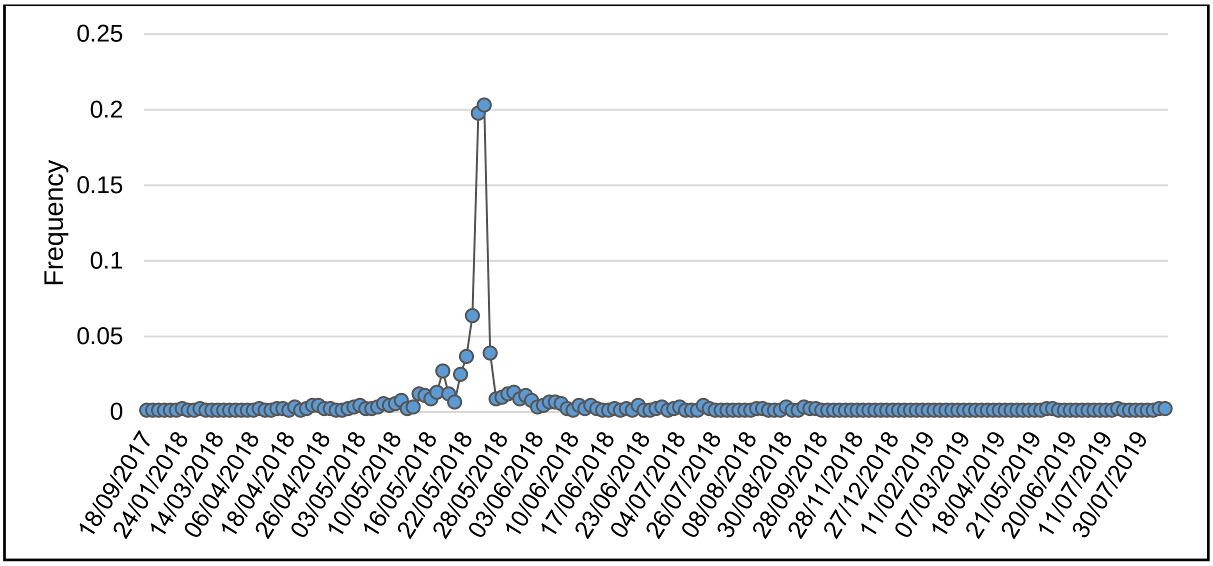
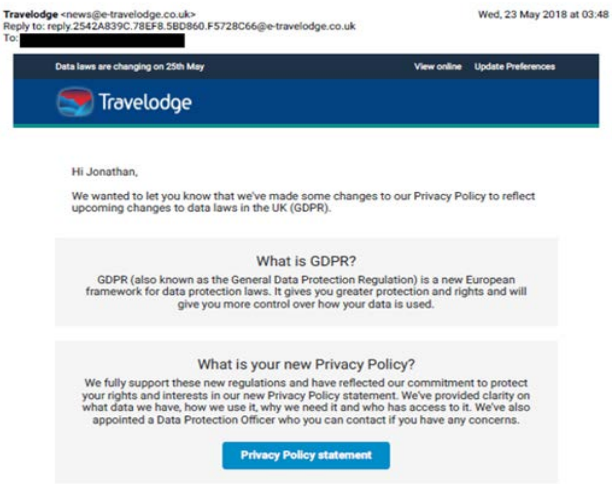
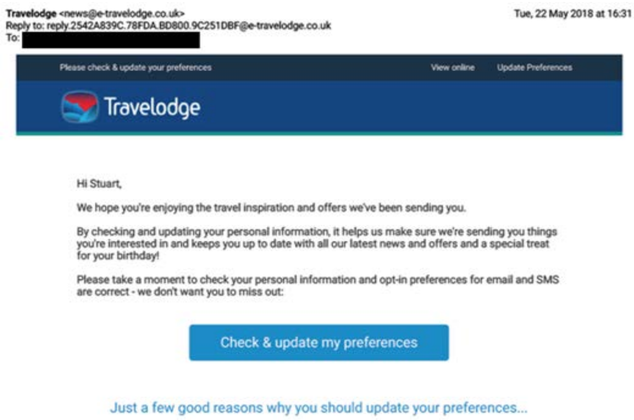


Figure WA2: Example of Different GDPR Re-permission Emails Sent by the Same Company

Re-permission Email 1



Re-permission Email 2



Notes: Firms may have different reasons for sending different versions of re-permission emails (e.g., different targets, different areas/countries, different time periods). As we do not know the reasons for this choice, we conducted a robustness check to test the sensitivity of our results by including the 6.6% of emails sent by the same company in our main analyses.

Theory-based Approach

Table WA1: 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 email highlights and stresses how the user can control personal data. Coded as 0 if the email provides only basic information about the possibility to control. Example: “We are giving you more control over your information. We have added information about how we share information and why we do it. We are making it easier for you to control the information you provide us with. Our policy explains how you can make choices about your information and the measures we’ve put in place to keep your information secure.”
	Transparency: Five-level variable assessing the level of transparency of companies in describing their data related activities and data security standards.	1	No or Minimum level of transparency. The email provides only the 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 limited and generic.
		3	Average level of transparency. Information is provided with references and links to sources to help readers 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 data protection domain.
Persuasive Themes	**Framing: Gain / Losses**	0 - 1	Coded as 1 if the email indicates the presence of gain frame, loss frame or both in the email. Gain and loss framing are two different ways of presenting information in persuasive communication. Gain framing focuses on the positive outcomes of a behavior, while loss framing focuses on the negative consequences of not engaging in that behavior. Examples: Gain-Framed Message: "By opting in for privacy, you’ll continue to receive personalized recommendations and exclusive offers based on your interests, leading to a more enjoyable and relevant online experience." Loss-Framed Message: "Choosing not to opt-in for privacy may result in potential risks to your personal information. You might miss out on receiving personalized recommendations and exclusive offers tailored to your interests.
	Framing: Time Orientation	0 - 1	Coded as 1 if the email has some form of time orientation - past, present or future, and 0 otherwise. Time orientation is a persuasive technique that utilizes temporal aspects of a message to influence the recipient's decision-making. It can involve highlighting long-term consequences of a behavior or employing an urgency-framed message to emphasize the importance of acting promptly,

	Communication Themes	Levels	Definition
			creating a sense of urgency to encourage quick decision-making.
	Monetary Incentives	0 - 1	Coded as 1 if the email provides a monetary incentive such as discounts or offers in the communication. Example: “Reconfirm your consent to let us know that you wish to continue to receive our e-mails. To thank you for the time you've spent, we offer a 25% discount-for your next order.”
	Non-Monetary Incentives	0 - 1	Coded as 1 if the email provides a non-monetary incentive such as invitation to events or free trials in the communication. (Chandon, Wansink, & Lauren 2000). Examples: • Gifts. This could include things like gift cards, company merchandise, or tickets to events. • Recognition. This could include things like public praise, awards, or certificates of achievement. Opportunities for advancement. This could include things like training and development opportunities

Figure WA3 Examples of How Re-permission Emails were Coded Using the Coding Protocol

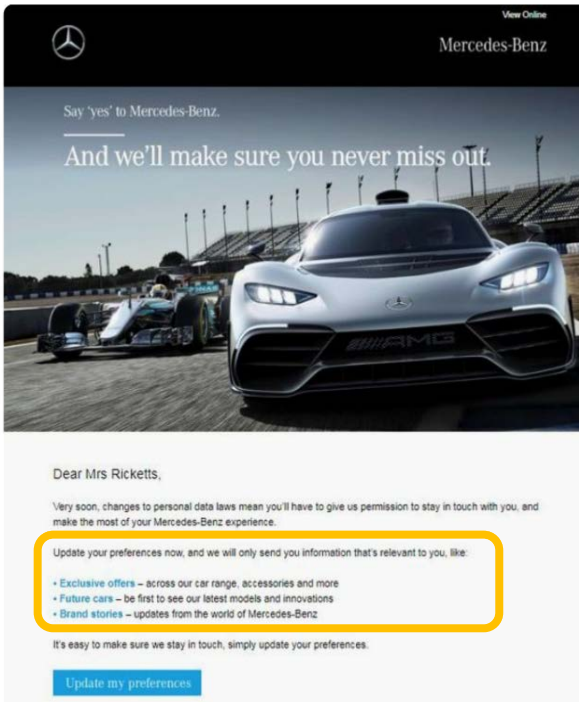
(Table WB1)

Transparency

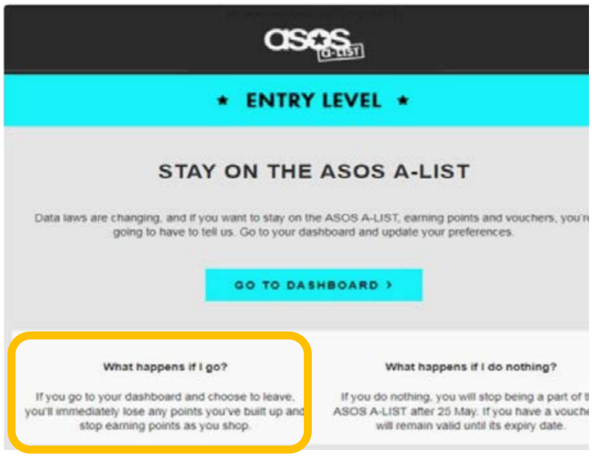
Level 1: Minimum Level of Transparency

Level 5: Very High Level of Transparency

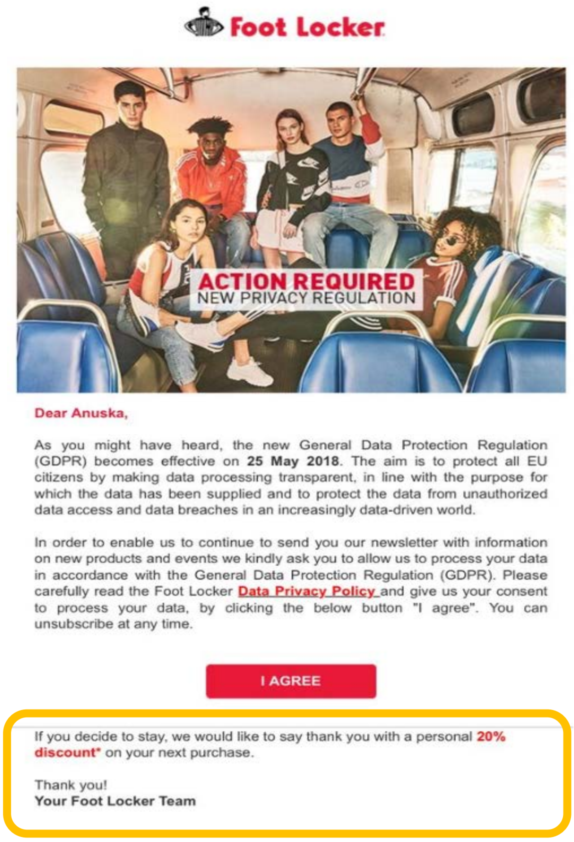
Gain Frame (1=present, 0=absent)

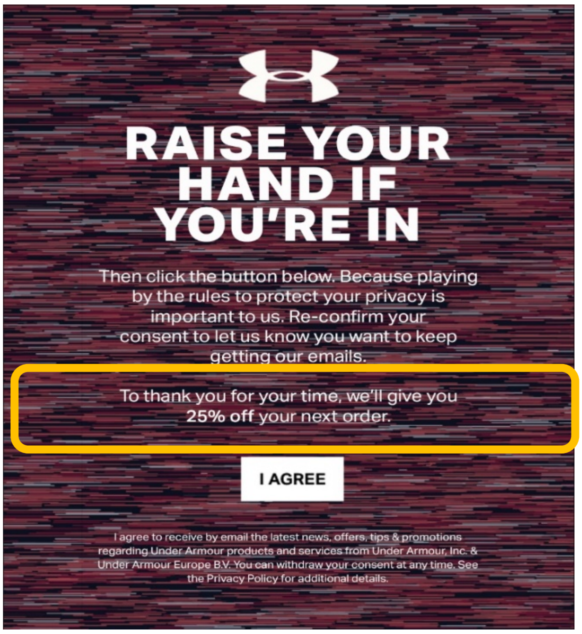


Loss Frame (1=present, 0=absent)



Monetary Incentives (1=present, 0=absent)





Web Appendix B: Analyzing Re-permission Emails Data-driven Unsupervised Approach

BERTopic: Workflow to Determine the Topics

After collecting the dataset of re-permission emails related to GDPR, we proceeded using the steps outlined below.

Step 1. Text Preprocessing

This step involved removing personal email references, new line characters, URLs (both HTTP and HTTPS), punctuation, and converting text to lowercase. The text was then tokenized into individual words. We also removed stop words and applied lemmatization using NLTK and Spacy.

Step 2. Word Frequency Analysis

We created a frequency distribution of words after removing stop words and plotted the most frequent words for an initial understanding of the dataset.

Step 3. Tokenization and Lemmatization

We applied tokenization and removed stop words using Gensim's simple_preprocess. Then, we performed lemmatization to keep only nouns, adjectives, verbs, and adverbs using Spacy. We created a dictionary of lemmatized words and filtered extremes to remove rare and overly common tokens.

Step 4. Embedding and Dimensionality Reduction

We used BERT to generate embeddings for each email, capturing semantic meanings. To make the embeddings manageable for clustering, we applied UMAP (Uniform Manifold Approximation and Projection) to reduce the dimensionality of the embeddings while preserving their structure.

Step 5. Topic Modeling and Clustering

BERT uses topic modeling to identify latent topics in a dataset. We used HDBSCAN (Hierarchical Density-Based Spatial Clustering of Applications with Noise) to cluster the document embeddings into topics. After initial topic clustering, we applied hierarchical clustering with Ward's linkage to group similar topics into broader categories (Grootendorst, 2020). The hierarchy produced by Ward's method consists of sub-hierarchies named clades (Dogan and Birant, 2021). These clades of topics are extracted by BERT and can be used to assign names to different topics (Catapang, Kyle, and Cleofas 2022). After constructing a dendrogram of the BERT topic embeddings, the hierarchy of the topics found within the data is illustrated in Figure WB1.

Figure WB1: Dendrogram of the BERT Topic Embeddings

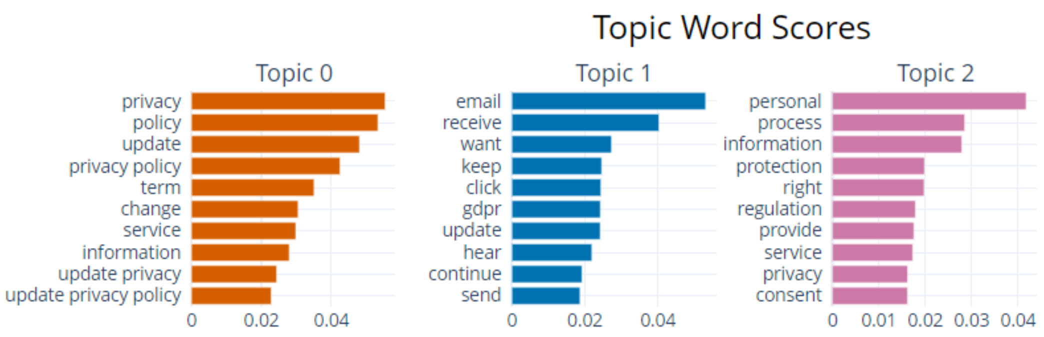
Step 6. Interpretation and Visualization

After exploring different possibilities, we decided to 'cut' the dendrogram at a specific threshold (depicted by the dashed line in Figure WB1), which yields three distinct groups. A three-group partition appeared to strike a balance between overly simplistic categorization and overly complex subdivisions, thus facilitating clearer interpretability. We also analyzed the probability that each document belonged to a given topic and inspected texts with extreme probabilities for clearer interpretation.

Step 7. Naming and Validation

We performed a partitioning of the original 13 latent topics identified by BERT into three macro topics. In the dendrogram, Topic 0 is visually represented by the green superclade and comprises latent topics 7, 4, 3, and 5. Similarly, Topic 1 is indicated by the red superclade and includes latent topics 0 and 8. Lastly, Topic 2 encompasses latent topics 10, 9, 6, 1, 2, 11, and 12. Although it is well known that the topics inferred by unsupervised methods are not always easily interpretable (Sievert and Shirley 2014, p. 64), we used several interactive visualization systems to facilitate an inspection of topics. For example, we used topic word score charts reported in Figure WB2.

Figure WB2: BERT Topic Word Scores



A topic words score chart is a visualization of the scores that BERT assigns to each word in a document, based on how likely that word is to be associated with a particular topic. The scores are represented as a bar chart, with the words that are most associated with the topic having the highest scores.

We began by inspecting Topic 1, which seems to be the most different from the others. Based on the words "receive," "want," "keep," and "click," Topic 1 seems to refer to “persuasive calls to action." It seems to include texts that are designed to encourage the receiver to take a specific action, such as clicking a button, or making an action. These kinds of text may appeal to a receiver's desire to receive something, to want something, to keep something, or to click something. They may also use words or phrases that are designed to create a sense of urgency or scarcity, such as "limited time" or "act now." Topic 1 therefore seems to be in line with the idea of using persuasive cues to persuade individuals to opt in. The distinction between Topic 0 and Topic 2 is more subtle as both seem to be focused specifically on privacy and the introduction of the GDPR. However, Topic 0 frequently includes the words “update” and “privacy”, whereas Topic 2 frequently includes “personal” and process”. To facilitate interpretation, we checked the probability that BERT associates with each email for Topic 0, 1, and 2 (see Table WB1 as an example). Topic probabilities represent the likelihood that a particular document belongs to a particular topic. For example, if a document has a topic probability of 0.8 for Topic 0, then there is an 80% chance that the document belongs to Topic 0. We, therefore, inspected the text classified as more extreme toward a topic, such as 100% and 0% for others (e.g., see email 1606 or 1610 in Table WB1), which allowed us to read a full text classified as fully Topic 0 or fully Topic 2.

Table WB1: Example of BERT Topic Probabilities Associated to Different Emails

Email ID	**Text**	**Topic 0**	**Topic 1**	**Topic 2**
1	Hi On 25 May 2018 The Gener...	22%	6%	17%
2	On May 25th, 2018, a new regulation is going t...	27%	4%	13%
3	Adestra \n\nMake sure you keep hearing from us...	32%	47%	20%
4	Good morning, \naizoOn - an innovation technol...	42%	13%	25%
5	algolia \n\nDear Smiles Davis, \nYou are recei...	21%	5%	74%
…				
1606	Confirmation, adding to the newsletter\nA requ...	0%	0%	100%
…				
1610	Updates to our Terms of Use and Privacy Policy...	100%	0%	0%

After this analysis, we conveyed that Topic 0 represents text aimed at providing a clear update about the changes in the GDPR. Indeed, the word "update" appears three times in the most frequently mentioned words in Figure WB1 for Topic 0. By contrast, Topic 2 seems to inform about the new GDPR but in a more legal style that fails to make the text clearer or specifically focused on the key new update about the privacy policy.

Based on these observations, we named Topic 0 as "Informative", Topic 1 as "Persuasive", and Topic 2 as "Legal Insights." We further explored the robustness of this classification by using the alternative, widely used, and unsupervised approach of LDA (described in the next section). Finally, we checked whether the results and our interpretation of both BERT topics and LDA topics aligned with the results of the manual content analysis.

Traditional LDA

In addition to using BERT, we conduct a robustness check by employing traditional LDA analysis to analyze the text of our sample of re-permission emails. LDA analysis represents a valid unsupervised topic model that has been employed across a wide variety of contexts for topic extraction in the marketing literature (see Hartmann and Netzer, 2023, and Shankar and Parsana, 2022 for a review on NLP approaches used in marketing).

We conducted the LDA topic modeling analysis by using the Gensim package available in Python. We first split the text into words (tokenization), then removed stop-words, and lemmatized the remaining words. We removed tokens with a frequency higher than 90% or lower than 1% given that they are too common or too rare to be useful for the analysis (Griffiths and Steyvers 2004; Lu et al. 2011). This yielded a final dictionary of 1,014 unique words. To identify the best number of topics (k), we estimated 100 models for each choice of k and generated their "coherence scores" (Kapadia 2019). We then calculated the mean and the median of the "coherence score" metric for each k. We find that k = 3 is the optimal choice based on the high coherence scores (Mean = 0.48; Median = 0.50) (AlSumait et al. 2009; Mimno et al. 2011; Puranam, Narayan, and Kadiyali 2017). We named Topic 1 "Informative," Topic 2 "Persuasive," and Topic 3 "Legal Insights." Results with the three topics identified and the likelihood of occurrence of each topic in the re-permission email text are provided in Table WB1, which also describes the procedure followed we used to interpret the topics.

Table WB2: Topics Revealed from LDA Analysis

Topic	**Bag of Words (Lemmatized)**	**Selected Label**	**Probability of topic *k* occurrence in the re-permission email *i***	
			Mean	**Standard Deviation**
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**	49.1%	0.319
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**	39.9%	0.323
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.	**Legal Insights**	10.7%	0.186

Note: The "bag of words" column reports thirty words with the highest probability of describing topic *k*. These words are listed in descending order of likelihood.

Interpretation of the LDA Results and the Three Topics

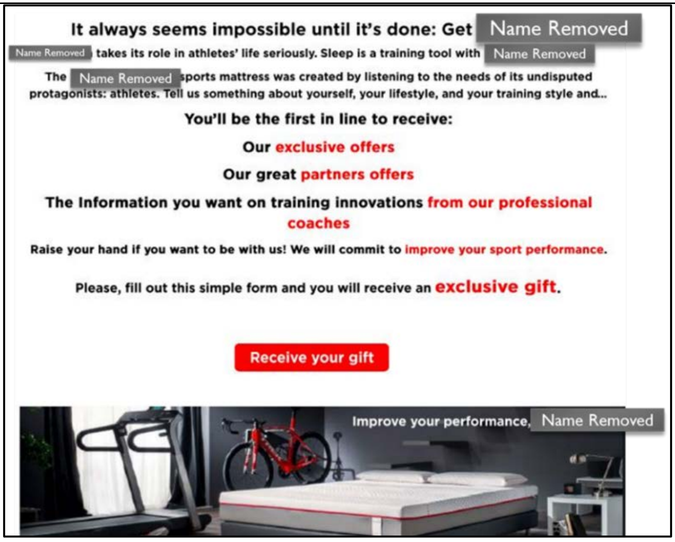
LDA is a probabilistic method. For each re-permission email, results provide a mix of topics that characterize the document (i.e. the email). Each word in the email is attributed to a specific topic with a probability. Each topic, therefore, is characterized by the set of words in the vocabulary that are more likely to describe it. Column 2 of Table WB2 reports the list of the first 30 words with the highest probability of describing each topic k.

We named Topic 1 "Informative" because it contains the most likely words: update, term, datum, change, transparency, and protect. We contend that these terms highlight a more informative focus of the communication (i.e., to inform users about the changes and the news brought about by the GDPR). By contrast, the set of words with the highest probability in Topic 2 differs. Here words include receive, click, offer, like, and stay. These words seem to suggest a communication designed to trigger a user's action (e.g., opt-in), providing an incentive (offer). Therefore, we named Topic 2 "Persuasive." Finally, we described Topic 3 as "Legal Insights" because we observe words that signal a more formal and legal style in communication (e.g., administrator, legal, authority, contract, obligation). Although it is not always easy to name topics inferred by LDA (Sievert and Shirley 2014, p. 64), we believe the identified labels well represent the three topics. To double-check our result, we randomly extract some emails from the dataset and check whether the probability associated with each topic fit with the text of the email. So, for example, if the re-permission email *i* obtained a probability of 0.8 for Topic 1, 0.2 for Topic 2, and 0.0 for Topic 3, then we should expect a highly persuasive email. This happens in all the cases that we randomly extracted. Additionally, we run the analysis described in the "Data-driven Unsupervised Approach" paragraph of the manuscript (see Table 5) to more formally cross-validate our theory-based vs. data-driven approaches.

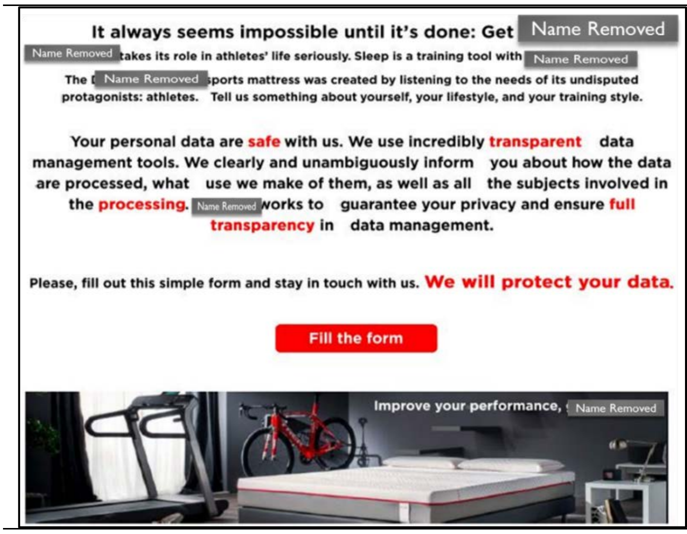
Web Appendix C: Field Test, and Lab Experiment

Figure WC1: Detailed Description of Exp Messages Used in the Field Test: G5 and G2 Taken as Examples

Panel A: Highly Persuasive Cues: Framing & Economic Incentive - (G5, Table 6)

Email sent to G5	***Description***
	The message contains three categories of persuasive cues: • Gain frame: "you'll be the first in line to receive…" • Time orientation: "you'll be the first, … we will commit to improve…" • Economic incentive: the gift is mentioned twice in the message. By clicking on the "receive your gift" button, the contact is re-directed in a web page describing the gift, in this page the prospect can also opt-in and fill the form. The gift mentioned is a 50% discount for the purchase of the "race number (bib)" for an upcoming sport event. The regular price of this kind of bib (gold bib) is 200 € and provides athletes access to better facilities at the race.

Panel B: Highly Informative Cues: Transparency, Clarity, Safety - (G2, Table 6)

Email sent to G2	***Description***
	The informative message mentioned the commitment of the company to protect personal data. Transparency, clarity, and safety of data protection is mentioned in this message and is absent in generic, moderately, and highly persuasive messages

Note: The text of the emails reported here is equivalent to the original text sent, but the layout and copy represented in this figure is just an example. The actual copy editing, and the name of the company are not revealed as the partnered company prefers to stay anonymous.
The economic incentive used (i.e. 50% discount for the purchase of the "race number") does not vary across conditions (i.e. G3, G4, G5, G6).

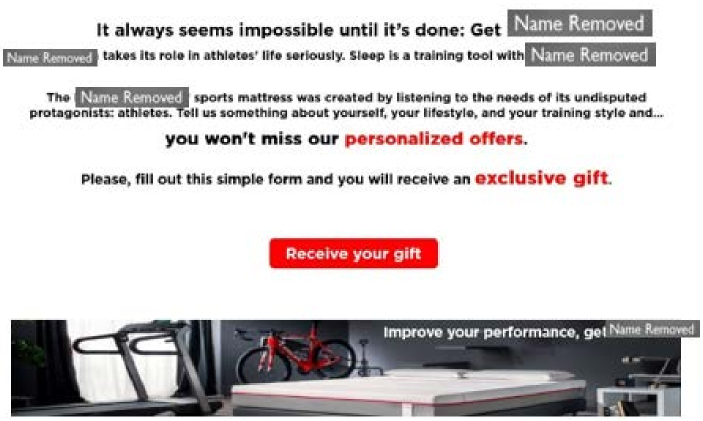
Figure WC2: Complete Set of Messages Used in the Six Experimental Groups

Experimental Groups — **Message**

G1 Generic Message

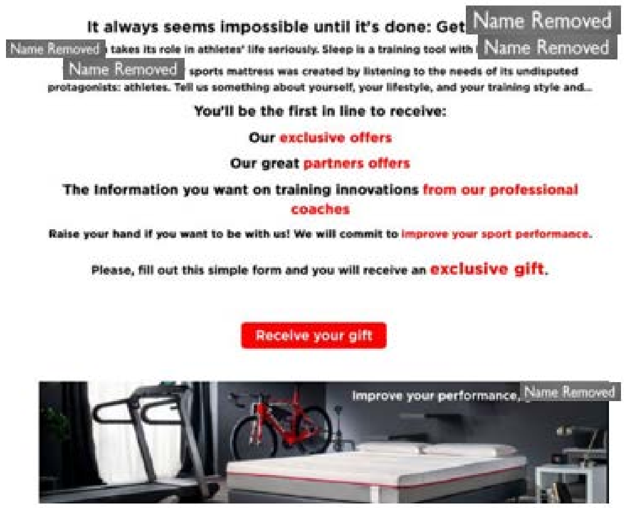
G2 Informative

G3 Moderately Persuasive

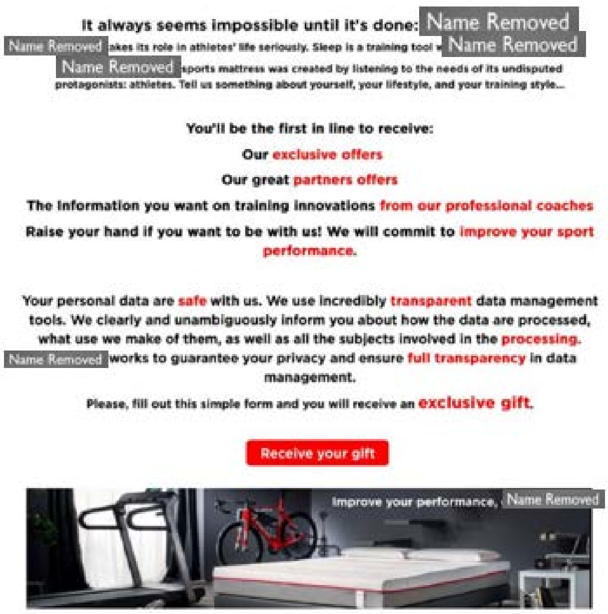


G4 Informative & Moderately Persuasive

G5 Highly Persuasive



G6 Informative & Highly Persuasive



Notes: For conditions G4 and G6 we control for order effects by reversing the order of appearance of the persuasive cues in the text.

Exploring Generalizability: A Lab Experiment on a Digital Product

We conducted a lab experiment (pre-registered at AsPredicted.org)[1] to assess the generalizability of our findings from Study 2 to industries beyond the one we examined in the field experiment. Our specific focus was on a predominantly digital firm, considering both product type and distribution channels. We selected Photobox as an illustrative example, clarifying at the outset of the survey that we chose this brand for demonstrative purposes, emphasizing that the survey is not affiliated with or endorsed by Photobox.

We recruited 305 participants from Prolific, aiming for approximately n=50 per between-subject cell. We recruited EU participants fluent in English. The experimental design described for the field experiment remained unchanged. Participants were randomly assigned to one of the six experimental groups outlined in Table 6. The text of the messages was kept very similar to that described in Figure WC2, with minor adjustments to suit the new firm. Additionally, we tailored the incentive as per the persuasive conditions (G3-G6), aligning it with a current promotion being run on the brand's website to enhance the realism of the scenario (See Figure WC3).

As pre-registered, we excluded eight participants who failed attention check question, leaving 296 participants. Table WC1 provides descriptive statistics for age and gender. Additionally, participants were asked to rate the frequency of their photo-taking habits on a 5-point scale (M_{photos}=3.25, SD_{photos}=1.09).

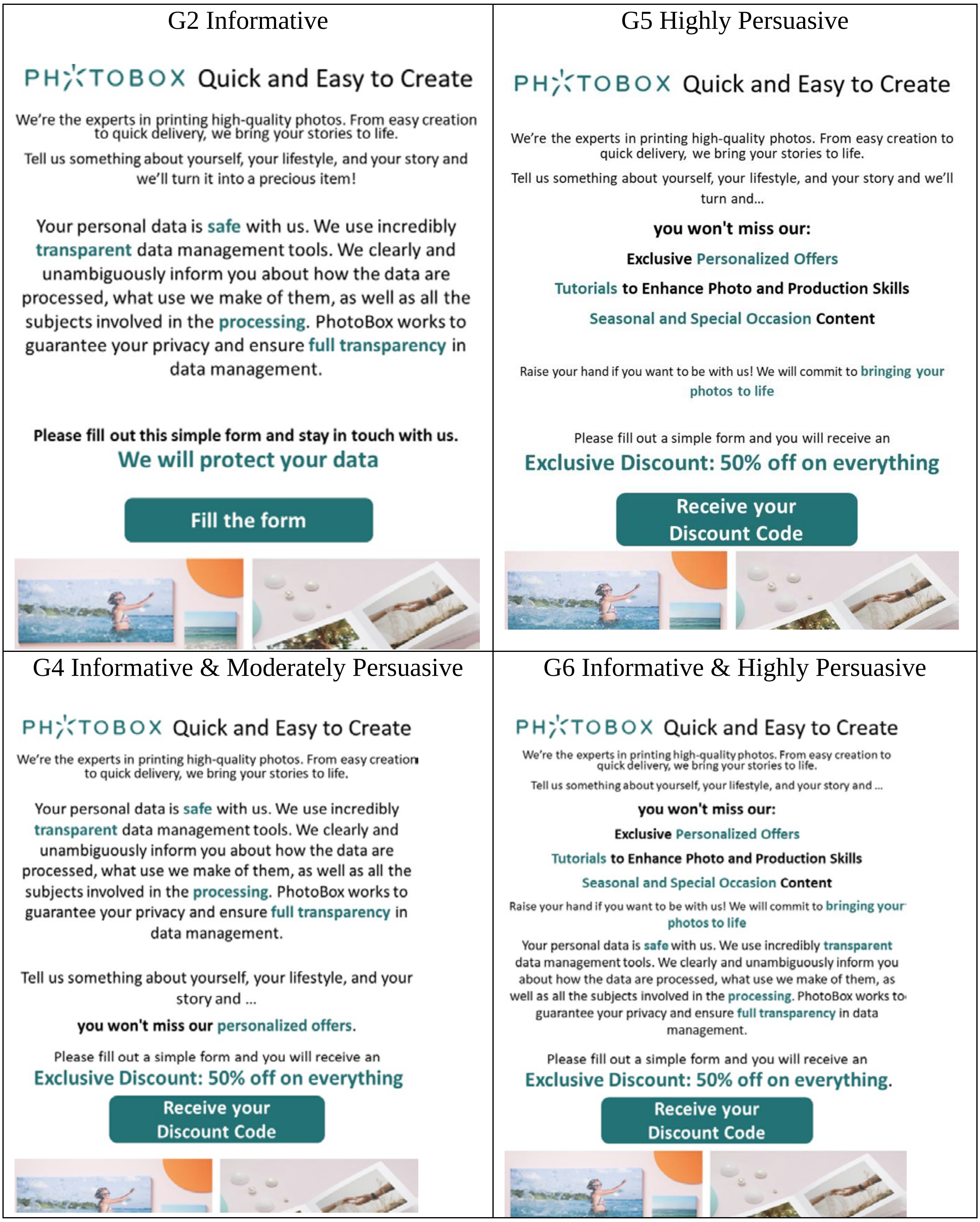
Table WC1: Descriptive Statistics for the N=296 Participants in the Lab Experiment

Age	Freq.	Percent	Gender	Freq.	Percent
18-25	116	39.19	Female	148	50.00
26-35	89	30.07	Male	147	49.66
36-45	42	14.19	Other	1	0.34
46-59	36	12.16			

[1] https://aspredicted.org/8t8hm.pdf

60+	13	4.39			
Total	296	100.00	Total	296	100.00

Figure WC3: Description of Messages Used in Lab Experiment: G2 - G6 Taken as Examples



Note: For conditions G4 and G6 we control for order effects by reversing the order of appearance of the persuasive cues in the text.

Participants were directed to envision themselves considering online photo printing services and initiating their exploration of Photobox, a popular photo printing company's website.

We measure intention to share personal data with two DVs. After exposure to one of the six messages, we assessed participants' intention to fill out a form by asking: 'How likely are you to fill out a form asking for your personal details (e.g. email) to share with PhotoBox?' (from 1 = 'extremely unlikely' to 7 = 'extremely likely'). Additionally, we measured intention to opt-in by asking: 'How likely are you to opt in to the terms and conditions outlined in the Photobox Privacy Policy?' (from 1 = 'extremely unlikely' to 7 = 'extremely likely').

Table WC2: Lab Experiment: OLS regression (n = 296) DV=Intention to Fill the Form

	Coefficient	t	p-value
Informative (G2)	0.23	0.59	0.557
Moderately persuasive (G3)	0.09	0.23	0.817
Highly persuasive (G5)	0.23	0.59	0.555
Moderately persuasive × Informative (G4)	**0.89**	**2.31**	**0.022**
Highly persuasive × Informative (G6)	**0.81**	**2.09**	**0.037**
Constant	3.85	14.04	0.000
N = 296, F(5, 290) = 1.96, pvalue= 0.085, R^2=0.03			

Table WC2 displays the results regarding the intention to fill out the form, showing that both conditions incorporating mixed cues (a persuasive cue—either highly or moderately persuasive—and an informative one) significantly enhance the likelihood to share personal information.

Table WC3 presents the results concerning the likelihood to opt-in. The findings align closely with those observed in Table WC2, albeit less strong. Only the condition combining moderate persuasiveness with informativeness showed statistical significance at the 10% level.

Table WC3: Lab Experiment: OLS regression (n = 296) DV=Intention to Opt-in

	Coefficient	t	p-value
Informative (G2)	0.21	0.59	0.553
Moderately persuasive (G3)	0.22	0.61	0.541
Highly persuasive (G5)	-0.02	-0.06	0.954
Moderately persuasive × Informative (G4)	**0.61**	**1.71**	**0.089**
Highly persuasive × Informative (G6)	0.52	1.43	0.153
Constant	4.17	16.23	0.000
N = 296, F(5, 290) = 1.06, pvalue= 0.3839, R^2=0.02			

We decided to delve deeper into whether the results were more pronounced for individuals with a greater interest in the product category. To achieve this, we specifically analyzed customers who rated their frequency of taking photos as >2. The results, as presented in Tables WC4 and WC5, confirm our previous findings and show a stronger effect of the mixed conditions for both DVs.

Table WC4: Lab Experiment: OLS regression (n = 199) DV=Intention to Fill the Form

	Coefficient	t	p-value
Informative (G2)	0.35	0.76	0.447
Moderately persuasive (G3)	0.36	0.81	0.421
Highly persuasive (G5)	0.31	0.64	0.523
Moderately persuasive × Informative (G4)	**1.27**	**2.68**	**0.008**
Highly persuasive × Informative (G6)	**1.27**	**2.70**	**0.008**
Constant	3.79	11.59	0.000
N = 199, F(5, 193) = 2.59 , R^2=0.06			

Table WC5: Lab Experiment: OLS regression (n = 199) DV=Intention to Opt-in

	Coefficient	t	p-value
Informative (G2)	-0.09	-0.21	0.835
Moderately persuasive (G3)	0.65	1.57	0.118
Highly persuasive (G5)	-0.05	-0.12	0.907
Moderately persuasive × Informative (G4)	**0.91**	**2.11**	**0.036**
Highly persuasive × Informative (G6)	**0.82**	**1.91**	**0.058**
Constant	4.12	13.78	0.000
N = 199, F(5, 193) = 2.35, R^2=0.06			

The results of our lab experiment can be viewed as a partial replication of our earlier field experiment (see Table 7 in the manuscript). Although we didn't replicate the main effect of moderately persuasive cues, messages with mixed cues (combining persuasive and informative elements) proved effective. Notably, a purely informative message didn't increase opt-ins significantly compared to a generic one, similarly to what we obtained with the field experiment.

In essence, while the main effect wasn't replicated, the core message remains consumers respond to persuasive messages indicating. Additionally, our study highlights a category-specific effect. For digital and sensitive data products like personal photos, individuals respond positively to persuasive cues only when coupled with information about data collection and management.

Although we hesitate to categorize this as a straightforward replication due to variations in subject demographics, methods, timing and scenario-based incentive vs. actual incentive (in the field test, participants were offered a tangible incentive, whereas in the lab test, participants were asked to imagine receiving the incentive). Nevertheless, it is encouraging to observe a high degree of consistency across most of our findings.

Web Appendix D: Pre-Testing Experimental Conditions

Table WD1: Sample 1 (N=188) Anova Analysis Results

Experimental Cells			**Perceived Persuasiveness**		**Perceived Informativeness and Transparency**	
Group	**Description**	**N**	**Mean**	**SD**	**Mean**	**SD**
G1	Generic (Low Low)	48	2.44	0.85	2.00	0.84
G2	Informative	48	2.59	0.73	3.40	0.80
G3	Moderately Persuasive	44	3.46	0.56	2.00	0.81
G4	Informative & Moderately Persuasive	48	3.28	0.83	3.44	1.00
		188	One-way ANOVA F(3, 185)=21.01, p-value=0.000		One-way ANOVA F(3, 185)=42.00, p-value=0.000	

Figure WD1: Means Plots Sample 1 (N=188)

Average Perceived Persuasiveness Average Perceived Informativeness

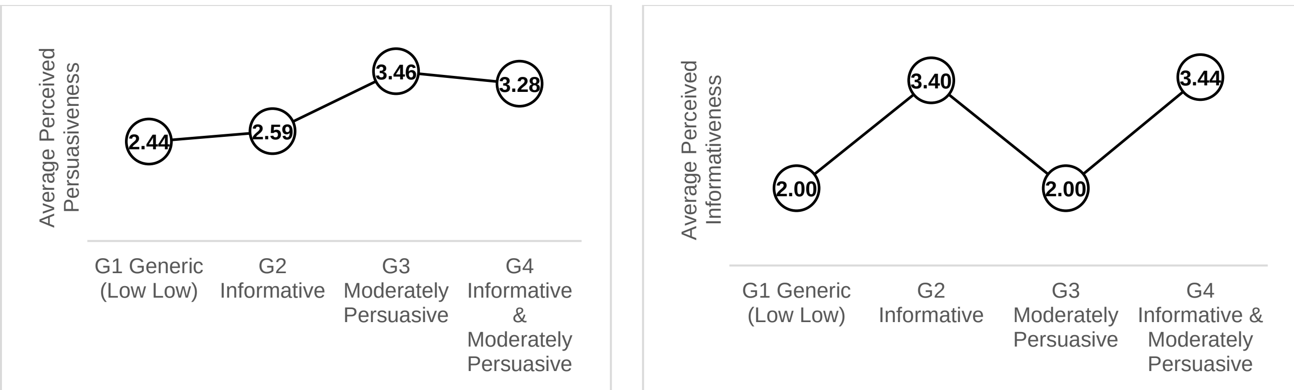


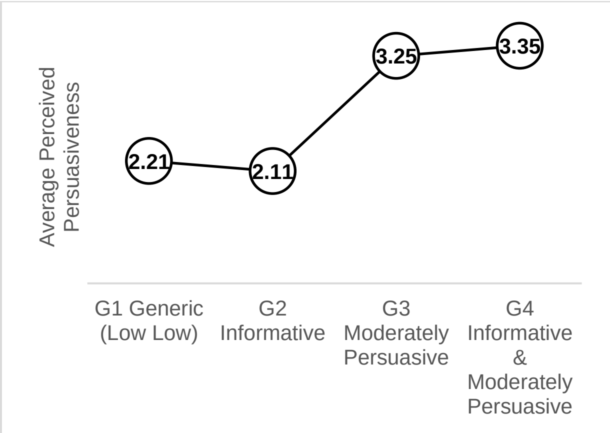
Table WD2: Sample 2 (N=191) Anova Analysis Results

Experimental Cells			**Perceived Persuasiveness**		**Perceived Informativeness and Transparency**	
Group	**Description**	**N**	**Mean**	**SD**	**Mean**	**SD**
G1	Generic (Low Low)	47	2.60	0.80	2.21	0.86
G2	Informative	49	2.51	0.77	2.11	0.86
G5	Highly Persuasive	46	3.45	0.77	3.25	1.12
G6	Informative & Highly Persuasive	49	3.46	0.70	3.35	0.92

Experimental Cells			Perceived Persuasiveness		Perceived Informativeness and Transparency	
Group	**Description**	**N**	**Mean**	**SD**	**Mean**	**SD**
		191	One-way ANOVA F(3, 188)=22.13, p-value=0.000		One-way ANOVA F(3, 188)=23.78, p-value=0.000	

Figure WD2: Means Plots Sample 2 (N=191)

Average Perceived Persuasiveness Average Perceived Informativeness



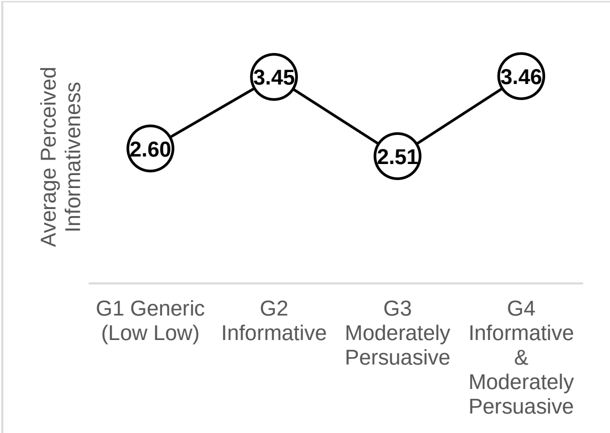


Table WD3: Items Used for the Manipulation Checks

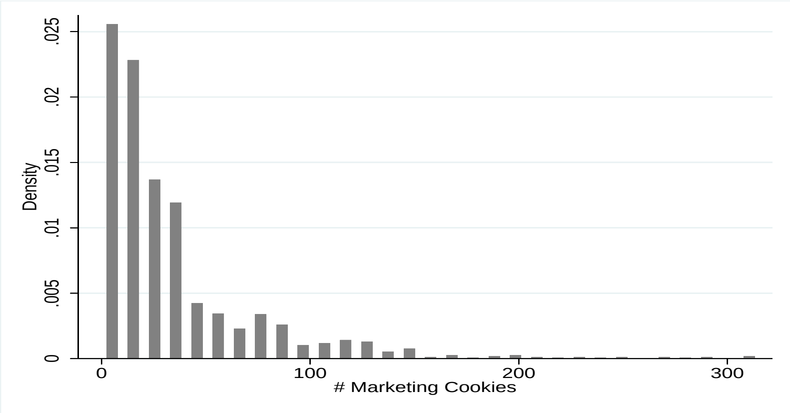
Perceived Informativeness and Transparency This company (Name of the company): • explains how the collected personal information will be used • informs whether personal information will be disclosed to a third party • protects my personal information • is transparent about the use of my personal information • gives me clear information about how my personal data are treated	Cronbach's Alpha Sample 1 Sample 2
Perceived Persuasiveness This company (Name of the 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	Cronbach's Alpha Sample 1 Sample 2

Note: A Likert scale is used where 1=strongly disagree and 5=strongly agree

Web Appendix E: Benefits and Costs of Using Persuasive Cues When Asking for Personal Data

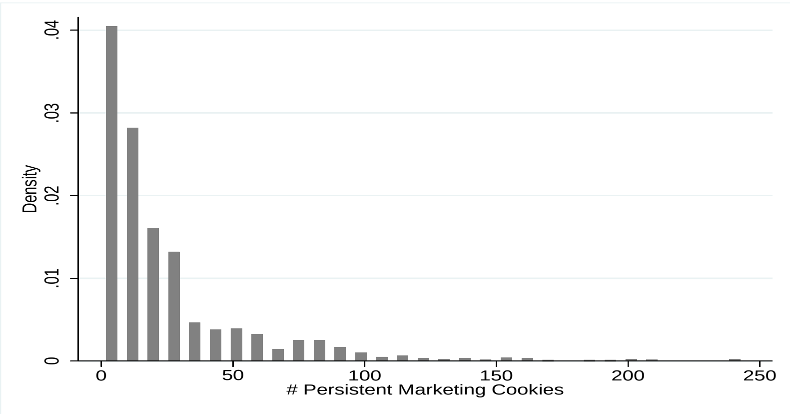
Figure WE1: Distribution Marketing Cookies

Panel A: # of Marketing Cookies of Firm's Website



Average	=	34.7
SD	=	42.1
Min	=	0
Median	=	21
Max	=	316

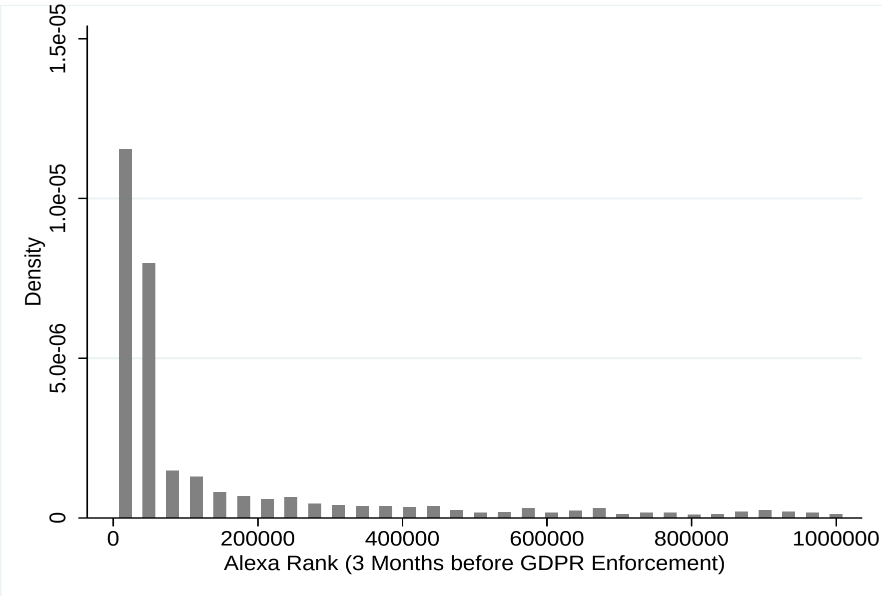
Panel B: # of Persistent Marketing Cookies of Firm's Website



Average	=	24.9
SD	=	32.1
Min	=	0
Median	=	14
Max	=	245

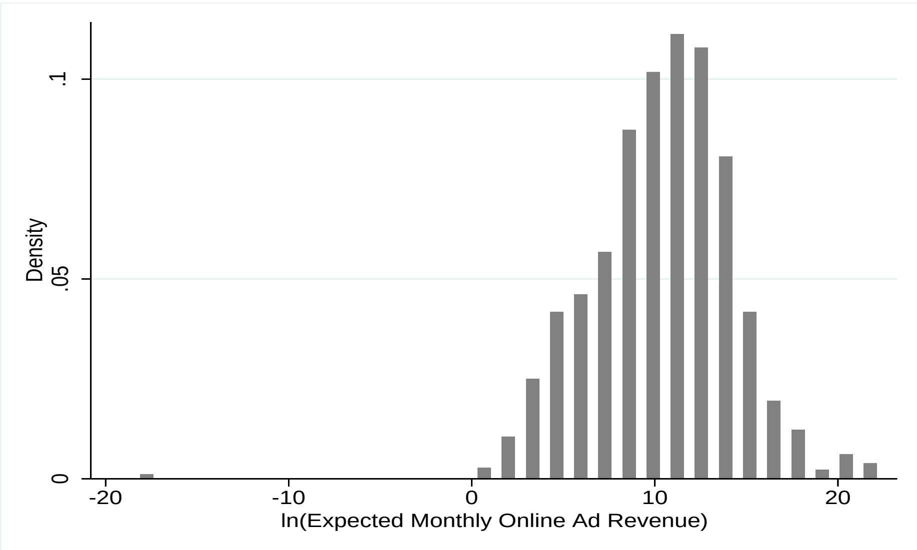
Notes: We used Cookiebot to scan up to 1,000 pages randomly selected from each of the 1,396 firm domains, resulting in an analysis of 980,182 pages from two distinct snapshots. These snapshots aimed to account for potential differences in sampling times and pages. Despite the variations, the two measurements for each firm were highly correlated (r = .92), indicating consistent extractions. However, Cookiebot noted that 35 of our sampled websites likely used dynamic cookies — cookies with unique names per user session. This could lead to overestimations in the cookie count. A sensitivity analysis, excluding these 35 sites, confirmed no significant impact on our main findings.

Figure WE2: Distribution and Summary Statistics for the Averaged Alexa Rank Scores for the Three Months Before the GDPR Enforcement



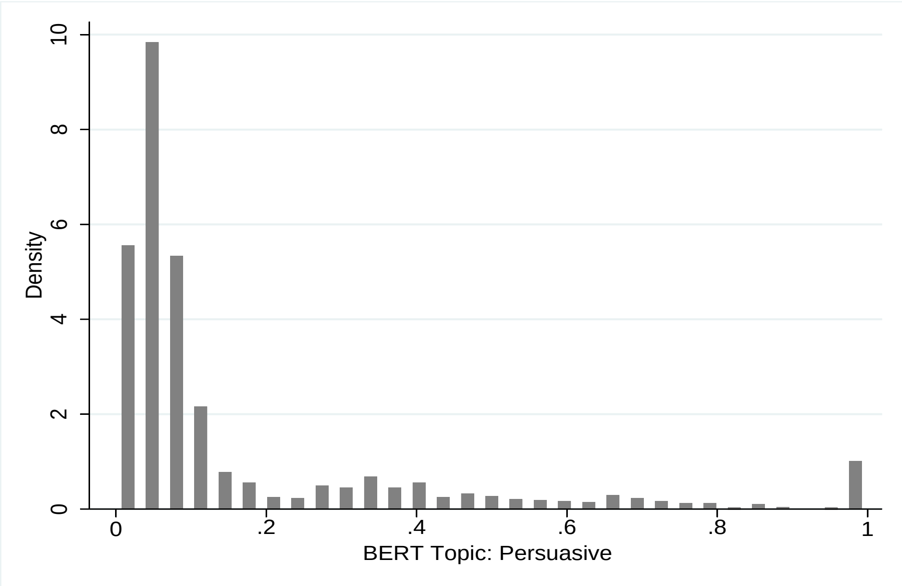
Average	=	146,505
SD	=	224,751
Min	=	1
Median	=	62,542
Max	=	1,016,361

Figure WE3: Distribution and Summary Statistics for the ln (Expected Online Ad Revenue)



Average	=	10.38
SD	=	3.88
Min	=	-18.4
Median	=	10.63
Max	=	22.43

Figure WE4: Distribution and Summary Statistics for the DV of Equation (3): *Persuasive$_{ij}$*



Average	=	0.163
SD	=	0.233
Min	=	0.000
Median	=	0.312
Max	=	1.000

DV of Equation (3): Persuasive$_{ij}$

Persuasive$_{ij}$ was obtained through the BERT analysis of the emails' text. It corresponds to Topic 1 in Figure WB2 of the Web Appendix B and can be interpreted as the probability that email *i* contains persuasive elements. It also represents the DV of equation (3) presented in the manuscript. We opt for a fractional logit model specification that is appropriate when the dependent variable is bounded between 0 and 1 and zeros and ones occur. A widespread alternative model used in literature to deal with proportions is Beta-regression (e.g., Buckley 2003; Ferrari and Cribari-Neto 2004; Hardin and Hilbe 2014). However, a limitation of this approach is that the dependent variable should be strictly greater than 0 and smaller than 1, meaning it cannot handle the extreme values of 0 and 1. In our sample, cases of 0 are not negligible, as Figure WE4 shows, and values that are almost 1 also occur. Therefore, we opted for a fractional logit model specification.

Web Appendix F: Controls Included in Equation 3

In this section, we describe the controls embedded in Equation 3, largely focused on firm characteristics. Following a brief definition of each variable, we furnish descriptive statistics tables associated with each control.

Company Size: We sourced the number of employees for each company in our sample primarily from Orbis.

Year of Foundation and Country: This was also extracted from Orbis. For missing data, we consulted supplementary sources like SimilarWeb, Crunchbase, and Owler.

Product Category: Each business was classified according to a specific product category, using Google's list of product categories as our reference protocol. To this end, we have introduced 20 dummy variables, each representing a distinct product category, with "Other" serving as the baseline. The categories include the following: 1) Apparel & Accessories, 2) Automotive, 3) Beauty and Wellness Products, 4) Books, Art & Game Items, 5) Computer, Software & Network Services, 6) Electronics, 7) Entertainment & Sports Services, 8) Finance & Insurance, 9) Food & Beverages, 10) Health & Beauty Services, 11) Home, Furniture & Garden, 12) Learning and Personal Development, 13) Magazines, Newspapers & Blogs, 14) Non-Profit & Charity, 15) Professional Services, 16) Retail, 17) Social Media, 18) Software Products, 19) Transportation Services, and 20) Utilities: Telecom & Electricity.

Our coding efforts have provided us the flexibility to delve into finer categorizations. Initially, we identified 34 distinct categories. However, to avoid incorporating categories with sparse observations in the model, we decided to consolidate some of them. As an additional layer of robustness, we also sourced sector classification data from the Bureau van Dijk (BdV), which covers 99% of firms in our sample. We also retrieved data about NAICS and SIC codes (covering a smaller percentage of firms 95% and 92% respectively). NAICS, SIC, and BdV do not always overlap. We performed a sensitivity analysis as a robustness check, confirming our key findings' consistency across different versions.

While both classifications yield consistent results, we ultimately preferred the product categories over the BdV system, given its more granular classification. This allowed us to explore separate models with more precision.

Firm Type: We implemented controls to differentiate firms based on their product/service nature: Digital products, Physical products, Services.

Distribution Strategy: Firms were categorized based on their distribution channel: Purely digital firms, Firms with physical stores.

Industry Concentration Index (Herfindahl): A measure to gauge competitiveness computed considering the four largest firm percentage. A lower concentration ratio signals heightened competition in the industry.

Number of Non-marketing Cookies: Thanks to our collaboration with Cookiebot, we can differentiate between marketing and non-marketing cookies. Therefore, we included the number of non-marketing cookies as control.

Discrepancy between Declared vs. Observed Cookies: This control contrasts cookies declared in firms' privacy policies against those detected by Cookiebot. To create this variable for each firm in our sample, we downloaded the full privacy policy document from its website. After carefully reviewing these documents, we counted how many cookies each company declared. Consequently, we created a variable representing the discrepancy between cookies that firms declared in their respective privacy statements and those detected by Cookiebot. A substantial positive difference might potentially highlight businesses that use more cookies than they declared.

Difference (# of Days) Re-permission Email Sending vs. May 28th, 2018: We incorporate this variable as a control, recognizing the potential influence of the interval between policy updates and email communications on user responses. To quantify this, we created a variable reflecting the number of days differing from the 28th of May 2018, which is a critical GDPR implementation date. For instance, a value of -10 indicates that the email was dispatched 10 days prior to this deadline, a value of 0 denotes the exact date of the deadline, and +10 signifies that an email was sent ten days post-deadline.

Advertiser Dummy: A firm's website may act as either a publisher, an advertiser, or both. We created a variable that equals 1 when the firm acts only as an advertiser.

Handling of Missing Data:

To maintain the integrity of our analysis amid missing data, we employed dummy variables.

For comprehensive statistics and details for each of these controls, please refer to the subsequent tables.

Table WF1: Descriptive Statistics on the Size of Firms Sending Re-permission Emails

Number of Employees	**Freq.**	**Percent**	**Cum**
1-10	336	25.87	25.87
11-50	276	21.25	47.11
51-200	62	4.77	76.6
201-500	102	7.85	71.82
501-1000	137	10.55	87.14
1001-5000	219	16.86	63.97
5001-10000	43	3.31	90.45
>10000	124	9.55	100
Total	1,299	100	

Notes: For 7% of the sample, this information is unavailable. Consequently, we introduced as control a dummy variable that equals 1 when information is unavailable.

Table WF2: Descriptive Statistics on the Age of the Firms Sending Re-permission Emails.

Variable	Obs	Mean	SD	Min	Max
Firms'Age (in Years)	1,306	22.37	26.01	0	237

Notes: For 6.1% of the sample, this information is unavailable. Consequently, we introduced a dummy variable, set to 1 when unavailable, as a control.

Table WF3: Descriptive on the Countries and Continents of Firms Sending Re-permission Emails

Country of the Headquarter	Freq.	Percent
United Kingdom	391	28.01
United States	351	25.14
Italy	156	11.17
Poland	65	4.66
Spain	46	3.30
Germany	39	2.79
France	36	2.58
Portugal	27	1.93
Ireland	23	1.65
Netherlands	20	1.43
Sweden	18	1.29
Belgium	14	1.00
Other	165	11.82
Missing	45	3.22
Total	1396	100.00

Continent of the Heardquarter	Freq.	Percent
Europe	913	65.40
North America	370	26.50
Asia	52	3.72
Oceania	14	1.00
Africa	1	0.07
South America	1	0.07
Missing	45	3.22
Total	1,396	100.00

Table WF4: Descriptive Statistics on the Product Category of Firms Sending Re-permission Emails

Service			Products		
Category	**Freq.**	**Percent**	**Category**	**Freq.**	**Percent**
Compute, Soft. & Network Services	168	18.90	Software Products	86	16.96
Retail	138	15.52	Electronics	85	16.77
Entertainment & Sport Services	136	15.30	Apparel & Accessories	74	14.60
Professional Services	96	10.80	Beauty and Wellness Products	55	10.85
Learning and Development	79	8.89	Food, Beverages & Tobacco	55	10.85
Finance & Insurance	76	8.55	Book, Art & Game Items	43	8.48

Service			Products		
Category	**Freq.**	**Percent**	**Category**	**Freq.**	**Percent**
Transportation Services	49	5.51	Magazines, Newspapers & Blogs	38	7.50
Health & Beauty Services	27	3.04	Home, Furniture & Garden	34	6.71
Non-Profit, Charity	26	2.92	Automotive	13	2.56
Social Media	12	1.35	Other	24	4.73
Utilities: Telecom, Electricity	11	1.24			
Other	71	7.99			
Total	**889**	**100.00**	**Total**	**507**	**100.0**

Table WF5: Descriptive Statistics on the Firm Type of Firms Sending Re-permission Emails

Firm Type	Freq.	Percent
Service	889	63.68
Digital Product	230	16.48
Physical Product	277	19.84
Total	**1396**	**100.00**

Table WF6: Descriptive Statistics on the Distribution Strategy of Firms Sending Re-permission Emails

Distribution Strategy	Freq.	Percent
Digital-only sales channel	928	66.48
Digital & Physical sales channels	72	5.16
Physical store-only sales channels	396	28.37
Total	1396	100.00

Table WF7: Descriptive Statistics on the Industry Concentration Index (Herfindahl) of the Firms Sending Re-permission Emails

Variable	Obs	Mean	SD	Min	Max
Concentration Index (Herfindahl)	1,371	0.09	0.06	0.01	0.38

Notes: For 1.73% of the sample, this information is unavailable. Consequently, we introduced a dummy variable that equals 1 when information is unavailable.

Table WF8: Descriptive Statistics on the Number of Non-Marketing Cookies Used by Firms Sending Re-permission Emails in Their Websites

Variable	Obs.	Mean	SD	Min	Max
Non Marketing Cookies	1,396	130.6	2209.1	0.0	82297.5

Table WF9: Descriptive Statistics on the Discrepancy between Number of Declared and Observed Cookies

Variable	Obs.	Mean	SD	Min	Max
Discrepancy between Number of Declared and Observed Cookies	355	59.3	228.9	-1051	2964

Notes: In our sample, 74% of firms did not specify the cookies they utilized. Due to this lack of clarity, we added a dummy variable as a control, that equals 1 when such information was absent. Owing to the significant portion of firms without available cookie declarations, we conducted a robustness check both with and without these variables. The results remained consistent across both scenarios.

Table WF10: Descriptive Statistics on the Time Discrepancy between GDPR Policy Date of Implementation and Email Timeline

Variable	Obs.	Mean	SD	Min	Max
Time Discrepancy between GDPR date and Email Timeline	1,120	14.9884	81.2023	-252	437

Notes: For 25% of the sample, this information is unavailable. Consequently, we introduced a dummy variable, that equals 1 when information is unavailable.

Table WF11: Descriptive Statistics on the Advertiser Dummy Variable

Type	Freq.	Percent
Advertiser Only=0	102	7.31
Advertiser Only=1	880	63.04
Missing Information	414	29.66
Total	1,396	100.00

Appendix G: Fractional Logit Regression (Equation 3)

Table WG1: Fractional Logit Models' Results – DV = Data-Driven BERT Persuasive Topics (Equation 3): Complete Results

Panel A: Full Model

	(1)		(2)		(3)		(4)	
Variables	Coef.	z	Coef.	z	Coef.	z	Coef.	z
ln(# Marketing Cookies)					0.017	1.920	0.020	2.140
ln(Expected Monthly AdRevenues$_i$) [b]			0.039	2.320			0.040	2.380
Website Popularity (3 months pre-GDPR)	0.000	0.750	0.000	1.520	0.000	0.680	0.000	1.480
# Data Breaches (pre-GDPR)	0.170	0.970	0.080	0.440	0.175	1.010	0.083	0.460
Website Popularity * # Data Breaches (pre-GDPR)	0.000	-3.710	0.000	-3.410	0.000	-3.720	0.000	-3.420
Control Variables								
Re-permission email's content: Informative [a]	-2.957	-13.560	-2.975	-13.720	-2.968	-13.600	-2.986	-13.790
Europe (0/1)	-0.105	-0.980	-0.090	-0.840	-0.105	-0.980	-0.088	-0.820
Firm Size [b]	-0.030	-1.220	-0.045	-1.720	-0.030	-1.210	-0.043	-1.650
Tenure (Firm Age) [b]	0.000	0.130	0.001	0.350	0.000	0.100	0.001	0.400
Advertiser (0/1)=1 [b]	0.136	0.850	0.191	1.180	0.118	0.750	0.173	1.070
Number of Non-Marketing Cookies	-0.002	-2.570	-0.002	-2.500	-0.001	-2.170	-0.001	-2.040
Discrepancy Declared Cookies vs. Observed Cookies	0.002	2.560	0.002	2.490	0.001	2.150	0.001	2.030
Difference (# of Days) Re-per. Email Sending vs. May 28th 2018	-0.001	-1.750	-0.001	-1.730	-0.001	-1.780	-0.001	-1.780
Physical Product (0/1)	0.451	2.160	0.432	2.080	0.429	2.040	0.394	1.960
Service (0/1)	-0.688	-1.930	-0.669	-1.850	-0.782	-2.110	-0.793	-2.110
Offline Presence (0/1)	0.011	0.100	0.029	0.270	0.018	0.170	0.036	0.340
Concentration Index (Four Largest Firm %)	-0.184	-0.220	-0.354	-0.420	-0.182	-0.210	-0.346	-0.410
Automotive (0/1)	-1.072	-2.090	-1.045	-2.000	-1.197	-2.320	-1.198	-2.290
Apparel & Accessories (0/1)	-0.717	-1.940	-0.667	-1.770	-0.829	-2.230	-0.797	-2.130
Beauty and Wellness Products (0/1)	-1.091	-2.740	-1.033	-2.560	-1.216	-3.030	-1.156	-2.870
Book, Art & Game Items (0/1)	-0.605	-1.570	-0.556	-1.430	-0.729	-1.880	-0.706	-1.820
Electronics (0/1)	-0.475	-1.360	-0.448	-1.260	-0.622	-1.740	-0.621	-1.720
Food & Beverages (0/1)	-0.555	-1.480	-0.513	-1.350	-0.665	-1.760	-0.636	-1.680
Home, Furniture & Garden (0/1)	-0.326	-0.770	-0.252	-0.590	-0.473	-1.090	-0.407	-0.920

	(1)		(2)		(3)		(4)	
Variables	Coef.	z	Coef.	z	Coef.	z	Coef.	z
Magazines, Newspapers & Blogs (0/1)	-0.294	-0.710	-0.280	-0.670	-0.425	-1.020	-0.559	-1.350
Software Products (0/1)	-0.498	-1.380	-0.451	-1.230	-0.654	-1.750	-0.641	-1.710
Retail (0/1)	0.391	1.980	0.409	2.090	0.344	1.760	0.357	1.850
Non-Profit, Charity (0/1)	0.486	1.390	0.501	1.460	0.418	1.200	0.428	1.250
Professional Services (0/1)	0.272	1.200	0.293	1.290	0.232	1.030	0.251	1.120
Transportation Services (0/1)	-0.011	-0.040	0.012	0.040	-0.063	-0.220	-0.048	-0.170
Social Media (0/1)	0.228	0.590	0.143	0.360	0.248	0.630	0.163	0.410
Utilities: Telecom, Electicity (0/1)	-0.799	-2.210	-0.783	-2.210	-0.868	-2.390	-0.867	-2.440
Compute, Software & Network Services (0/1)	0.149	0.730	0.139	0.700	0.116	0.580	0.103	0.530
Entertainment & Sport Services (0/1)	0.827	4.390	0.840	4.460	0.758	4.070	0.765	4.110
Finance & Insurance (0/1)	-0.467	-1.870	-0.443	-1.790	-0.531	-2.140	-0.517	-2.090
Health & Beauty Services (0/1)	0.061	0.190	0.097	0.300	0.028	0.090	0.062	0.200
Constant	-0.106	-0.250	-0.593	-1.260	-0.002	0.000	-0.490	-1.020
DV=Persuasive$_{ij}$[a]		N=1506 LL= -499.29 BIC= 1313.21		N=1506 LL= -497.66 BIC= 1324.59		N=1506 LL= -498.72 BIC= 1319.40		N=1506 LL=-496.91 BIC= 1330.407

Panel B: Results Distinct by Firm Type

	Service		Product		Digital Product		Physical Product	
Variables	Coef	z	Coef	z	Coef	z	Coef	z
ln(# Marketing Cookies)	0.024	1.980	0.013	0.830	-0.026	-1.370	0.059	2.580
ln(Expected Monthly AdRevenues$_j$) [b]	0.034	1.670	0.055	1.860	0.003	0.080	0.067	1.660
Website Popularity (3 months pre-GDPR)	0.000	1.640	0.000	0.420	0.000	-0.270	0.000	0.330
# Data Breaches (pre-GDPR)	0.086	0.350	-0.154	-0.520	-0.570	-1.460	1.039	1.650
Website Popularity * # Data Breaches (pre-GDPR)	-0.000	0.260	-0.000	-2.490	-0.000	-1.690	-0.000	-3.110
Control Variables								
Re-permission email's content: Informative [a]	-3.015	-10.530	-2.962	-9.040	-2.850	-5.360	-3.110	-7.230
Europe (0/1)	0.002	0.010	-0.198	-1.200	-0.363	-1.460	0.116	0.480
Firm Size [b]	-0.026	-0.800	-0.094	-2.130	-0.128	-1.590	-0.101	-1.760
Tenure (Firm Age) [b]	-0.001	-0.410	0.003	1.040	-0.009	-0.890	0.003	0.770
Advertiser (0/1)=1 [b]	0.086	0.430	0.278	0.990	0.180	0.480	0.065	0.140
Number of Non-Marketing Cookies	0.000	0.190	-0.002	-2.130	0.000	0.280	-0.002	-1.810

	Service		Product		Digital Product		Physical Product	
Variables	Coef	z	Coef	z	Coef	z	Coef	z
Discrepancy Declared Cookies vs. Observed Cookies	0.000	0.050	0.002	2.120	0.000	0.070	0.002	1.790
Difference (# of Days) Re-per. Email Sending vs. May 28th 2018	-0.001	-1.380	-0.002	-1.260	-0.005	-1.400	-0.001	-0.980
Physical Product (0/1)	na	na	0.515	2.430	na	na	na	na
Service (0/1)	na	na	na	na	na	na	na	na
Offline Presence (0/1)	0.127	0.930	-0.103	-0.610	-0.182	-0.540	-0.240	-1.200
Concentration Index (Four Largest Firm %)	0.379	0.330	-1.402	-1.130	1.165	0.660	-1.560	-0.740
Automotive (0/1)	na	na	-1.022	-1.830	na	na	-0.313	-0.460
Apparel & Accessories (0/1)	na	na	-0.656	-1.710	na	na	0.021	0.040
Beauty and Wellness Products (0/1)	na	na	-0.997	-2.400	na	na	-0.344	-0.640
Book, Art & Game Items (0/1)	na	na	-0.591	-1.470	-0.744	-1.450	-0.008	-0.010
Electronics (0/1)	na	na	-0.481	-1.270	-0.579	-1.180	0.143	0.250
Food & Beverages (0/1)	na	na	-0.471	-1.190	na	na	0.256	0.500
Home, Furniture & Garden (0/1)	na	na	-0.119	-0.260	0.014	0.020	0.592	0.830
Magazines, Newspapers & Blogs (0/1)	na	na	-0.446	-1.04	-0.466	-0.87	0.545	0.820
Software Products (0/1)	na	na	-0.476	-1.200	-0.399	-0.830	0.159	0.230
Retail (0/1)	0.342	1.780	na	na	na	na	na	na
Non-Profit, Charity (0/1)	0.406	1.170	na	na	na	na	na	na
Professional Services (0/1)	0.209	0.940	na	na	na	na	na	na
Transportation Services (0/1)	-0.168	-0.570	na	na	na	na	na	na
Social Media (0/1)	0.129	0.310	na	na	na	na	na	na
Utilities: Telecom, Electicity (0/1)	-1.222	-2.640	na	na	na	na	na	na
Compute, Software & Network Services (0/1)	0.085	0.430	na	na	na	na	na	na
Entertainment & Sport Services (0/1)	0.693	3.720	na	na	na	na	na	na
Finance & Insurance (0/1)	-0.560	-2.260	na	na	na	na	na	na
Health & Beauty Services (0/1)	0.056	0.180	na	na	na	na	na	na
Constant	-1.315	-3.170	-0.529	-0.820	0.404	0.520	-0.967	-1.020
DV=$Persuasive_{ij}$[a]		N=968 LL= -310.86 BIC= 862.36		N=538 LL= -182.833 BIC= 585.741		N= 237 LL= -71.74 BIC= 280.19		N=301 LL= -105.74 BIC= 354.16

[a] *Persuasive*$_{ij}$ and *Informative*$_{ij}$ used in this regression were identified through the BERT analysis in Study 1.

[b] We included dummy variables to control for missing values (e.g., missing firm age, missing advertiser).

Table WG2: Logit Models' Results – DV = Theory-based Persuasive Themes

	DV: Incentives[A]		DV: Framing[A]		DV: Overall Persuasive[A]	
Variables	Coef.	z	Coef.	z	Coef.	z
ln(# Marketing Cookies)	**0.069**	**2.120**	0.005	0.440	**0.022**	**1.720**
ln(Expected Monthly AdRevenues$_j$) [B]	**0.084**	**2.180**	**0.058**	**2.620**	**0.040**	**1.980**
Website Popularity (3 months pre-GDPR)	-0.000	-0.220	0.000	0.200	0.000	0.320
# Data Breaches (pre-GDPR)	-0.072	-0.140	-0.100	-0.440	-0.050	-0.240
Website Popularity * # Data Breaches (pre-GDPR)	**-0.000**	**-1.860**	-0.000	-0.110	**-0.000**	**-1.740**
Control Variables						
Re-permission email's content: Informative [a]	-0.294	-3.770	-0.218	-5.530	-0.274	-7.300
Europe (0/1)	-0.374	-1.400	-0.102	-0.740	-0.094	-0.720
Firm Size [B]	0.028	0.450	-0.092	-2.880	-0.068	-2.260
Tenure (Firm Age) [B]	-0.001	-0.240	0.000	0.190	0.000	-0.080
Advertiser (0/1)=1 [B]	0.779	1.410	-0.125	-0.560	0.044	0.210
Number of Non-Marketing Cookies	-0.001	-0.600	-0.002	-1.980	-0.002	-2.070
Discrepancy Declared Cookies vs. Observed Cookies	0.001	0.620	0.002	1.980	0.002	2.080
Difference (# of Days) Re-per. Email Sending vs. May 28th 2018	-0.001	-0.510	0.000	0.260	0.000	-0.030
Physical Product (0/1)	0.393	0.700	0.324	1.070	0.439	1.530
Service (0/1)	0.438	0.360	0.192	0.410	0.349	0.770
Offline Presence (0/1)	0.051	0.210	0.043	0.310	0.035	0.270
Concentration Index (Four Largest Firm %)	-1.372	-0.660	-1.491	-1.400	-1.405	-1.390
Automotive (0/1)	0.829	0.620	1.309	1.620	1.246	1.870
Apparel & Accessories (0/1)	0.644	0.550	-0.032	-0.060	-0.043	-0.090
Beauty and Wellness Products (0/1)	-0.305	-0.250	0.002	0.000	-0.083	-0.160
Book, Art & Game Items (0/1)	0.708	0.590	-0.013	-0.020	0.063	0.120
Electronics (0/1)	0.858	0.730	-0.006	-0.010	0.168	0.360
Food & Beverages (0/1)	0.288	0.240	-0.015	-0.030	-0.052	-0.100
Home, Furniture & Garden (0/1)	-1.372	-0.880	-0.147	-0.250	-0.187	-0.340
Magazines, Newspapers & Blogs (0/1)	0.626	0.470	-0.519	-0.910	-0.404	-0.730
Software Products (0/1)	-0.420	-0.340	0.159	0.320	0.030	0.060
Retail (0/1)	0.310	0.690	0.162	0.660	0.131	0.560
Non-Profit, Charity (0/1)	-0.819	-0.920	-0.254	-0.570	-0.224	-0.520

	DV: Incentives[A]		DV: Framing[A]		DV: Overall Persuasive[A]	
Variables	Coef.	z	Coef.	z	Coef.	z
Professional Services (0/1)	-0.947	-1.590	-0.074	-0.280	-0.264	-1.030
Transportation Services (0/1)	-1.653	-1.480	0.522	1.510	0.147	0.460
Social Media (0/1)	0.655	0.520	0.084	0.160	0.016	0.030
Utilities: Telecom, Electricity (0/1)	na	na	-0.092	-0.150	-0.432	-0.720
Compute, Software & Network Services (0/1)	-0.888	-1.600	-0.061	-0.260	-0.258	-1.150
Entertainment & Sport Services (0/1)	0.492	1.070	0.210	0.840	0.441	1.870
Finance & Insurance (0/1)	-0.411	-0.650	0.136	0.460	-0.055	-0.200
Health & Beauty Services (0/1)	-0.657	-0.740	1.059	2.270	0.678	1.760
Constant	-2.305	-1.620	0.471	0.770	na	na
/cut1					-0.59	
/cut2					1.33	
/cut3					2.55	
/cut4					4.23	
/cut5					7.03	
		N=1506		N=1506		N=1506
		LL= -437.94		LL= -981.7		LL= -1679.81
		BIC= 992.08		BIC=2292.78		BIC= 3513.44

[A] The specific topics in the re-permission emails (incentive, framing, overall persuasive) utilized in these regressions are derived from the unsupervised, theory-based method outlined in the "Analysis of GDPR Re-Permission Emails" section. Both $Incentive_{ij}$ and $Framing_{ij}$ denote the presence of the particular theme, represented as dummy variables where 1 signifies the theme's presence in email I; these variables were estimated using logit models. Meanwhile, $OverallPersuasive_{ij}$ is an ordinal variable tallying the various persuasive themes in email *i*, estimated using an ordered logit. A value of 0 indicates no themes were used, 1 signifies one persuasive theme (e.g., monetary incentives), 2 implies two persuasive themes (e.g., monetary incentives and loss framing), and so on.
[B] We included dummy variables to control for missing values (e.g., missing firm age, missing advertiser).

Table WG3: Fractional Logit Models' Results – DV = Data-Driven BERT Persuasive Topics (Equation 3): Using Sectors over Product Categories

Macro Category	Variables	(1) Coef	(1) z	(2) Coef	(2) z	(3) Coef	(3) z	(4) Coef	(4) z
Re-permission email's content	*Informative*$_{ij}$[a]	-2.87	-13.08	-2.92	-13.26	-2.88	-13.17	-2.93	-13.34
Benefits	ln(# Marketing Cookies)					0.02	2.35	0.02	2.27
	ln(Expected Monthly *AdRevenues*$_j$) [b]			0.05	2.59			0.04	2.51
Costs	Website Popularity (3 months pre-GDPR)	0.00	1.44	0.00	2.18	0.00	1.36	0.00	2.07
	# Data Breaches (pre-GDPR)	-0.14	-0.62	-0.18	-0.81	-0.18	-0.82	-0.22	-0.99
	Website Popularity * # Data Breaches (pre-GDPR)	0.00	-3.09	0.00	-2.80	0.00	-3.07	0.00	-2.79
Controls	Europe	-0.09	-0.78	-0.05	-0.41	-0.09	-0.76	-0.05	-0.41
	Firm Size [b]	-0.03	-1.41	-0.06	-2.25	-0.04	-1.45	-0.06	-2.25
	Tenure (Firm Age) [b]	0.00	1.13	0.00	1.37	0.00	1.16	0.00	1.39
	Advertiser (0/1)=1 [b]	0.01	0.04	0.07	0.42	0.01	0.08	0.07	0.44
	Sectors [c]								
	Media and Entertainment	-0.28	-1.43	-0.31	-1.60	-0.25	-1.29	-0.28	-1.46
	Professional Services	0.10	0.75	0.10	0.75	0.10	0.74	0.10	0.74
	Retail Trade	0.26	1.83	0.26	1.85	0.26	1.81	0.26	1.82
	Software and IT Services	-0.02	-0.13	-0.05	-0.27	-0.01	-0.04	-0.03	-0.17
	Travel Tourism and Hospitality	0.30	1.88	0.31	1.95	0.32	1.98	0.33	2.05
	Social Media	0.65	1.79	0.43	1.11	0.82	2.21	0.60	1.54
	Concentration Index (Four Largest Firm %)	0.23	0.24	0.16	0.17	0.14	0.15	0.07	0.08
	Constant	-0.81	-3.01	-1.32	-4.01	-0.87	-3.24	-1.35	-4.12
DV=*Persuasive*$_{ij}$[a]		N=1506 LL= -498.80 BIC= -10453.15		N=1506 LL= -497.59 BIC= -10448.25		N=1506 LL= -497.62 BIC= -10448.21		N=1506 LL=-496.49 BIC= -10443.13	

[a] *Persuasive*$_{ij}$ and *Informative*$_{ij}$ used in this regression were identified through the BERT analysis in Study 1.

[b] We included dummy variables to control for missing values (e.g., missing firm age, missing advertiser).

[c] We identified the sectors using BdV code, which covers 99% of firms in our sample. However, we also retrieved data about NAICS and SIC codes (covering a smaller percentage of firms 95% and 92% respectively). NAICS, SIC, and BdV do not always overlap. Therefore, we performed a sensitivity analysis as a robustness check, confirming our key findings' consistency across different versions.

Table WG4: Fractional Logit Models' Results – DV = Data-Driven BERT Persuasive Topics (Equation 3): Results Distinct by Distribution Strategy

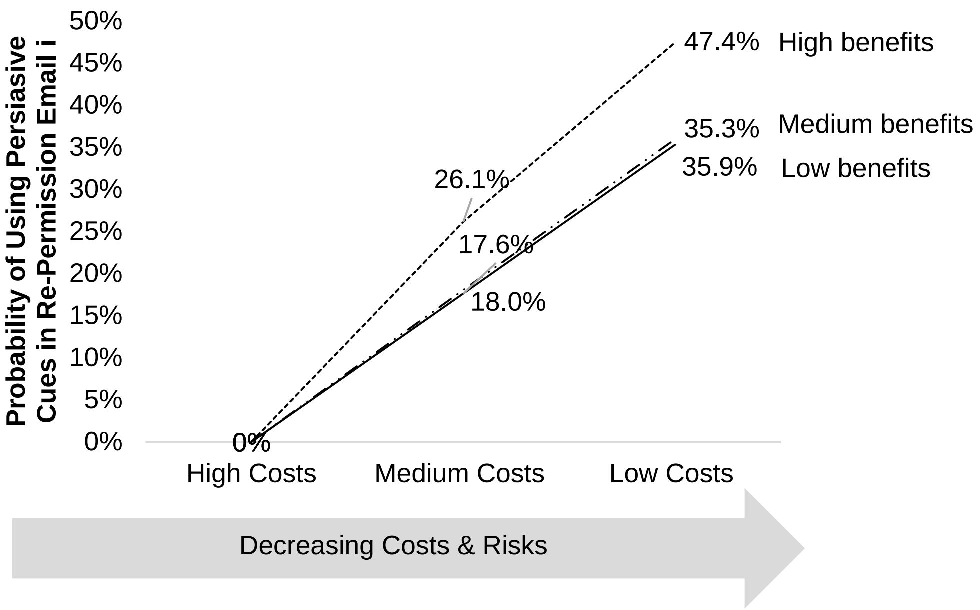
	Digital-only sales channel		**Physical sales channels=Yes**	
Variables	Coef.	z	Coef.	z
ln(# Marketing Cookies)	0.016	1.600	0.019	1.060
ln(Expected Monthly AdRevenues$_j$) [b]	0.011	0.570	**0.104**	**3.320**
Website Popularity (3 months pre-GDPR)	**0.000**	**2.070**	0.000	-0.060
Number of Data Breaches (pre-GDPR)	-0.019	-0.080	0.384	1.320
Website Popularity X Number of Data Breaches (pre-GDPR)	**-0.000**	**-2.230**	**-0.000**	**-3.500**
Control Variables				
Informative [a]	-2.849	-10.690	-3.294	-8.880
Europe (0/1)	-0.125	-0.960	0.089	0.460
Firm Size [b]	-0.023	-0.690	-0.084	-2.020
Tenure (Firm Age) [b]	-0.006	-1.820	0.005	1.840
Advertiser (0/1)=1 [b]	0.315	1.600	-0.229	-0.680
Number of Non-Marketing Cookies	-0.001	-0.990	-0.001	-1.330
Discrepancy Declared Cookies vs. Observed Cookies	0.001	1.180	0.001	1.320
Difference (# of Days) Re-per. Email Sending vs. May 28th 2018	-0.002	-1.800	-0.001	-0.570
Physical Product (0/1)	0.248	0.820	0.193	0.460
Service (0/1)	-0.993	-2.590	-0.101	-0.110
Offline Presence (0/1)	na	na	na	na
Concentration Index (Four Largest Firm %)	1.042	0.990	-3.355	-2.320
Automotive (0/1)	-0.420	-1.070	-0.822	-0.920
Apparel & Accessories (0/1)	-0.493	-1.140	-0.446	-0.550
Beauty and Wellness Products (0/1)	-0.526	-1.140	-1.136	-1.390
Book, Art & Game Items (0/1)	-0.826	-2.050	0.157	0.180
Electronics (0/1)	-0.782	-2.010	-0.080	-0.090
Food & Beverages (0/1)	-0.843	-1.760	0.116	0.150
Home, Furniture & Garden (0/1)	-0.564	-0.920	0.265	0.300
Magazines, Newspapers & Blogs (0/1)	-0.595	-1.210	-0.078	-0.090
Software Products (0/1)	-0.729	-1.930	-0.245	-0.250
Retail (0/1)	0.419	1.830	0.297	0.890
Non-Profit, Charity (0/1)	1.320	2.550	-0.219	-0.490
Professional Services (0/1)	0.141	0.510	0.521	1.420
Transportation Services (0/1)	0.170	0.750	-0.344	-0.940
Social Media (0/1)	0.480	1.180	na	na
Utilities: Telecom, Electicity (0/1)	-0.468	-0.850	-1.211	-2.020
Compute, Software & Network Services (0/1)	0.226	1.060	-1.827	-5.880
Entertainment & Sport Services (0/1)	0.885	3.740	0.634	2.120
Finance & Insurance (0/1)	-0.354	-1.230	-1.059	-2.710
Health & Beauty Services (0/1)	0.319	0.990	-1.002	-1.650
Constant	-0.315	-0.610	-0.951	-0.870
DV=Persuasive$_{ij}$ [a]		N=997 LL= -308.39 BIC= 920.60		N=509 LL= -177.46 BIC= 629.15

[a] *Persuasive$_{ij}$* and *Informative$_{ij}$* used in this regression were identified through the BERT analysis in Study 1.

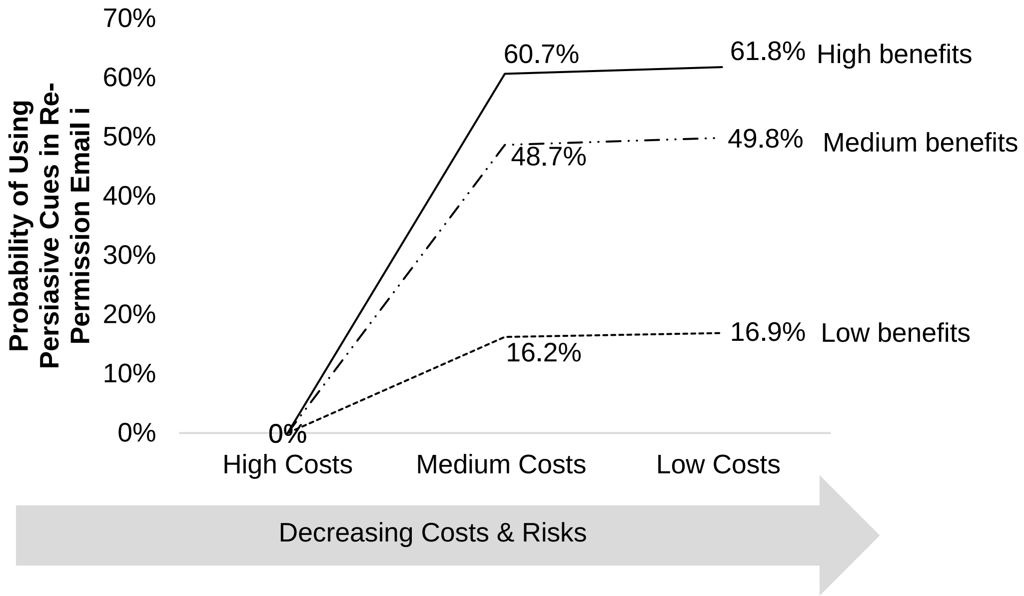
[b] We included dummy variables to control for missing values (e.g., missing firm age, missing advertiser).

Figure WG1: Scenario Analysis: Variation in Costs and Benefits of Using Persuasion in Opt-in Requests

Panel A: Firms Selling Digital Products (Software), having Only Digital Stores



Panel B: Firms Selling Physical Products (Furniture) and Having Physical Stores



Notes: For benefits, we used the variables *number of cookies* and *ad revenues*. For costs& risks, we used the *number of data breaches* and *website popularity*. To simulate the level for each variable, we took the value of the variables at the 5th, 50th, and 95th percentiles, respectively.

References

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

Buckley, Jack (2003), "Estimation of Models with Beta-Distributed Dependent Variables: A Replication and Extension of Paolino's Study," *Political Analysis*, 11 (2), 204–205.

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

Dogan, Alican, and Derya Birant (2022), "K-centroid link: A Novel Hierarchical Clustering Linkage Method," *Applied Intelligence*, 1-24.

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

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

Grootendorst, Maarten (2020), "BERTopic: Leveraging BERT and c-TF-IDF to Create Easily Interpretable Topics," Zenodo.

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.

Hartmann, Jochen, and Oded Netzer (2023), "Natural Language Processing in Marketing," *Artificial Intelligence in Marketing*, Vol. 20, 191-215.

Kapadia, Shashank (2019), “Evaluate Topic Models: Latent Dirichlet Allocation (LDA),” *Towards Data Science,* August 19. URL: https://towardsdatascience.com/evaluate-topic-model-in-python-latent-dirichlet-allocation-lda-7d57484bb5d0

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

Mimno, David, Hanna M. Wallach, Edmund Talley, Miriam Leenders, and Andrew McCallum (2011), “Optimizing Semantic Coherence in Topic Models,” *EMNLP 2011: Proceedings of the Conference on Empirical Methods in Natural Language Processin*g, 262–72.

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.

Shankar, Venkatesh, and Sohil Parsana (2022), “An Overview and Empirical Comparison of Natural Language Processing (NLP) Models and an Introduction to and Empirical Application of Autoencoder Models in Marketing,” *Journal of the Academy of Marketing Science*, 50 (6), 1324-1350.

Sievert, Carson, and Kenneth Shirley (2014), “LDAvis: A Method for Visualizing and Interpreting Topics,” Proceedings of the Workshop on Interactive Language Learning, Visualization, and Interfaces, Jason Chuang, Spence Green, Marti Hearst, Jeffrey Heer, and Philipp Koehn (eds.), Association for Computational Linguistics, Stroudsburg, PA, 63–70.