What is Online Privacy Worth? Evidence from an Information Provision Experiment*

Preliminary and incomplete

Guy Aridor† Samuel Goldberg‡ Jialong Li§

March 26, 2026

Abstract

We quantify the consumer welfare effects of privacy regulation via a two-month field experiment with nearly 1,600 participants who browse the web with a custom extension that tracks their behavior and experimentally varies the search costs and complexity of learning about website privacy practices. Our experimental variation identifies the parameters of a structural model of website choice and information acquisition that combines revealed browsing behavior with conjoint-elicited stated preferences. We find that comprehension, not access, is the binding friction: making policies easier to find dramatically increases search but produces no learning, while providing simplified, personalized privacy summaries meaningfully improves beliefs. Stated preferences overstate revealed preferences by approximately 34%, providing the first quantification of the privacy paradox. Using our estimates to evaluate two key provisions of privacy regulation: information disclosure and direct practice restrictions, we find that both generate welfare gains but both are inherently limited because consumers' valuations of website quality substantially exceed their privacy costs – implying that consumer-side privacy regulation generates modest benefits relative to the firm costs documented in prior work.

*We thank seminar/conference audiences at Boston University Questrom, Digital Economics Paris Digitization Seminar, George Mason Mini-symposium on Privacy, Marketing Science, Northwestern Kellogg, Prosocial Design Web Extension Research Group, Stanford GSB, UC Berkeley Haas, and Wharton for helpful comments and suggestions. Funding for this project was provided by the GMU Law & Economics Center Program on Economics & Privacy, Marketing Science Institute, and Stanford BGS. We thank Matt Rheault, Shan Sikdar, and Stanford DARC with assistance programming the Chrome Extension as well as Tony Kukavica for excellent research assistance. The research in this article was approved by the Institutional Review Board at Northwestern University (STU00219374). The experimental design and analysis were pre-registered in the AEA registry as AEARCTR-0016037. All errors are our own.

†Kellogg School of Management, Northwestern University. Email: guy.aridor@kellogg.northwestern.edu.

‡Stanford Graduate School of Business. Email: sggold@stanford.edu.

§Stanford Graduate School of Business. Email: markli@stanford.edu.

1 Introduction

The modern digital economy relies on the collection of consumer data for personalizing online experiences (Goldfarb and Tucker, 2019; Farronato, 2025). As data use has expanded, however, consumers have become increasingly concerned about online privacy (Goldfarb and Tucker, 2012), motivating large-scale protections such as the GDPR in the European Union and Apple's App Tracking Transparency (ATT). While existing empirical work suggests these policies may substantially harm firms by limiting data collection and use (Goldberg et al., 2024; Aridor et al., 2023, 2025), we know little about how much consumers value privacy and the benefits these regulations provide. The central challenge is that policy is often driven by stated preferences for privacy, yet a persistent gap between stated and revealed preferences, known as the "privacy paradox" (Acquisti et al., 2016; Athey et al., 2025), together with limited and noisy signals of privacy preferences in observed behavior, makes it difficult to recover consumer privacy valuations at policy-relevant scale.

This paper characterizes the welfare effects of consumer privacy regulation online using a field experiment and a structural model of browsing behavior. We argue that consumers are imperfectly informed about privacy practices due to two key informational frictions: a search cost of accessing privacy policies and a comprehension cost of understanding them. Consumers' incentives to acquire information about a given site depend on the long-run value of learning its privacy practices and the associated reduction in risk from resolving this uncertainty. When choosing between websites, consumers trade off website quality, which may depend on their data sharing choices, against the level of privacy provided, with these informational frictions shaping their expectations over privacy practices.

Our experiment involves nearly 1,600 participants who install a custom developed Chrome extension for two months; the extension collects browsing behavior and experimentally manipulates access to privacy information, and third-party tracking, to quantify the relative importance of each of these forces.[1] In our primary intervention, we vary both the ease of access and the complexity of privacy information. For participants in the *saliency* group, we reduce search costs by making the privacy policy more easily accessible as the participant browses. For participants in the *information* group, we both reduce search costs *and* the complexity of privacy information by providing users with binary privacy attributes; which we collect by hand for the top 500 websites online.[2] Finally, our experiment ends with a cookie-deletion intervention in which we exogenously remove third-

[1] A key concern is that our measurement itself is privacy-invasive. In Section 4.5 we discuss how we directly measure and mitigate selection into the recruiting platforms we use (Prolific, CloudResearch Connect), selection into the baseline survey, and selection into the extension portion of the study.

[2] Each policy is read by three trained raters overseen by law students (see Section 4.1.1 for more details). We show that there is considerable variation in privacy practices across websites within the same category, providing scope for information provision to affect choices.

party cookies associated with advertising and analytics, providing variation in privacy-dependent site quality.

Our reduced form results suggest significant information frictions in privacy choices, and provide sufficient variation to identify our structural model that captures these different forces. First, we document that our saliency intervention induces search (visiting the privacy policy), but not learning (knowing the privacy practices). In contrast, our information intervention both induces search and learning. To assess this, our endline survey asks participants to recall the privacy practices of their five most visited websites. We find that the participants in the information treatment are 7.1 percentage points more likely to be correct about site privacy practices than both the saliency and control groups.[3] While participants in the saliency group may not learn, they do search. Participants in the saliency group viewed privacy policies nearly 17 times more than those in either the information or control group. The combination of these facts points to the important role of information complexity in limiting learning in privacy settings. Finally, we demonstrate that changes in beliefs do translate into small behavioral changes. Participants in the information group exhibit a modest shift, primarily on the extensive margin, to sites more aligned with the participants' privacy preferences and a slight decrease in data sharing. Furthermore, following Haaland et al. (2025), qualitative coding of self-reported changes during the experiment suggests small behavioral adjustments to the intervention.

We next estimate an empirical analogue of our model of website choice in order to fully quantify the welfare benefits of privacy online and of different privacy regulation provisions. As in our theoretical model, participants trade-off the direct utility of consumption and the costs associated with privacy over which there is uncertainty with both search and complexity frictions involved in acquiring information. In addition to the *revealed* preference data of actual browsing choices during our experiment, our model leverages measures of participant's *stated* preferences over privacy attributes – elicited using a conjoint survey in the first stage of our study (Allenby et al., 2019). We rely on these stated preferences to map utilities to dollars and to estimate the gap between stated and revealed preferences for privacy – an object of extensive debate within the privacy literature (Norberg et al., 2007; Adjerid et al., 2018; Athey et al., 2025). We find that the stated preference measures overstated revealed preferences by approximately 34% – directionally consistent with the literature but provides the first quantification for how large this gap is.

We use the model estimates to evaluate how much consumers value privacy online, showing that at baseline consumers experience disutility equivalent to paying $6.71 per month from the privacy-invasive practices they experience. However, this is small compared to the positive utility they gain from the quality of the websites they use. We use our rich privacy characteristic space to

[3] In the intervention we persistently provide the top two most informative attributes for the participant based on their beliefs and stated preferences.

characterize which privacy attributes participants care the most about. Our privacy characteristic space contains practices involving collection (e.g., financial, location, biographical data), usage (e.g., sharing with advertisers, law enforcement, financial transactions), and control (e.g., data deletion, anonymization) and our estimates suggest that consumers have disutility from websites engaging in collection and usage, but positive utility from control. In terms of absolute magnitudes, we find that consumers assign the most weight to collecting and sharing financial data, collecting sensitive information and off-site data, and sharing data with law enforcement. Additionally, consumers do not have large disutility from collection of browsing and demographic data. These results paired with our survey evidence suggest that most of the privacy disutility comes from the potential downstream consequences (e.g., identity theft, hacking) involved with data sharing.

We then use our model to evaluate the welfare benefits from two distinct aspects of privacy regulation: information provision and direct regulation of privacy practices.

We begin with the full information benchmark, in which participants are exogenously provided with true privacy information for all websites online. We estimate full information leads to a welfare increase of 79% relative to the status-quo, or \$0.67 per month per participant. Still, an important element of disclosure regulation online is that information is made salient only on visited websites. This motivates our targeted information counterfactual, in which we remove both search costs and complexity – leading participants to acquire full information only for the sites they browse. Our targeted information provision provides a welfare gain of 55% over the status quo, or 70% of the full information benchmark. This amounts to a welfare increase of \$0.47 per month per participant. Finally, to disentangle the complexity and search cost mechanisms, we estimate a counterfactual in which we remove complexity and allow participants to endogenously search for privacy information as they browse. Removing complexity leads to a small welfare gain of 12% over the status quo. The small gain is rationalized by the small likelihood of search: only 2.1% of participant-website interactions include search. These results highlight how addressing both complexity and search costs is necessary for effective information provision in privacy.

We next examine how direct regulation of privacy practices, in the spirit of regulations like the GDPR, impact welfare. A key provision of the GDPR is the right to opt out of the collection and sharing of: special-category data (sensitive information), purpose limitation and data minimization (collecting location and offsite data), and third-party sharing restrictions (sharing with advertisers, partners, and social media platforms). We estimate an upper bound welfare effect of banning these practices across all consumers, showing that a joint ban leads to \$1.22 per month increase in welfare for consumers or an increase of 143% over the status quo. Interestingly, the bundle restriction leads to larger gains than the sum of restricting each of the attributes individually – highlighting complementarity that is driven by consumer reallocation of usage of different websites. The largest gain comes from restriction of offsite browsing data as consumers value this and

it is relatively prevalent online. Critically, these estimates are an upper bound because we do not allow for websites to endogenously adjust quality (or their privacy attributes) in response to our intervention. Further, its important to note that these restrictions have been shown to be associated with large costs and revenue losses for firms (Goldberg et al., 2024; Aridor et al., 2023, 2025), suggesting such quality adjustments could be large.

Overall, these results have several important policy implications. First, consumers have reasonably poor understanding of the prevalence of privacy practices online which is primarily driven by the complexity and length of privacy policies. Second, consumers have meaningful valuations for privacy online, but they are dwarfed by how much consumers value their most frequented sites. Third, direct regulation of practices has the largest scope for welfare gains – though this is the channel most likely to lead to firm revenue losses – but there are also meaningful welfare gains from information provision alone, given the lack of information by consumers. Thus, regulations targeting the information disclosure environment are likely to generate meaningful welfare gains without as much scope for firm revenue losses as direct regulation of practices.

2 Related Work

Our paper primary contribution is to two significant strands of literature on the economics of privacy and information disclosure.

Economics of Privacy. Our paper contributes to a growing literature on the economics of privacy (Acquisti et al., 2016; Goldfarb and Que, 2023) by empirically measuring the value of privacy online.[4] A key concern of this literature has been the evaluation of privacy regulations, with an emphasis on understanding how privacy regulation trades off benefits to consumers with costs to firms. Many studies have documented substantial regulatory costs due to privacy regulation (Goldfarb and Tucker, 2011; Johnson et al., 2020; Goldberg et al., 2024; Aridor et al., 2023; Wernerfelt et al., 2025; Aridor et al., 2025; Alcobendas et al., 2026).[5]

Still, understanding the potential benefits of privacy regulation to consumers has proven difficult. Most work focused on characterizing the structure of privacy preferences using surveys or lab experiments, as opposed to generating policy-relevant estimates of valuations in the field. There has been a significant focus on a perceived gap between stated and revealed preferences, often re-

[4]There is also a substantial literature in computer science on using computational methods to analyze privacy policies (for example (Amos et al., 2021; Wagner, 2022; Tang et al., 2023)) which we omit here.

[5]For a more general overview of the economic consequences of privacy regulation, see Johnson (2024). Privacy regulation has also been shown to influence firm data usage (Demirer et al., 2024), decrease VC investment (Jia et al., 2021), increase web tracker concentration (Johnson et al., 2023; Peukert et al., 2022), and adversely impact app entry and exit (Janssen et al., 2022).

ferred to as the "privacy paradox," (Spiekermann et al., 2001; Acquisti and Gross, 2006; Norberg et al., 2007; Athey et al., 2025) and the highly context-specific nature of privacy preferences (Acquisti et al., 2013). Other research has worked to disentangle intrinsic preferences for privacy from instrumental preferences (Lin, 2022). Recent work has explored whether the observed wedge between stated and revealed preference is due to imperfectly informed consumers (Tsai et al., 2011; Adjerid et al., 2018, 2019; Bian et al., 2025; Collis et al., 2023). This paper builds upon this work by constructing measures of both stated and revealed preferences, relying on our experimental intervention and structural model to estimate the degree to which stated preferences are aligned with revealed privacy preferences. Combining our model and experiment allows us to estimate consumer welfare under different counterfactual regulatory designs.

Within this literature, the two closest papers are Bian et al. (2025) and Fradkin et al. (2024). Bian et al. (2025) studies the market level effects of the iOS privacy nutrition labels, showing that information transparency has a large impact on consumer application download choices and closes the gap between stated and revealed preferences. Our paper differs by relying on experimental manipulations of information transparency that allows us to more clearly elucidate mechanisms, directly quantify the gap between stated and revealed preferences, and provide welfare estimates. Fradkin et al. (2024) similarly uses an extension-based field experiment to study consumer privacy preferences, but is complementary to our paper by studying the welfare effects of different consent dialog designs as opposed to measuring the underlying privacy preferences.

Information Provision, Obfuscation, & Field Experiments. A large literature has examined the ways in which firms use information frictions to manipulate consumers, even when disclosure is mandatory (Dranove and Jin, 2010; Jin et al., 2022). This issue arises in many markets such as healthcare, financial earnings reports (deHaan et al., 2020), e-commerce (Brown et al., 2010; Blake et al., 2021), and college admissions (Luca and Smith, 2015). Our setting is a canonical example of complex disclosure; firms are legally forced to disclose their privacy practices, but obfuscate the objective truth in legalese. By experimentally manipulating the magnitude of information frictions our paper also contributes to the literature that characterizes how regulation can assist consumers in making informed choices under information frictions. Existing work has studied policy remedies in the specific context of frictions in food choice (Barahona et al., 2023), school choice (Agte et al., 2024), health care plan choices (Handel and Kolstad, 2015; Brown and Jeon, 2024), and the fuel economy market (Allcott and Knittel, 2019).

Our experimental intervention directly manipulates the level of complexity in disclosure, relying on an information provision design (for a survey of these types of experiments see Haaland et al. (2023)) that randomizes both the search and complexity costs involved with privacy information – the two key frictions in our setting. We integrate this information intervention into a Chrome

Extension so that it affects the actual web browsing habits of experimental participants for a period of two months. Methodologically, this follows a recent paradigm that has been used to study digital platforms by using researcher-generated variation to speak to policy issues with our paper doing this at the scale of the entire internet, as opposed to an intervention targeting one platform (Levy, 2021; Aridor, 2025; Beknazar-Yuzbashev et al., 2022; Allcott et al., 2025; Fradkin et al., 2024) – see Aridor et al. (2026) for an overview.

3 Theoretical Framework

3.1 Model of Website Choice and Information Acquisition

Consider a risk-averse, forward-looking participant who, in each period t, chooses a website j to visit (or takes an outside option). Preferences for a website are determined by the (known) quality dimension of the website and the participant's current beliefs about the site's privacy practices. If a participant picks website ($j \neq 0$), her deterministic flow payoff is given by:

$$u_{ij} = \lambda_i^q Q_j(\theta_{ij}^b) + \lambda_i^p E_{\theta_{ij}^b}[\mu_j^*] \tag{1}$$

$$U_{ij} = u_{ij} - \frac{\gamma}{2} u_{ij}^2 \tag{2}$$

where θ^b is participant i's prior belief over site j's privacy type, Q_j is website j's observable quality, and γ is the participant's coefficient of absolute risk aversion.

Given these preferences, a natural question is whether participants acquire information to refine their beliefs about privacy practices. A distinctive feature of the online privacy setting is that participants can typically only observe privacy information *after* visiting a website – unlike most information disclosure environments where the relevant information is available prior to purchase. This timing has direct implications for our experimental design. Accordingly, we model a two-stage process: participants first choose which website to use in a given period, and then decide whether to acquire and process the privacy information of that site.

Our model incorporates two information frictions. The first is a cost, denoted c_{ij}, that the participant must pay to find and attend to the privacy policy. This cost has two components: a search cost c_j^s that captures the difficulty of locating the privacy policy on website j, and an attention cost c_i^a that captures the cognitive cost of attending to privacy information, even when it is directly accessible. In the default regime online in the United States, this captures the combined cost of finding the privacy policy, accessing it, and choosing to read it.

Once the participant pays the cognitive cost, she updates her belief to the true state μ^* with probability α_j, and with probability $1-\alpha_j$ her belief remains at her prior θ^b. We interpret α_j as the

probability of success in acquiring the true privacy information, or the *information complexity*. A lower α_j represents a more complex privacy policy and increases the likelihood that a participant searches but learns nothing. Importantly, α_j is chosen by the firm – firms control the complexity and clarity of their privacy disclosures, providing a potential channel through which they can influence participant information acquisition.

Once a participant visits a website, she decides whether to acquire information by trading off the future, potentially discounted, value of information with the current cost of acquisition. We formalize this with the following definitions:

Definition 3.1 (Search)**.** A participant is said to *search* for information if she chooses to acquire information and pays the cost c_{ij}.

Definition 3.2 (Learn)**.** A participant is said to *learn* if she searches and successfully updates her prior – that is, she pays the cost c_{ij} and draws the true information from the privacy policy. A search that does not result in learning is said to *fail*.

We refer to this general process, and choices about search and learning, as the participant's *information acquisition process*.

3.2 Experimental Hypotheses

Our experiment is designed to identify each of these information frictions and consists of two treatment groups:

1. **Saliency**: Reduce the overall cost of search ($\downarrow c_j$)
2. **Information**: Reduce the cost of search ($\downarrow c_j$) and the complexity of information ($\uparrow \alpha_j$)

The combination of these two treatments – one that varies only search costs and another that additionally varies complexity – allows us to separately identify the role of search and learning in information acquisition.

With this design in hand, we state the following hypotheses. These hypotheses – while of independent interest – provide direct tests for whether the intervention is strong enough to change beliefs and behavior needed to separately identify each of these components needed for the structural model discussed in Section 8.

Hypothesis 1 (Information Acquisition). We first examine whether reductions in information frictions change information acquisition behavior and, in turn, participant beliefs. If participants

attend to and internalize the provided information, belief accuracy should be highest in the information group, intermediate in the saliency group, and lowest in the control group. Formally:

$$|E_j[\theta^b_{ij}|i \in I] - \mu^*_j| \leq |E_j[\theta^b_{ij}|i \in S] - \mu^*_j| \leq |E_j[\theta^b_{ij}|i \in C] - \mu^*_j|$$

This is the most direct test of whether the information intervention "worked," as it measures whether participants attended to and internalized the information that was provided. Our rich data on information acquisition and experimental variation allows us to go further and distinguish whether the success (or failure) of the information intervention hinges on search – finding and attending to the information – or learning – internalizing the information conditional on attending to it.

Hypothesis 2 (Search & Learning). The core mechanism we wish to identify is whether reduced search costs alone suffice for learning, or whether reducing complexity is additionally necessary. Among treated participants:

Hypothesis 1 (Search & Learning)**.** Among participants that are treated:

(i) Search behaviors increase in the treatment groups.

(ii) The probability of learning, conditional on search, is the same in the salience and control groups.

(iii) Conditional on search, the probability of learning is greater in the information group than in the salience group.

Parts (ii) and (iii) are the key tests: if the saliency group searches more but learns no more than the control group, while the information group learns more conditional on search, this isolates complexity as the binding friction.

Hypothesis 3 (Changes in Data Sharing Behavior). Even if the information intervention successfully transmits information, it only matters insofar as it materializes in changes in participant choices. This is especially important in the privacy context, where a central debate is whether participants genuinely value privacy or merely state that they do (Adjerid et al., 2018; Athey et al., 2025). While our model focuses on website choice, a natural behavioral margin is data sharing: participants may respond to privacy information by directly withholding information or using privacy-preserving actions (e.g., opting out, deleting cookies). Treated participants may shift their data sharing behaviors along these dimensions.

Hypothesis 4 (Visits & Website Heterogeneity). The most direct channel through which participants may respond – and the one embedded in our model – is changes in browsing behavior. If privacy preferences matter (i.e., $\lambda_i^p > 0$), then learning about privacy practices should lead to changes in website usage. While clean characterizations of how aggregate browsing will change depend on the distribution of privacy practices and beliefs, a clear implication of our model is that websites with better (worse) privacy practices should see an increase (decrease) in visits. Denoting $v_j^i(t)$ as visits by participant i to site j in period t:

$$\begin{aligned} &E_j\left[v_j^i(t_{\text{post}}) - v_j^i(t_{\text{pre}}) \mid j = \theta_j^H,\, i \notin C,\, v_j^i(t_{\text{pre}}) > 0\right] \\ &> E_j\left[v_j^i(t_{\text{post}}) - v_j^i(t_{\text{pre}}) \mid j = \theta_j^H,\, i \in C,\, v_j^i(t_{\text{pre}}) > 0\right] = 0 \end{aligned}$$

This effect should be more pronounced in the information group than in the saliency group, given Hypothesis 1.

To summarize, our experimental design maps directly to these hypotheses: the saliency treatment varies c_j^s, the information treatment additionally varies α_j, and the combination allows us to separately identify search and learning as mechanisms through which disclosure operates.

4 Experiment Description and Data

Implementing this experiment requires two novel elements: a comprehensive dataset of website privacy practices and a browser extension capable of both observing participant behavior and delivering experimental treatments at scale. We introduce each building block before detailing the experimental design. Additional details on the datasets and associated technology are provided in Appendix A.

4.1 Auxiliary Data

4.1.1 Privacy Practices Dataset

Our experiment relies on accurate and comprehensive information about website privacy characteristics. We construct a novel dataset of privacy indicators taken directly from website privacy policies for the top 500 most visited websites in the United States. To our knowledge, this is the first dataset that systematically codes the full breadth of privacy-relevant attributes from website privacy policies at this scale.

Privacy policies are legal documents presenting the terms and rights associated with using a website, and they serve as the primary source of participant information on a website's privacy practices. These policies are often obtuse, difficult to find, and require significant expertise to

understand (Srinath et al., 2023). We leverage a team of privacy experts, law students and trained raters to carefully analyze and digest these policies into binary characteristics of website behavior. Table A.1 presents an exhaustive list of the nineteen collected privacy attributes.

To ensure accuracy, each policy is read by three independent raters who answer a series of binary questions corresponding to different privacy attributes. In addition, we elicit rater certainty on a five-point scale for each attribute. Raters are trained using twelve practice policies randomly chosen from the top 500 websites, which they score alongside a law student. With three raters and nineteen characteristics, this amounts to 57 different privacy indicators per domain. Privacy policies take on average 17.8 minutes to read, with some exceeding 60 minutes; in total, the rating process has cumulatively taken 383.8 human hours. Additionally, in Appendix A we cross-reference the tracker data collected by the Chrome Extension with the privacy policy data. We focus on attributes that we can reliably measure using the tracker data, such as "Collecting Browsing on-site data", and show that these two measures are in general accordance.

4.1.2 Experimental Extension

Our primary data is collected by a custom-developed browser extension based on the same underlying content detection as the popular AdBlockPlus extension. The extension serves two purposes: passive observation of participant behavior and delivery of experimental treatments.

On the observation side, the extension collects several categories of data for each participant. First, it records the total seconds spent on each subdomain (e.g., `drive.google.com`), aggregated to the daily level, with robust detection of user idleness to ensure accurate time measurement. Second, it collects full browsing history, which we primarily use to detect visits to privacy policies. Third, it captures the identity and quantity of third-party requests and associated cookies for each website using the same underlying technology as AdBlockPlus.[6] Fourth, it logs interactions with cookie banners and privacy policies.[7]

On the treatment side, the extension enables us to manipulate the information environment by providing salient access to privacy policies, displaying interpretable summaries of privacy practices, and clearing third-party cookies. Additional technical details are available in Appendix A.

[6] As we document in Appendix A, we rely on a Javascript implementation of the AdBlockPlus parser paired with the 'Easylist', 'Easylist Cookie', and 'Easylist Privacy' detection lists from `easylist.to`. Every network request on the browser runs through this list and we log the information for requests that AdBlockPlus would have blocked.

[7] We implement interactions with cookie blockers by using the 'Easylist Cookie' list and logging this when a participant mouses over a DOM element associated with a cookie banner based on the list.

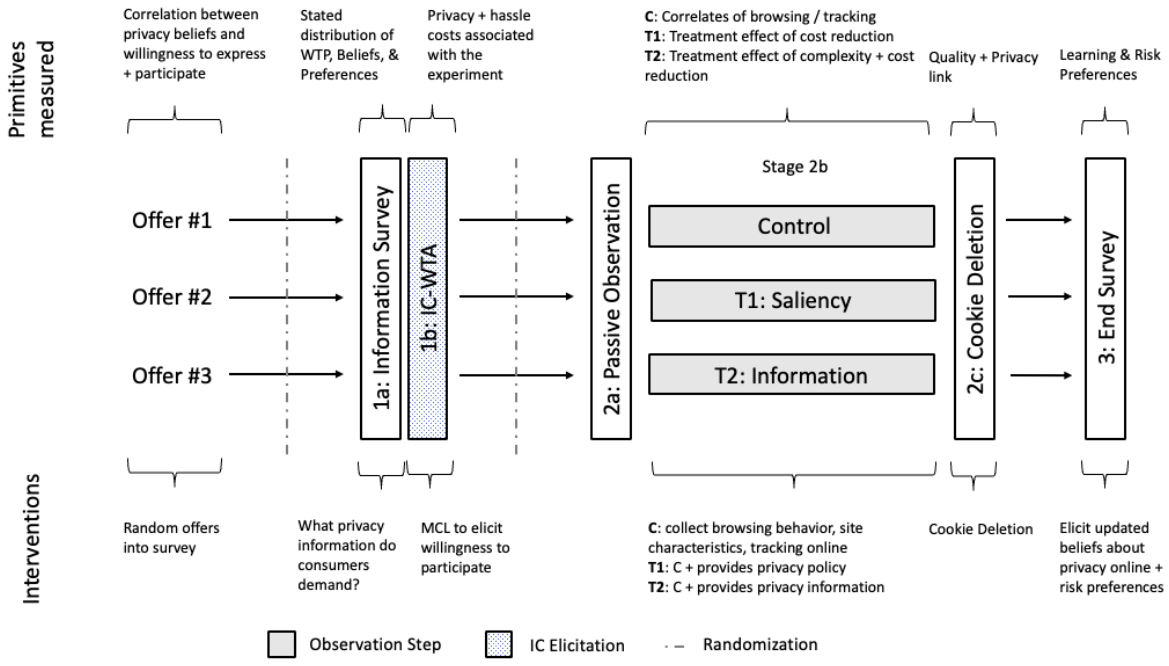
4.2 Experimental Design and Implementation

In this section, we introduce and motivate our experimental design and detail the experimental timeline. Figure 1 provides an overview of the experiment. Below, we explain each piece in detail.

4.2.1 Recruitment & Baseline Survey

We recruit participants from the online research platforms CloudResearch Connect and Prolific. To minimize selection based on privacy preferences, the recruitment materials make no mention of privacy; participants are told only that we are conducting "a study to understand attitudes about different websites."[8] Participants first complete a screening survey that asks whether they use Google Chrome as their primary browser and if they are willing to provide their email address for a follow-up study.

Figure 1: Experiment Overview



NOTES: Schematic representation of the experimental design. The timeline proceeds as follows: screening survey with randomized compensation, baseline survey including belief elicitation and conjoint preference measurement, willingness-to-accept elicitation via multiple price list, passive observation period (2 weeks), treatment randomization into Control, Saliency, or Information conditions, intervention period (4 weeks), cookie deletion phase (2 weeks), and endline survey. The primitives measured at each stage are indicated in the top row, with corresponding interventions in the bottom row.

[8] We advertise and randomly offer three different levels of compensation for participating in the survey, \$3, \$4, and \$5. Randomization of offers induces exogenous variation in selection into the experiment, which we exploit in Section 4.5.

The study begins with a baseline survey that accomplishes four things. First, we replicate a series of questions on privacy practices and attitudes from PEW (McClain et al., 2023) to assess the representativeness of our sample on privacy attitudes. Second, for each of the 19 privacy attributes in our dataset, we elicit participants' beliefs about their prevalence online. Third, we use a partial-profile conjoint (Green et al., 2001; Allenby et al., 2019) to elicit participants' willingness to pay for each privacy practice. Because privacy preferences vary by context, we randomize the website category in which the conjoint is framed.[9] Fourth, we use a switching multiple-price list (Andersen et al., 2006) to incentive-compatibly elicit participants' willingness to accept (WTA) compensation for participating in the observation stage of the study.[10]

Second stage eligibility. Following the baseline survey, a random offer from the multiple-price list is chosen for each participant. If their stated WTA is less than or equal to the random offer, they are offered a spot in the second stage and asked to install the browser extension. Eligible participants receive an email containing an installation link and an individualized experiment ID that serves both as validation of successful installation and as a link to our backend-generated identifier, ensuring that we can track each participant across the entire intervention period.[11]

4.2.2 Information Intervention & Treatment Assignment

Our intervention provides participants with comprehensive and credible information on the privacy practices of the websites they browse. The primary randomization includes a *control*, a *saliency*, and an *information* group. The control group installs the extension, which passively collects time spent, privacy information acquisition, and third-party trackers

Saliency Treatment. Participants in the saliency group are provided with direct and easy access to website privacy policies. The extension introduces a prominent pop-up in the upper right corner of the browser when a participant arrives on a website. If the participant clicks this link, they are directed to the privacy policy page (see Figure 2a for an example on the New York Times website). Mapping back to our conceptual framework, this treatment reduces search frictions associated with finding privacy information online but does not change the attention costs associated with comprehension nor the complexity of the information. In the language of our model, this induces experimental variation in c_j^s.

[9] See Appendix B for additional information on the conjoint.

[10] Participants are given an example and complete comprehension questions to ensure they understand the procedure for both the conjoint and the WTA elicitation.

[11] In the case when the participant has multiple devices, we are able to identify each device individually but the participant across these devices and aggregate across devices for the analysis.

Figure 2: Experimental Intervention



(a) Saliency Intervention



(b) Information Intervention

Information Treatment. Participants in the information group are given three pieces of privacy information through a transparent and salient pop-up in the upper right corner of the browser, providing experimental variation in α_j (see Figure 2b). Participants can click this box to get additional information on further privacy attributes, as well as definitions of each attribute.[12]

A key feature of our information treatment is personalization: we select the privacy attributes that are most informative for each individual participant. Two of the three displayed attributes are held constant across websites and personalized using each participant's conjoint estimates and elicited beliefs. Specifically, we calculate the expected information gain from attribute k for individual i as:

$$V_{ik} = E[u_{\text{post}}] - E[u_{\text{prior}}] = \theta^b_{ik}(1 - \theta^b_{ik})\beta_{ik} \tag{3}$$

where θ^b_{ik} is individual i's belief that a random website in the top 500 engages in practice k and β_{ik} is their preference weight from the conjoint. The formula captures the intuition that information is most valuable when a participant both cares about the attribute (high $|\beta_{ik}|$) and is uncertain about its prevalence (beliefs near 0.5). We take the two attributes with highest V_{ik} as individual i's personalized treatment. The third attribute is chosen at random and varies across websites.

4.3 Cookie Deletion State & Endline

Third-Party Cookie Deletion. After four weeks of treatment, we begin the final portion of our intervention, during which we induce exogenous variation in data sharing by randomizing the deletion of third-party advertising and analytics cookies. We identify these cookies using the the custom AdBlockPlus parser built into the Chrome Extension that allows us to identify when a network request is sent to a third-party domain and continually delete these cookies.[13] Treatment timing is staggered: we randomize within block to have 2 of the 3 participants (for even blocks) or 1 of the 3 participants (for odd blocks) begin cookie deletion at the start of week 7, while the remaining participants start at the beginning of the final week of the study. This staggered rollout generates variation in both the presence and duration of cookie deletion, which we exploit for identification of cookie-value parameters in the structural model (Section 9).

Endline Survey. After the treatment period is completed, we conduct an endline survey to validate our treatment. For each participant, we elicit beliefs about the privacy practices of their top 2 preferred attributes on the five most visited websites during the intervention period. In addition,

[12] Participants can click to find extended descriptions of the attributes as well as see the full list of attributes, apart from the three presented. The additional information dialogs can be seen in Figure C.9.

[13] Analysis and categorization of the collected trackers from the extension (using the process described in Appendix A) suggests that these are predominantly third-party domains associated with data collection for advertising and analytics.

we implement a simple bonus lottery to measure risk preferences (Charness et al., 2013; Dave et al., 2010) and re-elicit privacy attitudes. The full details of questions asked in the surveys is documented in Appendix E.

4.4 Experimental Implementation and Sample

Timeline and Recruitment. We conducted our experiment from June 14th until August 23rd, 2025, recruiting participants based in the United States from both CloudResearch Connect and Prolific. Recruitment occurred in two waves, with the first wave running from June 14th until August 9th and the second from June 28th until August 23rd. Participants recruited in each wave followed the same timeline conditional on study start date.

Figure 3: Participant Flow through Study

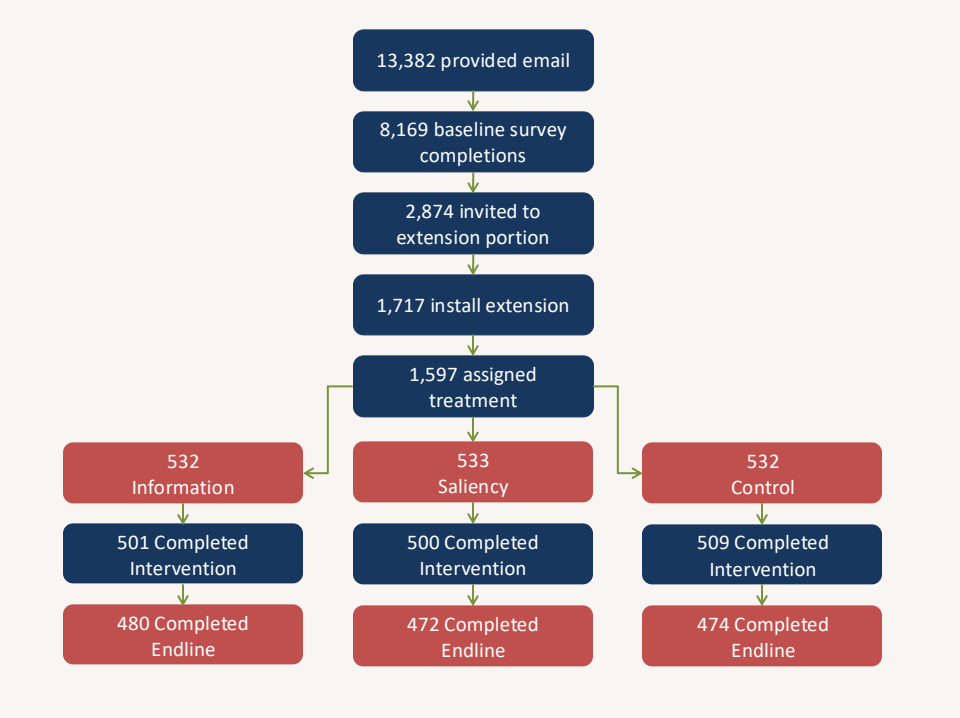


Figure 3 presents the flow of participants through the study. Of the 13,382 participants who provided their email and use Google Chrome, 8,169 (61%) completed the baseline survey. Of these, 2,874 were eligible for the extension phase (35%), 1,717 successfully installed the extension (60% of eligible), and 1,597 remained by the time treatment was assigned (93%). The drop from 1,717 to 1,597 is attributable to technical issues with older laptops, manual removal of participants with abnormal behavior, and general attrition. Throughout, we refer to the participants who completed

the baseline survey as the *survey* sample and those who completed the extension portion of the study as the *extension* sample. For the extension sample, the study proceeds as follows:

1. Baseline Period (two weeks)
2. Primary Information Intervention Period (four weeks)
3. Cookie deletion (two weeks)
    - First wave of third-party cookie deletion (two weeks)
    - Second wave of third-party cookie deletion (one week)
4. Endline Survey (one day after extension uninstall)

Experimental Sample. The average participant is 39.4 years old, 59.9% of the sample is female, and 58.6% of the sample has at least a college education. To assess demographic representativeness, we compare the demographic composition to the 2021 U.S. Census on age, education, gender, and income in Figure C.1. Overall, our sample is reasonably representative but skews younger, more female, and more educated than a fully representative sample. Importantly, the demographic composition is consistent between the survey and extension samples.

In the baseline period, the average participant spends 1.2 hours on websites with privacy information and 2.4 hours online.[14] The top websites used by participants are displayed in Table C.2 and correspond to domains consistent with the top most visited domains for which we collect privacy information – e.g., YouTube, Google, Facebook.

To ensure similar exposure to the information intervention across treatment groups, we block-randomize based on baseline usage. Specifically, we order participants by their total time spent on websites for which privacy information is available and, within each consecutive block of three participants, randomly assign treatments. In addition, we enforce balance on total time spent, demographics, privacy knowledge, WTA, beliefs, and conjoint category. Table C.1 confirms that this procedure achieves balance across treatment groups on each of these dimensions.

From the experimental sample, 501 (of 532), 500 (of 533), and 509 (of 532) participants remained in the information, saliency, and control groups, respectively, with their extensions installed at the end of the intervention.[15] Of these, 480, 472, and 474 completed the endline survey.[16] Standard balance checks provide no statistical evidence of differential attrition: chi-square tests across

[14] Since our participants are recruited from survey recruiting platforms, for this calculation and throughout the paper, we drop time spent on survey websites (e.g., qualtrics, surveymonkey) and recruitment platforms (e.g., prolific, cloudresearch, mturk, paidviewpoint). This ensures that we are focusing on 'leisure' sites as opposed to 'labor' sites. Including usage on these websites increases the average daily time spent online to 3.1 hours.

[15] When participants uninstall the extension, they are automatically opted out of the study.

[16] Some of the attrition between the final extension sample and the endline survey is driven by email delivery issues.

treatment groups yield $p = 0.37$ for the extension sample and $p = 0.96$ for the endline sample, and logistic regressions with attrition as the dependent variable confirm the absence of overall or pairwise differential attrition (Gerber and Green, 2012).

4.5 Selection into Study Based on Privacy Preferences

A key concern with our study is that we are studying privacy preferences using data collection that is potentially privacy-invasive. Our study is designed to precisely measure the magnitude of this selection. There are three points at which selection on privacy preferences might occur: into the recruiting platforms, into responding to the survey, and into installing the Chrome Extension. We address each in turn and find no evidence of meaningful selection on privacy preferences at the first two stages; at the third stage, our WTA measure allows us to precisely bound the extent of selection.

Selection into the survey. We mitigate possible selection by advertising the study generically, with no mention of privacy. To characterize any remaining selection, we randomize the advertised first-stage survey payments between \$3, \$4, and \$5. If privacy preferences influence the decision to participate, we would expect our measured privacy preferences and belief misspecification to monotonically vary with the survey payments. Figure C.2 compares aggregated belief error measures, our PEW index, and WTA across the different survey payments and reveals little systematic variation.

Selection into the platform. Online recruitment platforms may suffer from selection bias, as their users are likely more willing to share information online than the general population. To quantify this, we replicate a subset of questions from PEW's nationally representative survey on privacy preferences. Figure C.3 compares our results with PEW's. On average, our participants are more likely to immediately accept privacy policy terms and more likely to switch websites due to privacy, with similar beliefs about policy effectiveness. Given there is no clear ordering between the samples in terms of privacy sensitivity, we interpret this as indicating that platform-level selection based on privacy preferences is not dramatic.

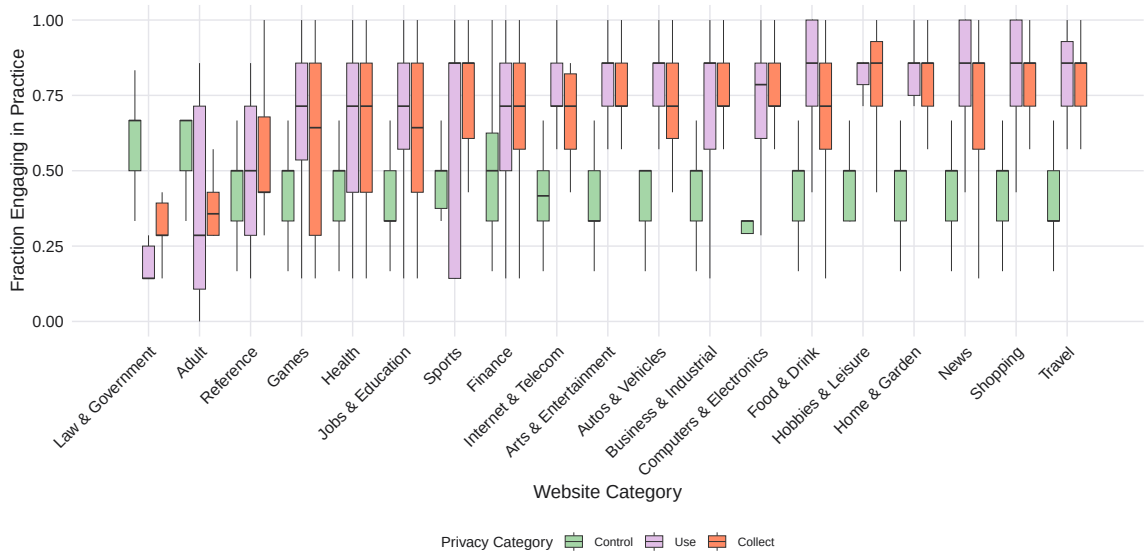
Selection into the extension. While the survey is advertised generically, the purpose of the experiment becomes clearer once participants are asked to install a browser extension that continuously monitors their browsing. We expect more privacy-sensitive participants to require higher compensation. The final part of the survey uses the incentive-compatible WTA mechanism to directly measure this. The median (mean) WTA in the invited sample is \$20 (\$19), compared to \$25 (\$52) in the full survey population. Differences arise from a cap on maximum compensation of

$40, which we impose to manage experimental costs. Interpreting WTA as a proxy for privacy preferences, our final experimental sample allows us to estimate treatment effects for the bottom 89% of participants in terms of the intensity of privacy preferences. Our estimates are informative about the large majority of the population but may not generalize to the most privacy-sensitive individuals. Figure C.4 compares the conjoint estimates between the full survey sample and the extension sample finding that, while the privacy valuations for the survey sample are higher for most attributes, the relative ordering is the same and the difference is not dramatic.

5 Privacy Practices, Beliefs, and Stated Preferences

For our information intervention to influence participant choices, three conditions must hold: there must be meaningful heterogeneity in privacy practices across websites within plausible substitute sets, participants must have uncertainty about the privacy practices of sites, and the attributes on which we provide information must be ones participants care about. In this section, we document that all three conditions are satisfied.

Figure 4: Privacy Practices across Website Categories



NOTES: Distribution of privacy practice counts by website category. Privacy practices hand-coded from privacy policies of top 500 US websites and aggregated into three indices: collect (8 attributes), control (5 attributes), and use (7 attributes). Each website receives a score equal to the count of practices engaged in within each category. Website categories obtained from the Klazify API. Box plots display median, interquartile range, and outliers.

5.1 Privacy Practice Heterogeneity

We begin by characterizing variation in privacy practices across websites. We categorize each website using the Klazify API and aggregate each domain into a single score for each of the three privacy practice categories described in Table A.1: *collect* (8 attributes relating to data collection), *use* (7 attributes relating to how data is shared), and *control* (5 attributes relating to user protections). Based on our conjoint estimates (discussed below), engaging in more "control" practices is generally utility-improving for participants, while engaging in more "collect" or "use" practices is utility-decreasing.

Figure 4 presents box plots documenting the variation in these indices across website categories. Two patterns stand out. First, there is considerable heterogeneity *within* categories — even among websites offering similar content, privacy practices vary substantially. This is precisely the variation that informed participants could act on. Second, while median scores are relatively similar across categories, two notable exceptions emerge: "Law & Government" and "Adult." These categories consist of websites that naturally collect more sensitive information (e.g., legal records, health data), and accordingly tend to have stricter privacy practices. This pattern provides useful validation that our hand-coded data captures meaningful differences.

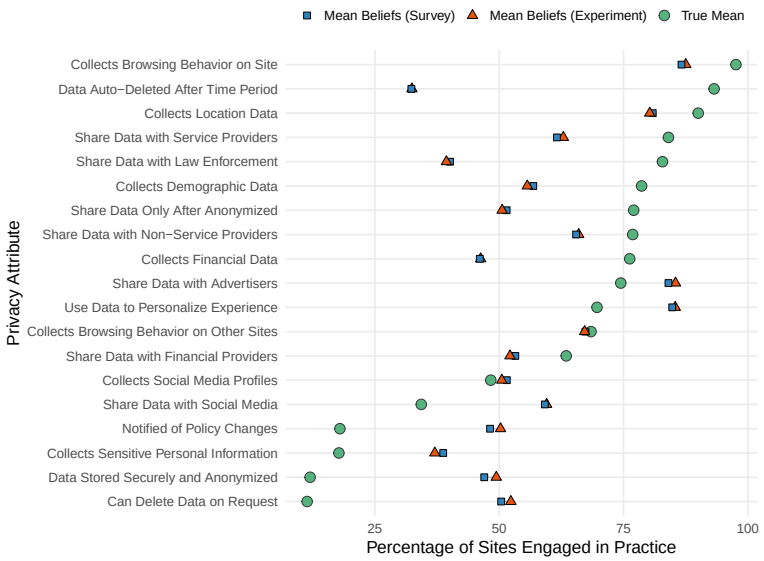


Figure 5: Aggregate Beliefs about Privacy Practices

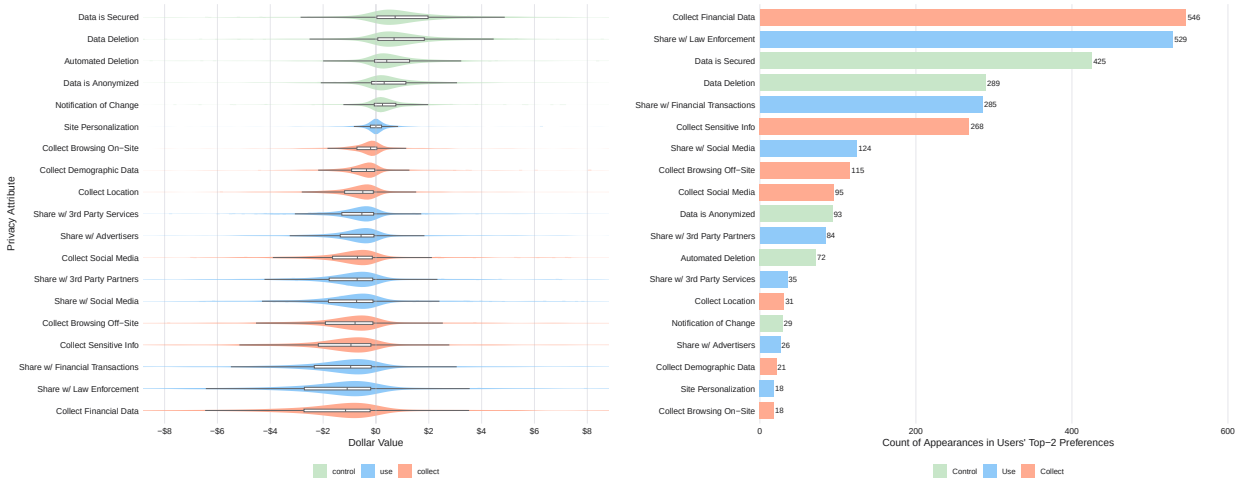
NOTES: Mean prior beliefs versus ground truth prevalence. Beliefs elicited on 5-point Likert scale, converted to percentage points via the transformation $(r-1)\times 25$, where $r \in \{1,2,3,4,5\}$. Ground truth computed as the proportion of websites engaging in each practice.

5.2 Beliefs about Privacy Practices

We next characterize what participants believe about privacy practices and how these beliefs compare to our collected data. In the baseline survey, we ask each participant: "For a random website in the top 500 online, how likely do you think it is they engage in [PRIVACY PRACTICE]?" for each of the 19 attributes.

Figure 5 presents mean beliefs alongside the true prevalence of each attribute.[17] Participants' beliefs are clearly misspecified for most attributes, though the direction of misspecification varies. For instance, participants correctly perceive that on-site browsing behavior and location data are among the most commonly collected attributes, but they underestimate the extent of this collection. Conversely, participants substantially overestimate the prevalence of data collection for advertising purposes — consistent with public discourse around online privacy focusing disproportionately on advertising. Column (1) of Table C.5 shows that the only significant demographic predictor of misspecification is education as individuals with a bachelor's degree or above have a lower degree of misspecification.

Figure 6: Privacy Valuations and Most Valued Information



(a) WTP for Privacy Attributes

(b) Most Valued Information

NOTES: Both panels use the extension sample of participants who installed the browser extension. Panel (a) presents the estimated monthly willingness to pay for different privacy attributes, computed using the estimated part-worth for each attribute and the price coefficient. Green rows denote "Control" attributes, blue rows denote "Use" attributes, and orange rows denote "Collect" attributes. Appendix B contains the estimation details. Panel (b) shows the distribution of top-2 personalized attributes. Information value is computed per Equation (3). The full conjoint sample version is in Figure C.5.

[17] Figure 5 displays beliefs for both the extension and survey sample showing that the selection into the experiment does not induce large differences in participants' beliefs about the prevalence of privacy practices online.

5.3 Stated Privacy Preferences

To understand which privacy attributes participants value most, we estimate willingness to pay (WTP) for each attribute from the conjoint task completed in the baseline survey. Figure 6a presents the estimated monthly WTP.[18] Control attributes — data deletion, data security, anonymization — are associated with positive WTP, consistent with participants viewing these protections as valuable. In contrast, collect and use attributes are associated with negative WTP: participants would need to be compensated for using a website that collects financial data, shares information with law enforcement, or collects sensitive personal information. Column (2) of Table C.5 shows that education, being over 60, and being female are associated with valuing privacy more highly.[19]

The most valued attributes are those associated with potential off-platform consequences: collecting and sharing financial information (approximately $6–8/month), sharing with law enforcement, collecting sensitive information, and data security. This pattern is consistent with participants' self-reported concerns about data sharing (Table C.8), which emphasize loss of data control, identity theft, and hacking as the most important consequences.[20]

5.4 Distribution of Personalized Treatment Attributes

The personalized attributes provided to each participant in the information treatment reflect the combination of these two inputs — preferences and belief uncertainty — through Equation (3). Attributes appear in a participant's personalized set when the participant both cares about the attribute and believes there is meaningful variation in practices across websites.

Figure 6b presents the distribution of the top 2 personalized attributes across participants. Collecting and sharing financial data, sharing data with law enforcement, data security, data deletion, and collecting sensitive information are the most prominent. Conversely, attributes that participants believe are nearly universal (such as collecting location or on-site browsing data) or that they do not value highly rarely appear in the personalized set. This distribution characterizes the attributes for which our estimated treatment effects reflect the impact of providing information. An important implication is that we provide each participant with information on the attributes they are most likely to act on.

Finally, we examine the cross-sectional relationship between consumers' belief accuracy about

[18] We derive WTP by dividing the estimated part-worth for each attribute by the price coefficient. Appendix B contains the full estimation details.

[19] Reassuringly, preferring not to answer the question on gender is associated with the largest predicted increase in privacy preference intensity.

[20] Bian et al. (2023) document that Apple's App Tracking Transparency allowed consumers to meaningful reduce these harms through changing their data sharing practices – indicating that information on privacy practices could help consumers take action to reduce the likelihood of these consequences.

privacy practices and their preference intensity from the conjoint experiment.[21] Table C.4 reports the results. Column (1) shows that respondents who underestimate the prevalence of a practice have 0.016 higher part-worths than the correctly informed (0.04 SD of the part-worth distribution). Column (2) confirms that overestimating prevalence is associated with lower part-worths: a one standard deviation increase in misspecification corresponds to a 0.01 SD decrease in preference intensity. Column (3) shows this is driven entirely by underestimation – those who overestimate prevalence are no different from the correctly informed. In other words, the consumers who would penalize invasive practices most in their choices are precisely those who are unaware of how common those practices are.

6 Information Acquisition and Beliefs

This section evaluates Hypotheses 1 and 2 — whether and to what extent participants in the intervention engaged in information acquisition and changed their knowledge of privacy practices. The results reveal a sharp asymmetry: reducing search costs dramatically increases search behavior but produces no improvement in knowledge, whereas reducing information complexity leads to meaningful learning. We document this in two steps, first examining beliefs directly (Section 6.1) and then decomposing the underlying search and learning mechanisms (Section 6.2).

6.1 Beliefs and Knowledge of Privacy Practices for Most Frequented Sites

If participants attended to and internalized the information provided during the experiment, they would be most knowledgeable about the privacy practices of the sites they visit most often and the attributes they care most about. Accordingly, we personalize the endline survey to elicit beliefs for each participant's top two preferred attributes on their five most visited sites during the intervention period. We then compare belief accuracy across treatment conditions using the following regression:

$$|\theta^b_{ijk} - \mu^*_{jk}| = \beta T_i + \eta_j + \eta_b + \epsilon_{ijk}$$

where η_b represents fixed effects for the experimental block and η_j represents fixed effects for the website. The dependent variable measures the absolute distance between the participant's self-reported likelihood that website j engages in practice k and the truth (which is 0 or 100 by construction). We also consider binarized measures of correctness: a weak measure that excludes neutral responses of 50, and a strict measure that classifies a participant as correct if their belief exceeds 50 when (and only when) the website engages in the practice. The continuous measure

[21] Part-worths are converted so that positive values uniformly indicate disutility from the practice.

captures the magnitude of misspecification; the binary measures capture whether participants get the direction right.

Table 1: Belief Correctness on Top Sites – Provided Information

Model:	|Belief − Truth| (1)	Correctness (Weak) (2)	Correctness (Strict) (3)
Saliency Treatment	-0.3060	-0.0098	0.0076
	(0.8241)	(0.0139)	(0.0120)
Information Treatment	-4.971***	0.0667***	0.0715***
	(0.8354)	(0.0137)	(0.0116)
Fixed-effects			
Website FE	Yes	Yes	Yes
Block FE	Yes	Yes	Yes

Clustered participant standard-errors in parentheses
*Signif. Codes: ***: 0.01, **: 0.05, *: 0.1*

NOTES: OLS estimates from Equation (3). Includes website and randomization block fixed effects. Column (1): Dependent variable is $|\text{Belief}_{ijk} - \text{Truth}_{jk}|$, where beliefs are scaled to $\{0, 25, 50, 75, 100\}$ and truth is binary $\{0, 100\}$. Column (2): Binary indicator equal to 1 if response is correct, excluding neutral responses of 50. Column (3): Strict correctness, equal to 1 if $(\text{Belief} > 50) = (\text{Truth} = 100)$. The beliefs are the likelihood of the website engaging in the specified privacy practice for the top 5 most visited sites during the intervention and for the top 2 most informative attributes according to Equation (3).

Table 1 presents the results. The information treatment produces a clear improvement in belief accuracy across all specifications. Column (1) indicates that participants in the information group are 5.0 percentage points closer to the true privacy practices than those in the control group. To put this in context, the median participant in both the control and saliency groups has a belief error of 50 — indicating that even for their most visited websites and most valued attributes, participants are essentially at the midpoint of the scale. Against this baseline, a 5 pp improvement represents a meaningful shift, though substantial misspecification remains. Columns (2) and (3) use the binarized correctness measures and show that information group participants are 6.7–7.2 percentage points more likely to correctly identify a site's privacy practices.

Equally striking is the null result for the saliency group: the point estimates are close to zero and, if anything, slightly negative. Making privacy policies dramatically easier to find — as we show below, the saliency treatment increases privacy policy visits by a factor of 17 — produces no detectable improvement in knowledge about privacy practices. This result has direct implications for disclosure policy: regulatory approaches that focus on making privacy policies more accessible, without also reducing their complexity, are unlikely to meaningfully inform consumers.

We probe the robustness of these results along two dimensions. First, Table C.6 estimates the same specification for a random attribute not in the participant's top two personalized attributes. While directionally consistent — the information group shows a statistically significant improvement relative to the control group — the estimated effect size is substantially smaller. This suggests

that participants primarily internalize the highlighted information from the popup rather than seeking out additional details, but their exposure to the random information also weakly shifted their beliefs. Second, we evaluate the CDF of belief errors across treatment groups (Figure C.7). The distributions for the saliency and control groups are nearly identical, while the information group produces a visibly left-shifted and steeper CDF, implying both higher average accuracy and lower within-group dispersion.

A natural question is whether more time spent on a website leads to better knowledge of its privacy practices. We partition (website, participant) tuples within each experimental group into terciles based on baseline visit frequency and plot mean belief correctness across conditions and terciles (Figure C.6). Two patterns emerge. First, within the control and saliency groups, there is little variation across terciles — even heavy users of a website know little about its privacy practices. Second, participants in every exposure tercile of the information group are significantly more accurate than participants in any tercile of the other groups. Within the information group, accuracy increases modestly with exposure, but this within-condition gradient is small relative to the overall treatment effect. The dominant source of variation is between treatments, not within them.

6.2 Quantifying Search and Learning

The preceding results establish *what* happened: the information group learned about privacy practices while the saliency group did not. This section establishes *why*, by separately measuring search (finding and attending to information) and learning (internalizing information conditional on attending to it). This decomposition is at the heart of Hypothesis 2.

We define *search* as a visit to a privacy policy (measurable across all treatment groups) or engagement with the privacy information dialog (measurable in the information group). This definition undercounts search in the information group, since a participant may read the popup without clicking on it; this biases against finding a search effect for that group. We define *learning* using the belief measures from Section 6.1.

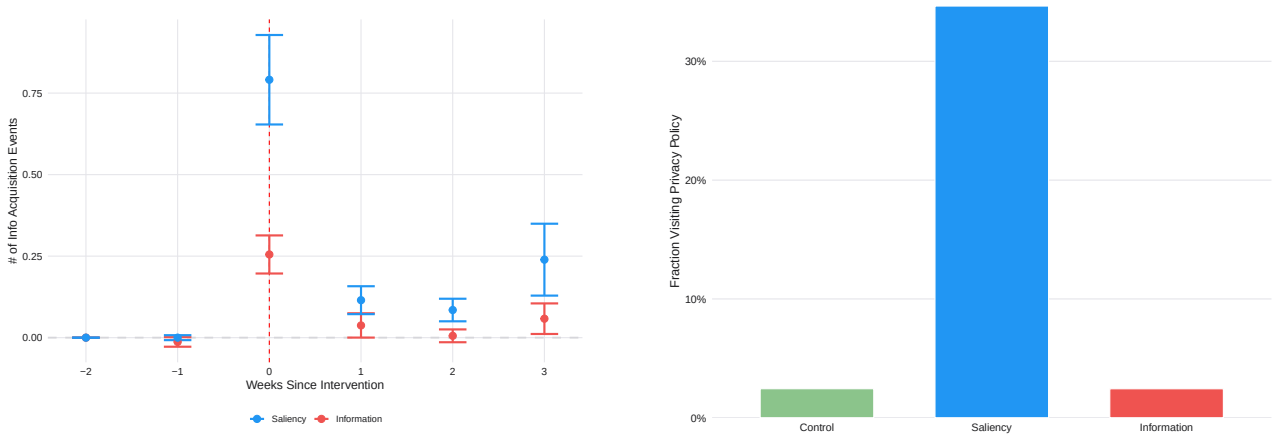
Does the Intervention Increase Search?

We test Hypothesis 2(i) using the following difference-in-differences specification:

$$y_{ijt} = \beta \cdot \Big(\mathbb{1}\{\text{Post}\} \cdot T_i \Big) + \eta_t + \eta_i + \eta_b + \epsilon_{ijt} \tag{4}$$

where y_{ijt} denotes whether participant i engaged in search on website j at time t, $\mathbb{1}\{\text{Post}\}$ is an indicator for the post baseline period, and η_t, η_b, η_i are time, experimental block, and individual

Figure 7: Treatment Effect on Search Behaviors



(a) Total Sites

(b) % Participants Viewed ≥ 1 Privacy Policy

NOTES: Panel (a): Weekly difference-in-differences estimates from Equation (5), with $T_i \in \{\text{Info}, \text{Saliency}\}$, individual fixed effects, and randomization block fixed effects. Standard errors clustered at the participant level. Time window: weeks -2 to $+3$ relative to intervention. Panel (b): presents the fraction of participants who visited at least one privacy policy following the start of the information intervention across the three different treatment groups.

fixed effects.[22]

Results are presented in Figure 7a. We find that the saliency group persistently engages in more search throughout the treatment period. In the first week of the intervention, participants in the saliency group visited 0.8 more privacy policies than the control group, with this effect being positive and ranging from 0.1 to 0.25 in the following weeks of the intervention. Figure 7a also implies that the information group engaged in additional information acquisition – beyond the presented dialogs – which typically meant clicking to see all the privacy information for a given website or viewing the extended descriptions of the privacy information.

To understand whether these effects are driven by a small number of participants, Figure 7b plots the fraction of participants in each treatment group who *ever* visit a privacy policy. Only 2% of participants in the control group do so. In the saliency group, this rises to nearly 34% — a 17-fold increase on the extensive margin. Yet even with this dramatic reduction in search costs, nearly two-thirds of saliency group participants never visit a single privacy policy. In the information group, the rate is approximately 5%, reflecting the fact that the popup already provides the key information without requiring a policy visit.[23]

[22] We estimate this regression over the time periods until the cookie-deletion intervention, since this intervention can change behavior and we do not implement it at the same time for each participant.

[23] As additional evidence of search, our extension logs engagement with cookie banners. Table C.7 reports a statistically significant increase in the number of distinct domains where saliency group participants interacted with cookie banners, providing further confirmation of Hypothesis 2(i).

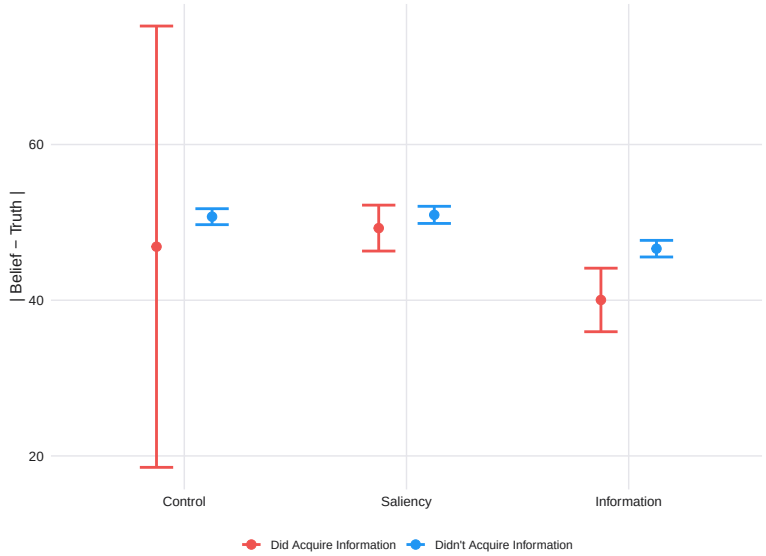


Figure 8: Beliefs by Search Behavior

NOTES: Weekly difference-in-differences estimates from Equation (5), with $T_i \in \{\text{Info}, \text{Saliency}\}$, individual fixed effects, and randomization block fixed effects. Standard errors clustered at the participant level. Time window: weeks -2 to $+3$ relative to intervention. Dependent variable is mean belief error $|\text{Belief} - \text{Truth}|$ conditional on engaging in privacy-related search (policy visit or dialog interaction).

Does Search Lead to Learning?

The critical question is whether this additional search translates into learning. Figure 8 plots mean belief error conditional on whether the participant engaged in information acquisition on each site, separately by treatment group.

Several patterns emerge. First, across all conditions, participants who engage in search have somewhat more accurate beliefs about the website's privacy practices than those who do not. Second, and most importantly, this difference is small and statistically insignificant for the saliency group: participants who searched for and read actual privacy policies are no more knowledgeable than those who did not. This result is consistent with privacy policy complexity being the binding constraint — even motivated searchers cannot extract useful information from these documents.

Within the information group, participants who engaged in additional search beyond the popup have more accurate beliefs than those who relied on the popup alone. This holds even relative to the information group's higher baseline accuracy: non-searchers in the information group are more knowledgeable than searchers in the saliency group, consistent with the popup transmitting information even to participants who do not actively engage with it.

We note an important caveat: we are underpowered to precisely estimate the search–learning relationship in the control group, since very few control participants seek out privacy policies, which results in wide confidence intervals. This limits our ability to formally test Hypothesis 2(ii). However, the joint pattern across all three groups — saliency participants search more but learn no

more than control participants, while information participants learn substantially more whether or not they search — provides strong evidence for Hypothesis 2(iii) and for complexity as the binding friction.

The results of this section deliver a clean decomposition of the two information frictions in our model. Reducing search costs produces a dramatic increase in information-seeking behavior: the saliency treatment increases privacy policy visits by a factor of 17. Yet this search is almost entirely unproductive — saliency group participants are no better informed about the privacy practices of their most visited websites than control group participants. In contrast, reducing complexity through simplified, personalized disclosure produces meaningful improvements in belief accuracy, even among participants who do not actively engage with the information. The central lesson is that the binding constraint on privacy information acquisition is not access but comprehension: participants can find privacy policies but cannot learn from them.

7 Effect on Data Sharing and Website Choices

The results of Section 6 establish that the information treatment successfully transmits knowledge about privacy practices. The central question is whether this learning translates into changes in behavior. We examine two margins: data sharing (Section 7.1) and website choice (Section 7.2). We then corroborate these objective measures with self-reported behavioral changes from the endline survey (Section 7.3).

Before presenting results, it is useful to set expectations about plausible magnitudes. Our conjoint estimates (Figure 6a) imply that moving from a website that infringes on all three of a participant's most valued attributes to one that does not is worth approximately $3 per month at the median. Meanwhile, the most popular websites in our data — YouTube, Google, Amazon, Facebook — have daily time-spent averages between 0.4 and 0.9 hours. Given the survey compensation rate of $3 for 15 minutes of work, this implies a lower bound on the valuation of these sites of at least $8.[24] This crude approximation illustrates a significant gap: for the most frequently visited websites, quality-based valuations substantially exceed privacy costs. This means that large behavioral responses on the intensive margin — participants abandoning their most-used sites — are unlikely even if the information treatment is fully effective. The more plausible margin of adjustment is on the extensive margin, where participants shift among less-visited or marginal sites.

[24] It additionally dramatically dwarfs the overall valuation for these sites discussed in Brynjolfsson et al. (2019).

7.1 Data Sharing

We first examine whether participants change their data sharing practices in response to the intervention. Table C.9 presents treatment effects on four self-reported measures from the endline survey: whether a participant deleted cookies, explicitly opted out of data sharing, withheld data due to privacy concerns, or changed privacy settings on at least one website.[25]

The one statistically significant result is that participants in the information group are 6.9 percentage points more likely to have opted out of data sharing on at least one website (column 2, $p < 0.10$), consistent with a 20% increase relative to the control group mean. Point estimates for the remaining measures are positive for both treatment groups but not statistically distinguishable from zero. The saliency group shows no significant effects on any measure.

Overall, there is weak evidence of a small change in data sharing behavior in the information group, concentrated in the opt-out decision, and no detectable response in the saliency group. The limited behavioral response likely reflects the high costs of changing data sharing practices relative to the modest privacy valuations documented in Section 5.3: even informed participants may rationally choose not to incur the hassle of modifying settings on individual websites.

7.2 Website Choice and Usage

We now examine whether changes in beliefs lead to changes in the websites participants visit and how they allocate their time. Our model implies that the direction and magnitude of behavioral effects depend on beliefs (θ^b), the intensity of privacy preferences (λ^P), quality differences across websites within a market, and the distribution of privacy practices. Consequently, aggregate effects are theoretically ambiguous and reduced-form magnitudes are difficult to interpret. We present results for both the intensive and extensive margins and interpret them in light of these considerations.

7.2.1 Intensive Margin

We first ask: holding the set of sites fixed, how does browsing behavior change? We construct a balanced panel for each participant across their full set of pre-treatment visited sites and estimate:

$$y_{ijt} = \beta \cdot (\mathbf{1}\{\text{Post}\} \cdot T_i) + \eta_j + \eta_i + \epsilon_{ijt} \tag{5}$$

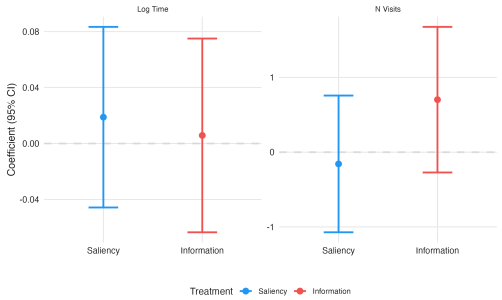
where η_j and η_i are site and participant fixed effects.

Figure 9a presents results for log time spent and the number of visits. Across both dependent variables, there is no evidence of changes in the intensive margin of browsing behavior. Impor-

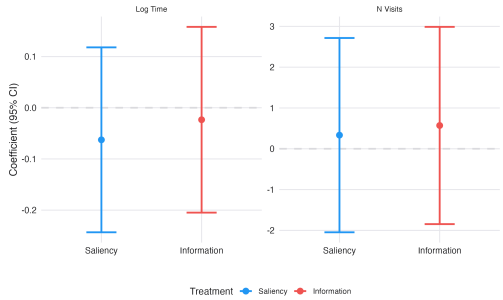
[25] This question was only added for the second wave of participants.

tantly, our estimates are precise: the largest effects within our confidence intervals would imply changes of less than one minute in time spent or two additional visits over a two-week period — less than 0.1% of average time spent online.

Figure 9: Browsing Behavior Treatment Effects: Intensive Margin



(a) Difference-in-differences ATE



(b) Triple Difference-in-differences: Heterogeneous ATE by Privacy Score

NOTES: Panel (a): Difference-in-differences estimates from Equation (6). Outcomes: log time spent and visit count for individual i on website j in period t. Includes website and individual fixed effects. Sample restricted to balanced panel of websites visited in both baseline weeks ($t \in \{-2, -1\}$). Standard errors clustered at the participant level. Panel (b): Triple-difference estimates with privacy heterogeneity from Equations (7)–(8). Privacy score P_{ij} computed per Equation (4), weighting site-level privacy attributes by individual i's conjoint coefficients for personalized attributes $k \in \{1, 2\}$. Outcomes winsorized at 5th and 95th percentiles.

This null is unsurprising given the valuation gap discussed above. The sample used here consists primarily of sites that participants spend significant time on and thus likely have strong preferences for. Observing large changes in these regressions would require participants to move away from their highest-value sites, which our WTP estimates suggest is unlikely.

Our model further implies that even if aggregate time on a site does not change, how participants allocate time *across* sites may shift. To test this, we construct site-user-level privacy scores by weighting site-level privacy attributes by individual-level conjoint utility estimates. We then estimate a heterogeneous treatment effect specification:

$$y_{ijt} = \beta(\mathbf{1}\{\text{Post}\} \cdot T_i \cdot P_{ij}) + \eta_j + \eta_t + \eta_i + \epsilon_{ijt} \tag{6}$$

where P_{ij} is our individual measure of privacy invasiveness. Figure 9b presents the results and mirrors the aggregate findings: there is no evidence that participants differentially reallocate time toward or away from sites based on their privacy characteristics.

7.2.2 Extensive Margin

Given the valuation gap, the more plausible margin of adjustment is the extensive margin: how participants change the *set* of sites they visit, particularly among less-visited or marginal sites. We

Figure 10: Browsing Behavior Treatment Effects: Extensive Margin



(a) Assortment Changes within Category

(b) Experienced Privacy within Category

NOTES: First-difference estimates from Equation (9), where $\Delta Y_{ic} \equiv Y_{ic,\text{post}} - Y_{ic,\text{pre}}$ for individual i in website category c. Includes category fixed effects. Standard errors clustered at the participant level. Panel (a): Concentration measures. HHI computed as $\sum_j s_{ijc}^2$, where s_{ijc} denotes individual i's time share on website j within category c. Top-site share defined as $\max_j s_{ijc}$. Panel (b): Privacy measures. Sites classified as new if $s_{ijc,\text{pre}} = 0$ and $s_{ijc,\text{post}} > 0$; dropped if the reverse. Privacy differential computed as $\bar{P}_{i,\text{new}} - \bar{P}_{i,\text{dropped}}$, averaging site-level privacy scores within each group. Time-weighted privacy score computed as $\sum_j s_{ijc} \cdot P_{ij}$.

assess this at the participant–category level using:

$$Y_{ic}^{\text{post}} - Y_{ic}^{\text{pre}} = \beta \cdot T_i + \eta_c + \varepsilon_{ic} \tag{7}$$

where the dependent variable is the within-user pre–post difference and η_c is a category-level fixed effect.

Figure 10a presents results for two measures of browsing concentration: the Herfindahl–Hirschman Index (HHI) at the user–category level and the share of time spent on a participant's top 2 sites. Both regressions show small decreases in concentration for the information group, with p-values of 0.063 (HHI) and 0.065 (top-2 share). While only marginally significant, the direction is consistent with participants diversifying their browsing in response to the information treatment.

Figure 10b examines whether this diversification is privacy-motivated. We construct two measures. First, the *privacy differential* between newly visited and dropped sites, capturing whether sites that participants begin visiting have better privacy practices than the sites they stop visiting. Second, the change in the average privacy score of each participant's portfolio of websites. The information group shows positive effects on both measures: the treatment accounts for approximately 11% of the standard deviation in the privacy differential and 3% in the portfolio privacy score. These are modest magnitudes, but they indicate that the direction of extensive-margin reallocation is toward more privacy-conscious sites.

7.3 Self-Reported Behavioral Changes

We complement the objective measures above with qualitative evidence from free-form responses in the endline survey (Haaland et al., 2025). Participants were asked: "Has this experiment changed your view on privacy online? Have you changed any online behaviors in response to this experiment?" We recruit MTurk raters to classify each response into four categories: behavioral change (site choice or data sharing), heightened awareness, fatalism (resignation that data collection is inevitable), and no change.[26]

Figure C.8 presents the results. The information group shows elevated rates of both behavioral change and heightened awareness relative to the control group, consistent with the objective evidence of modest behavioral shifts. The saliency group is statistically indistinguishable from the control group on all dimensions and, if anything, shows higher rates of fatalism — consistent with the complexity mechanism documented in Section 6.2. Participants who were directed to privacy policies but could not extract useful information appear to have concluded that privacy management is futile.[27]

Cumulatively, the results of this section paint a consistent picture. Participants in the information group make small but detectable adjustments: they are somewhat more likely to opt out of data sharing and they modestly shift their browsing toward more privacy-conscious sites on the extensive margin. However, the intensive margin of browsing — where participants spend the bulk of their time — is essentially unchanged. This pattern is consistent with the valuation gap between website quality and privacy costs: for the sites that dominate participants' browsing, quality-based preferences are simply too strong for privacy information, even when personalized and simplified, to meaningfully alter behavior.

These reduced-form results, while informative about the direction and approximate magnitude of behavioral responses, face important limitations. The theoretical ambiguity noted at the outset of this section — that aggregate effects depend on the joint distribution of beliefs, preferences, quality, and privacy practices — means that reduced-form estimates are difficult to map into welfare-relevant quantities. Moreover, the reduced-form approach cannot separately identify how much of the limited behavioral response reflects low privacy valuations versus high search or comprehension costs that persist even under treatment. Disentangling these forces requires a structural model that jointly estimates preferences, information frictions, and the value of website quality. We develop such a model in the next section.

[26] Three raters coded each response. In cases where the majority coding was tied, the researchers broke ties based on their own interpretation.

[27] One participant in the saliency group captured this sentiment: "I would click on the data privacy link, and my eyes would just immediately glaze over at the length and legalese, and I just go into every website with the expectation that I will have no power over my data."

8 Empirical Model

We consider the empirical analog of the model detailed in Section 3.

Flow Utility. The model in Section 3 supposes that choices rely on quality and privacy components, which we characterize separately. We parametrize the quality component as follows:

$$\lambda_i\big(\bar{Q}_j + \delta_j D_{it}\big)$$

where $\bar{Q}_j$ represents the latent website quality, $D_{it} \in \{0,1\}$ indicates whether consumer i's cookies are active on day t, δ_j captures the cookie's contribution to site value, and λ_i is a type-specific quality weight. This parametrization decouples the aspects of website quality that are impacted by data sharing choices so that it is clear which aspects are impacted by our cookie-blocking intervention.

The privacy component of consumer utility includes the preferences that she has over the different attributes. We suppose that she has prior beliefs about each site's vector of K binary privacy practices $\mu_j^* \in \{0,1\}^K$, denoted $\theta^0_{ijk} \equiv \Pr(\mu^*_{jk} = 1 \mid \mathcal{I}_{it})$. We consider that the consumer is risk-averse and has CARA preferences, so that the the certainty-equivalent of the privacy component is given by:

$$\text{CE}_{it}(\boldsymbol{\theta}_{ijt}, \boldsymbol{\beta}_i, \gamma_i) = \sum_{k=1}^{K} \beta_{ik}\, \theta_{ijkt} - \frac{\gamma_i}{2} \sum_{k=1}^{K} \beta_{ik}^2\, \theta_{ijkt}(1-\theta_{ijkt}) \tag{8}$$

where β_{ik} is consumer i's preference weight for attribute k and γ_i is risk aversion. The second term penalizes uncertainty about privacy practices and generates the incentive to acquire information.

Put together, the flow utility that consumer i receives from visiting website j on day t is given by the combination of these two:

$$u_{ijt} = \lambda_i\big(\bar{Q}_j + \delta_j D_{it}\big) + \text{CE}_{it}(\boldsymbol{\theta}_{ijt}, \boldsymbol{\beta}_i, \gamma_i) \tag{9}$$

We consider a nested logit formulation based on the categorization from Klazify used in Section 5. Each day, consumer i allocates browsing time across J sites plus an outside option ($j = 0$), with s_{ijt} denoting the share of time spent on site j on day t. Given i.i.d. Type I Extreme Value taste shocks ε_{ijt}, the nested logit yields choice probabilities

$$s_{ijt} = \frac{\exp\big(u_{ijt}/(1-\sigma)\big)}{\big(\sum_{j' \in n(j)} \exp(u_{ij't}/(1-\sigma))\big)^{\sigma}} \cdot \frac{\big(\sum_{j' \in n(j)} \exp(u_{ij't}/(1-\sigma))\big)}{\sum_n \big(\sum_{j' \in n} \exp(u_{ij't}/(1-\sigma))\big)^{1-\sigma}} \tag{10}$$

where $\sigma \in [0,1)$ governs within-nest substitution. Following Berry (1994), this inverts to give the

log relative share:

$$\ln s_{ijt} - \ln s_{i0t} = u_{ijt} + \sigma \ln \bar{s}_{j|n(j),it} \tag{11}$$

where $\bar{s}_{j|n(j),it}$ is the within-nest share.

Information Acquisition. After choosing which website to use, consumers then decide whether to acquire information on the privacy practices of that site. Consumers are forward-looking and value resolving uncertainty about privacy practices.

For notation we denote $\Omega_{it} = (\Theta_{it}, h_{it})$ as the consumer i's state at time t, comprising the full belief matrix $\Theta_{it} \in [0,1]^{J \times K}$ and visit history h_{it}. The consumer's dynamic program is

$$V(\Omega_{it}) = \mathbb{E}_{\varepsilon}\Big[\max_{j}\Big\{u_{ijt} + \varepsilon_{ijt} + \beta^d \, \mathbb{E}\big[V(\Omega_{i,t+1}) \mid j, \Omega_{it}\big]\Big\}\Big] \tag{12}$$

where β^d is the daily discount factor. Upon visiting site j the consumer's visit history updates, and they are presented with a choice of whether or not to engage in search for privacy information. Search is costly with search cost c_s^T depending on the treatment condition, T. If they choose to search for information beliefs are updated according to:

$$\theta_{ijk}^{\text{post}} = \theta_{ijk}^0 + \alpha_i^T \cdot (\mu_{jk}^* - \theta_{ijk}^0) \tag{13}$$

where we allow the comprehension rate $\alpha_g^T \in (0,1)$ to depend on the treatment condition T. Then, we can write the value of information about site j as the gain in expected lifetime utility from learning j's privacy practices:

$$\text{VOI}_{ij}(\Omega_{it}) = V(\Omega_{it}^{\text{informed}}) - V(\Omega_{it}) \tag{14}$$

where $\Omega_{it}^{\text{informed}}$ denotes the state in which the consumer is informed about site j's privacy attributes.

Taking expectations over the unknown truth μ_{jk}^* under the prior θ_{ijk}^0, the CARA–Bernoulli structure yields a closed-form expected per-period flow gain from search:

$$\mathbb{E}[\Delta u_{ij}^{\text{search}}] = \frac{\gamma_i}{2} \cdot \alpha_g^T \cdot \sum_{k=1}^{K} \beta_{ik}^2 \, \theta_{ijk}^0 (1 - \theta_{ijk}^0) \tag{15}$$

The flow gain is nonnegative and increasing in the preference-weighted belief variance: consumers value information most about sites where they face greater uncertainty over attributes they care about. The full VOI_{ij} aggregates this flow gain across all future visits to site j, weighted by the discount factor and the probability of visiting in each period.

To smooth the discrete search decision, we add a normally distributed taste shock $\varepsilon_{ij}^s \sim$

$N(0,1)$. Consumer i searches on site j if the expected value of information exceeds the cost:

$$\mathbb{E}[\text{VOI}_{ij}] > c_s^T + \varepsilon_{ij}^s \tag{16}$$

9 Estimation

We estimate our model via two-step GMM.

9.1 Parameterization

The model in Section 8 is defined over a rich state space: each consumer's beliefs form a $J \times K$ matrix of continuous probabilities, preferences and comprehension rates vary continuously across individuals, and consumers solve an infinite-horizon dynamic program each period. Direct estimation would require evaluating this dynamic program over a high-dimensional continuous state for each candidate parameter vector—an intractable problem with $J = 489$ sites and $K = 19$ attributes.[28] We introduce a sequence of parameterizations that reduce the dimensionality of heterogeneity, the choice set, and the belief state while preserving the model's key economic forces.

9.1.1 Type Classification

Following Bonhomme et al. (2022), we discretize the distribution of unobserved consumer heterogeneity into a finite number of types. Each consumer i belongs to a type $g_i \in \{1, \ldots, G\}$ with population share π_g. Types govern preferences (β_{ik}, λ_i) and information processing rates (α_i^{popup}, α_i^{search}) which are indexed by type g instead of consumer i. This reduces the preference space from $N \times K$ individual-level parameters to $G \times K$ type-level parameters, making GMM estimation feasible while providing a flexible nonparametric approximation to the continuous heterogeneity distribution.

We classify consumers using a Gaussian mixture model on PCA-reduced baseline survey data: elicited beliefs $(\theta_{i1}^b, \ldots, \theta_{iK}^b)$ about each privacy attribute and risk aversion γ_i. The number of types G is selected by BIC over $\{3, 4, 5, 6, 7\}$, yielding $G = 4$ in our preferred specification. Type assignments are treated as data in subsequent estimation stages, following the two-step approach in Bonhomme et al. (2022). Because our classification uses only pre-treatment survey variables—not browsing outcomes—the type assignments are orthogonal to the treatment variation that identifies preferences and comprehension.

[28] The choice set comprises 488 individually characterized websites from the top 500 by aggregate browsing time in the sample. All remaining websites are aggregated into a single composite option with no privacy characteristics, representing non-top-500 browsing. This composite option is distinct from the outside good ($j = 0$), which captures time not spent browsing.

9.1.2 Beliefs and Information

Consumer i holds population-level beliefs θ^b_{ik} about each attribute k, elicited from the baseline survey. For a site j in content category (nest) n, category-specific shifts adjust the prior:

$$\theta^0_{ijk} = \text{clamp}\big(\theta^b_{ik} + \eta_{nk},\ 0,\ 1\big) \tag{17}$$

where η_{nk} captures category-level differences in privacy practices (e.g., shopping sites track users more than education sites).

In our experiment, information arrives through two channels. First, any consumer can read site j's privacy policy at cost c^T_s with comprehension rate $\alpha^{\text{saliency}}_g$ over all K attributes. Second, when a participant in the information group visits site j, a privacy popup fires with comprehension rate α^{info}_g over the experimentally exposed attributes $\mathcal{K}_i$. Search costs are defined by $c^T_s = c_s - \Delta c_s \cdot \mathbf{1}\{\text{Treated}_i\}$, where Δc_s captures the treatment-induced cost reduction.

9.1.3 Cookie Value

The model allows cookies to contribute to site quality through δ_j in equation (1). In practice, the value a cookie provides depends on what the site does with the tracking data it collects. We parameterize this as

$$\delta_j = \delta_0 + \delta_1 \cdot \text{NumTrackers}_j + \delta_2 \cdot \text{Personalizes}_j + \delta_3 \cdot \text{ShareAds}_j \tag{18}$$

where $\delta_0 = 1$ by normalization. NumTrackers measures the extent of third-party tracking; Personalizes indicates whether the site uses cookies to customize content; and ShareAds indicates whether tracking data is shared with advertisers. This decomposition allows cookies to have both positive effects (personalization) and negative effects (invasive tracking), with the net value varying across sites. The cookie-deletion period at the end of our experiment provides exogenous variation in D_{it} to identify these parameters.

9.1.4 State Space Reduction

The consumer's state $\Omega_{it} = (\Theta_{it}, h_{it})$ includes a $J \times K$ matrix of continuous beliefs—intractable for direct computation. Three structural features yield a finite-state formulation.

First, because each privacy attribute μ^*_{jk} is binary, the belief θ_{ijkt} is a scalar probability, and the certainty equivalent (8) depends on beliefs only through two sufficient statistics: the privacy index $\sum_k \beta_{gk}\theta_{ijkt}$ and the preference-weighted belief variance $\sum_k \beta^2_{gk}\theta_{ijkt}(1-\theta_{ijkt})$. Furthermore, because we restrict prior beliefs θ^b_{ik} to be common across sites (with nest-specific shifts), the variance

governing the value of information is effectively a consumer-level quantity before any site-specific learning occurs.

Following this, two state variables remain: a visit stock $s \in 0, \ldots, 5$ (6 bins) tracking cumulative visits to site j, and an information level $\ell \in 0, 1, 2$ indicating whether the consumer is uninformed, popup-informed, or search-informed about the site's privacy practices. The Bellman recursion therefore requires only $6 \times 3 = 18$ states per (g, j) pair rather than a continuous belief grid.

Finally, under stationarity, the site-specific value function satisfies

$$W(s, \ell) = u(s, \ell) + \beta^d \big[P_\ell(s) \cdot W(s{+}1, \ell) + (1 - P_\ell(s)) \cdot W(s, \ell) \big] \tag{19}$$

where $P_\ell(s)$ is the CCP at state (s, ℓ) and $\beta^d = 0.997$. Rearranging yields a closed-form recursion over visit-count bins:

$$W(s, \ell) = \frac{u(s, \ell) + \beta^d \, P_\ell(s) \, W(s{+}1, \ell)}{1 - \beta^d + \beta^d \, P_\ell(s)},$$

with absorbing boundary $W(S{-}1, \ell) = u(S{-}1, \ell)/(1 - \beta^d)$. We solve from the top bin downward through $s = S{-}2, \ldots, 0$, obtaining a closed-form expression at each step. Because the per-period flow gain from information (15) is constant conditional on the information level, the expected value of information factors as

$$\mathbb{E}[\text{VOI}_{ij}(s)] = \frac{\gamma_i}{2} \cdot \alpha_g^{\text{salience}} \cdot \sum_{k=1}^{K} \beta_{gk}^2 \, \theta_{ijk}^0 (1 - \theta_{ijk}^0) \cdot M(s) \tag{20}$$

where $M(s)$ is the expected discounted visit count starting from bin s, defined by

$$M(s) = \frac{P(s)\big[1 + \beta^d \, M(s{+}1)\big]}{1 - \beta^d + \beta^d \, P(s)}, \qquad M(S{-}1) = \frac{P(S{-}1)}{1 - \beta^d}.$$

This dynamic multiplier converts the per-period flow gain into a present value, weighting each future period by both the discount factor and the probability of visiting site j.

9.2 GMM Estimation

Given these parameterizations, the full parameter vector is

$$\theta = \Big(\underbrace{\beta_{gk}}_{G \times K},\ \sigma,\ \delta_1,\ \delta_2,\ \delta_3,\ c_s,\ \Delta c_s,\ \underbrace{\alpha_g^{\text{info}},\ \alpha_g^{\text{salience}}}_{2G},\ \underbrace{\xi_g}_{G} \Big)$$

where ξ is a stated-to-revealed preference scaling parameter (defined below). With $G = 4$ and $K = 19$, the model has 94 free parameters. The type-specific quality weights λ_g are concentrated analytically from the cookie-deletion moments and do not enter the optimization. We estimate θ by two-step GMM with six blocks of moment conditions (171 moments total). Site quality $\bar{Q}_j$ is recovered via Berry inversion of pre-treatment log relative shares. We normalize $\bar{Q}_j$ so that the model-implied outside share matches the empirical daily outside share of 41%. This shifts $\bar{Q}$ down while leaving all GMM-estimated parameters unchanged, since these are identified from DiD and IV moment conditions independent of $\bar{Q}$. The baseline inclusive value is $\bar{V} = 0.81$.

Block A: Search Rates ($G + 1 = 5$ moments)

Among control and saliency consumers, the model predicts the search probability $P_i^{\text{search}} = \Phi(\text{VOI}_i - c_s + \Delta c_s \cdot \mathbf{1}[\text{saliency}_i])$, where VOI_i depends on α_g^{search}, β_g, and M. The per-type mean residuals and the saliency interaction identify $(c_s, \Delta c_s, \alpha_g^{\text{salience}})$.

Block B: Popup-Event Difference-in-Differences ($G \times K = 76$ moments)

For each information group consumer i and each first-visit site j during the treatment period, we construct the popup-event DiD $\Delta_{ij}^{\text{popup}} = \bar{y}_{ij,\text{post}} - \bar{y}_{ij,\text{pre}}$ in a 7 day window. The model predicts

$$\mathbb{E}[\Delta_{ij}^{\text{popup}}] = (1 - P_{ij}^{\text{search}}) \cdot M_{ij} \cdot \Big[\sum_{k \in \mathcal{K}_i} \beta_{g_i k} \, \alpha_g^{\text{info}} \, (\mu_{jk}^* - \theta_{ijk}^0) - \frac{\gamma_i}{2} \sum_{k \in \mathcal{K}_i} \beta_{g_i k}^2 \, w_{ijk} \Big] + \sigma \cdot \Delta \ln \bar{s}_{j|n(j)} \quad (21)$$

The $G \times K$ moments interact the residual with $Z_{ijk}^A = (\mu_{jk}^* - \theta_{ijk}^0) \cdot \mathbf{1}[k \in \mathcal{K}_i]$, the potential belief shock at full comprehension.

Block C: Endline Beliefs ($2G = 8$ moments)

The endline survey measures belief accuracy by treatment arm. For each type g, the cross-arm difference in the realized learning rate $\rho_{ijk} = (\theta_{ijk}^{\text{endline}} - \theta_{ik}^b)/(\mu_{jk}^* - \theta_{ik}^b)$ directly identifies the comprehension parameters:

$$\alpha_g^{\text{info}} = \mathbb{E}[\rho_{ijk} \mid T_i = 2, g] - \mathbb{E}[\rho_{ijk} \mid T_i = 0, g] \quad (22)$$

$$\alpha_g^{\text{salience}} = \frac{\mathbb{E}[\rho_{ijk} \mid T_i = 1, g] - \mathbb{E}[\rho_{ijk} \mid T_i = 0, g]}{P(\text{search} \mid T_i = 1, g)} \quad (23)$$

Block D: Cookie-Deletion Instruments (4 moments)

The cookie-deletion experiment provides exogenous variation in D_{it} (cookie status) that is orthogonal to privacy preferences. For consumers whose cookies are deleted, we observe the DiD:

$$\Delta_{ij}^{\text{cookie}} = \bar{y}_{ij,\text{post}} - \bar{y}_{ij,\text{pre}} \tag{24}$$

The model predicts $\Delta_{ij}^{\text{cookie}} = -\lambda_{g_i} \cdot \delta_j + \sigma \cdot \Delta \ln \bar{s}_{j|n}$, where the second term is the nested logit correction. Rather than optimizing λ_g directly, we concentrate it out analytically at each GMM iteration:

$$\hat{\lambda}_g(\delta) = -\frac{\sum_{i:g_i=g} \Delta_{ij}^{\text{cookie}}}{\sum_{i:g_i=g} \delta_j} \tag{25}$$

This exactly satisfies the type-mean moment conditions by construction. Four remaining instrument moments identify δ_1, δ_2, and δ_3:

$$\mathbb{E}\Big[\Big(\Delta_{ij}^{\text{cookie}} + \hat{\lambda}_g \delta_j - \sigma \Delta \ln \bar{s}\Big) \cdot Z_{ij}^{\ell}\Big] = 0, \quad \ell = 1, 2, 3, 4 \tag{26}$$

where $Z^1 = \text{NumTrackers}_j$, $Z^2 = \text{Personalizes}_j$, $Z^3 = \text{ShareAds}_j$, and $Z^4 = \mathbf{1}\{\text{early deletion group}\}$. The staggered timing of cookie deletion (early vs. late groups) provides an additional moment.

Block E: Nesting Parameter (2 moments)

The nesting parameter $\sigma \in [0, 1)$ governs within-nest substitution in the nested logit demand model. It is identified from pre-treatment variation in within-nest shares, following the BLP instrument strategy. After absorbing consumer and site fixed effects via Frisch–Waugh double-demeaning, the moment conditions are:

$$\mathbb{E}\big[\big(y_{ij}^{dd} - \sigma \cdot \ln \bar{s}_{j|n(j),i}^{dd}\big) \cdot Z_{ij}^{\ell}\big] = 0, \quad \ell = 1, 2 \tag{27}$$

where y^{dd} and $\ln \bar{s}^{dd}$ are double-demeaned log relative shares and within-nest shares, and the instruments are $Z^1 = \gamma_i \cdot |\text{Nest}_{n(j)}|$ (nest size $\times$ risk aversion) and $Z^2 = \gamma_i \cdot \overline{\text{Trackers}}_{n(j)}$ (within-nest mean tracker count $\times$ risk aversion). The interaction with consumer-level γ_i generates (i, j)-level variation that survives double-demeaning.

Block F: Conjoint Validation ($G \times K = 76$ moments)

We leverage our conjoint experiment to help identify β_g. Block F links these stated preferences to the revealed-preference parameters via a single scaling parameter:

$$m_{gk}^{F} = \frac{1}{N_g} \sum_{i:g_i=g} \left(\beta_{ik}^{\text{conj}} - \xi \cdot \beta_{gk} \right) = 0 \tag{28}$$

This block serves three purposes. First, the conjoint provides a dollar-denominated scale for privacy preferences, allowing us to express the β_{gk} in willingness-to-pay terms rather than in the abstract units of the browsing utility. Second, identifying all $G \times K = 76$ attribute-specific preference parameters from behavioral variation alone is a substantial power challenge. Block B identifies the product $\alpha_g^{\text{info}} \cdot \beta_{gk}$, and while Block C pins α_g^{info} independently, the remaining cross-attribute variation within a type is limited—particularly for attributes with low prevalence or small belief gaps $(\mu_{jk}^{*} - \theta_{ijk}^{0})$. The conjoint moments anchor the attribute-level pattern, with ξ absorbing any uniform scaling difference between stated and revealed magnitudes. Third, the wedge ξ is itself a first-order object of interest: it quantifies the gap between what consumers say they care about in a hypothetical setting and how they actually behave. When $\xi > 1$, stated valuations exceed revealed valuations, consistent with hypothetical bias documented across environmental, health, and public goods domains.

9.3 Identification

Identification relies on the experimental design and standard distributional assumptions. Random assignment of treatment arms, cookie-deletion timing, and the exposed attribute set $\mathcal{K}_i$ ensures that the variation used in Blocks A–D is orthogonal to preferences and unobserved heterogeneity. The logit error structure (i.i.d. Type I Extreme Value) delivers closed-form choice probabilities and the Berry inversion, while the normal search cost shock generates the probit search probability in Block A.

The key parameters are identified from distinct sources of experimental variation. The comprehension rates α_g^{info} and $\alpha_g^{\text{salience}}$ are point-identified from Block C: the information–control difference in endline belief accuracy identifies α_g^{info} directly, while the saliency–control difference, scaled by the search probability, identifies $\alpha_g^{\text{salience}}$. Given these comprehension rates, Block B separately identifies the privacy preferences β_{gk} through the browsing response to exogenous popup-induced belief shocks—the popup fires on the consumer's first treatment-period visit to each site, and under logit errors this timing depends only on i.i.d. draws, so pre/post differencing within (i, j) pairs removes time-invariant heterogeneity. Block F anchors the attribute-level scale of β_{gk} through ξ. The cookie-value parameters $(\delta_1, \delta_2, \delta_3)$ are identified from cross-sectional variation in site char-

acteristics among cookie-deletion episodes (Block D), with λ_g concentrated out analytically. The nesting parameter σ is identified from BLP-style instruments exploiting pre-treatment within-nest share variation (Block E), and the search costs $(c_s, \Delta c_s)$ from the search-rate differential across experimental arms (Block A).

9.4 Results

Tables 2 and 3 report our preferred estimates under nested priors and $G = 4$ types (BIC-selected).

Table 2: Parameter Estimates

Parameter	Symbol	Estimate
Nesting parameter	σ	0.404
Cookie value (trackers)	δ_1	-0.0006
Cookie value (personalization)	δ_2	0.033
Cookie value (ad-sharing)	δ_3	0.013
Stated-to-revealed scaling	$\bar{\xi}$	1.34
Baseline search cost	c_s	4.92
Treatment cost reduction	Δc_s	3.89

The nesting parameter $\hat{\sigma} = 0.404$ implies moderate within-nest correlation in unobserved tastes: the within-nest elasticity of substitution is $1/(1-\hat{\sigma}) = 1.68$, roughly 1.7 times the between-nest elasticity. Consumers substitute more readily among sites within the same content category (e.g., between news sites) than across categories (e.g., from news to shopping). This attenuates the between-nest reallocation response to privacy information, since learning that one's preferred news site has poor privacy practices shifts demand primarily toward other news sites rather than toward unrelated categories with better privacy records. Personalization ($\hat{\delta}_2 = 0.033$) contributes positively to the cookie-dependent quality premium, while the raw tracker count ($\hat{\delta}_1 = -0.0006$) and ad-sharing ($\hat{\delta}_3 = 0.013$) are small. Cookies are valued primarily for the personalization services they enable rather than for the tracking infrastructure itself. The magnitudes are modest relative to site quality and privacy disutility.

We estimate a baseline search cost $\hat{c}_s = 4.92$ which is large relative to the value of information. At estimated comprehension rates, the expected per-period flow gain from searching a site is $(\gamma_i/2) \cdot \hat{\alpha}_g^{\text{salience}} \cdot \sum_k \hat{\beta}_{gk}^2 \theta_{ik}(1-\theta_{ik})$ is between 0.01 and 0.05 for the median consumer—two orders of magnitude below $\hat{c}_s$. The treatment cost reduction $\Delta\hat{c}_s = 3.89$ implies the saliency intervention dramatically reduces effective search costs, yet this remains insufficient to generate substantial voluntary search.

The estimated privacy attribute preferences, $\hat{\beta}_{gk}$, exhibit similar qualitative patterns in the relative importance of attributes as the conjoint estimates discussed in Section 5.3. While the relative

ordering is the same, the magnitudes are different. The scaling parameter $\hat{\xi}$ ranges from 1.31 to 1.36 across types (mean 1.34), indicating that conjoint-stated privacy preferences overstate revealed behavioral preferences by approximately 34%. This magnitude is consistent with hypothetical bias in the environmental and health economics literatures, where stated valuations typically exceed revealed valuations by factors of 1.3–3.0 (Murphy et al., 2005; List and Gallet, 2001).

Table 3: Type-Specific Parameters

Type (g)	$\hat{\alpha}_g^{\text{info}}$	$\hat{\alpha}_g^{\text{salience}}$	$\hat{\xi}_g$	$\hat{\lambda}_g$	Share
1	0.051	0.130	1.316	0.356	16.1%
2	0.030	0.057	1.311	0.205	15.2%
3	0.005	0.005	1.357	0.342	23.0%
4	0.086	0.008	1.361	0.325	45.7%

Table 3 reveals substantial heterogeneity in comprehension. Type 4, the largest group (45.7%), has the highest popup comprehension ($\hat{\alpha}_4^{\text{info}} = 0.086$), absorbing roughly 8.6% of the information delivered. Type 3 effectively ignores popup information ($\hat{\alpha}_3^{\text{info}} = 0.005$, at the lower bound), reflecting negative information-treatment gains in endline beliefs for this type—control-group consumers learned more than treated consumers. Saliency comprehension varies more across types: Type 1 has the highest ($\hat{\alpha}_1^{\text{salience}} = 0.130$), reflecting genuinely higher search activity in the saliency arm, while Type 2 also shows meaningful saliency comprehension ($\hat{\alpha}_2^{\text{salience}} = 0.057$). Types 3 and 4 remain near zero ($\hat{\alpha}_g^{\text{salience}} = 0.005$–$0.008$), implying that voluntary privacy policy reading generates almost no usable learning for these groups. The quality weights $\hat{\lambda}_g$ range from 0.205 (Type 2) to 0.356 (Type 1), indicating that consumers differ substantially in the weight they place on site quality relative to privacy.

In Appendix D.1 we provide detailed discussion of model fit across our moments.

10 Welfare and Information Provision

10.1 The Value of Privacy

Before turning to counterfactuals, it is useful to quantify how much privacy-invasive practices cost consumers under the status quo. The per-period flow utility that consumer i of type g derives from visiting site j includes a privacy component

$$\Pi_{ij} = \sum_{k=1}^{K} \hat{\beta}_{gk}\, \theta_{ijk}^{0}$$

where θ^0_{ijk} is the consumer's prior belief about whether site j engages in practice k. $\hat{\beta}_{gk}$ are negative for invasive practices and positive for protective ones, Π_{ij} aggregates the consumer's perceived privacy burden at each site.

To convert utils to dollars, we exploit the conjoint experiment's monthly fee attribute. The conjoint estimates a price coefficient $\hat{\beta}^{\text{price}} = -0.529$ (conjoint utils per \$/month). Structural preferences are scaled relative to conjoint preferences by the stated-to-revealed parameter $\hat{\xi} = 1.34$. Thus, one util corresponds to $\hat{\xi}/|\hat{\beta}^{\text{price}}| = 1.34/0.529 = \$2.53$/month. This scaling is applied throughout the paper to dollarize both privacy costs and counterfactual welfare changes.

For now, we consider welfare with respect to a given set of beliefs, but later we follow the Train (2015) correction to adjust for the difference between perceived and experienced utility. Table 4 reports the mean privacy component across consumers indicating that under prior beliefs, the average consumer perceives to lose \$6.21/month from privacy-invasive practices but under truth, this rises to \$6.71/month—consumers underestimate the privacy cost by approximately \$0.50/month because their priors are overly optimistic on their most important privacy attributes (Section 5.3).

Table 4: Privacy Cost in the Status Quo (\$/month)

Type	At beliefs	At truth	Gap
Type 1	-6.09	-6.77	-0.68
Type 2	-4.44	-6.39	-1.95
Type 3	-7.70	-6.84	$+0.86$
Type 4	-6.09	-6.73	-0.64
Overall	-6.21	-6.71	-0.50

Notes: Privacy component $\Pi_{ij} = \sum_k \hat{\beta}_{gk}\theta^0_{ijk}$ averaged across consumers and sites. "At beliefs" uses consumer priors θ^0_{ijk}; "at truth" uses true attribute values μ^*_{jk}. Dollar conversion via the conjoint price coefficient (\$2.53/util). Gap = truth − beliefs; negative values indicate optimistic priors (truth is worse than believed).

The quality component of utility ($\hat{\lambda}_g \bar{Q}_j$) averages -8.55 utils (\$21.63/month); privacy accounts for roughly 29% of the total disutility from inside sites. Consumers browse despite these costs because the outside option forgoes both quality and privacy disutility; the Train-corrected status-quo welfare of \$0.85/month reflects the net gain from the browsing portfolio.

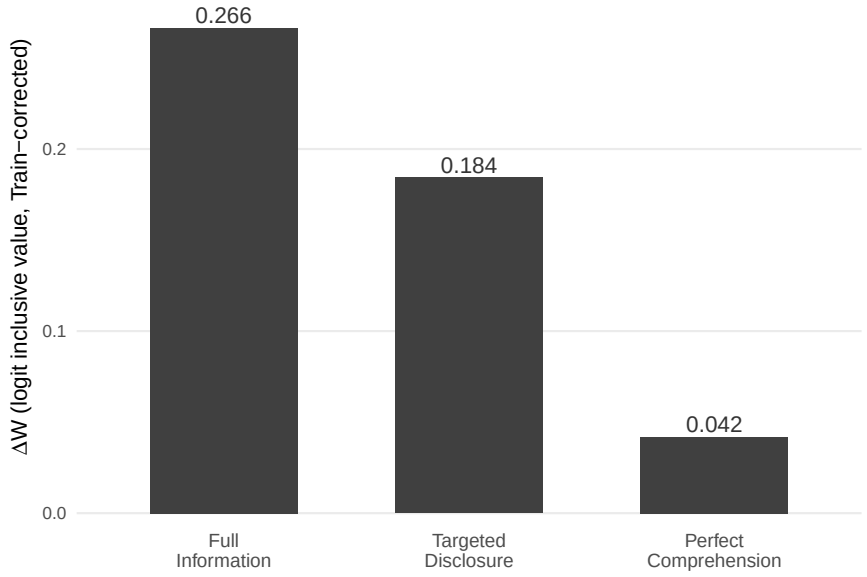
When we consider different counterfactuals we characterize the counterfactual change in total welfare $\Delta W_i = V_i^{\text{CF}} - V_i^{\text{SQ}}$. This can be driven by two channels: direct benefits from the restriction of practices and indirect benefits due to reallocation of browsing choices. We follow the Train (2015) correction in order to characterize the change in welfare based on experienced, as opposed to perceived, utility.

10.2 Information Provision

We consider three information counterfactuals that progressively relax the frictions preventing consumers from making informed browsing choices. The *full-information benchmark* eliminates all frictions simultaneously, establishing an upper bound on the value of privacy transparency. *Targeted disclosure* removes all frictions only for sites a consumer has already visited, isolating the value of correcting beliefs about familiar alternatives from the value of discovering new ones. *Perfect comprehension* removes the comprehension barrier while preserving search costs, asking whether consumers would learn enough through voluntary search if they could process privacy policies perfectly.

These counterfactuals change beliefs but hold privacy practices fixed, so the direct utility channel is zero. Under the Train (2015) correction, the entire welfare gain comes through the reallocation channel: consumers re-optimize their browsing portfolio given updated beliefs, and the welfare improvement is measured by the change in experienced utility evaluated at truth. We estimate status quo welfare of $\bar{W}^* = 0.337$ utils ($0.85/month), representing the mean consumer surplus from the browsing market relative to the outside option. All percentage welfare changes below are relative to this baseline. Figure 11 reports the Train-corrected welfare gains.

Figure 11: Welfare Decomposition: Information Provision Counterfactuals



Notes: This figure presents the welfare decomposition across the information provision counterfactuals (full information, targeted disclosure, perfect comprehension) according to the Train (2015) correction.

Full information. The full information counterfactual dictates mandatory disclosure which costlessly communicates each site's privacy practices ($\theta_{ijk} = \mu^*_{jk}$ for all i, j, k) and generates a large welfare gain: $+79\%$.

The mechanism for this result is straightforward: under status-quo beliefs, consumers cannot distinguish privacy-protective sites from privacy-invasive ones, so browsing is spread across sites with heterogeneous true practices. Full information reveals this heterogeneity, enabling consumers to shift toward sites that are better matched to their preferences. Across types, Type 3 benefits most, consistent with its near-zero baseline comprehension ($\hat{\alpha}_3^{\text{info}} = 0.005$)—these consumers learn the most from mandatory disclosure precisely because the comprehension and attention costs for these consumers were still high. Type 1 benefits least ($\$0.53$/month).

Targeted disclosure. The targeted disclosure counterfactual dictates mandatory disclosure which costlessly communicates each site's privacy practices for the sites that the consumer visited in the baseline period. This counterfactual provides a direct test of whether the full-information gains depend on consumers learning about the entire market or primarily reflect corrections to beliefs about familiar sites. This follows a key provision of our intervention and existing information disclosure policies online: they only target sites that consumers actually visit.

During the baseline period consumers visited an average of 18.9 characterized sites. Although these visited sites represent a small share of the 500 sites that we consider in the primary choice set, they are the sites where belief corrections translate most directly into browsing reallocation. The welfare gain is $+55\%$ – which captures 69% of the full-information gain – and reveals that most of the welfare value comes from correcting beliefs about sites consumers already visit—not from discovering new alternatives. The remaining gap quantifies the additional value of discovery: learning about unfamiliar sites enables reallocation that previously wasn't possible due to the disclosure environment.

Perfect comprehension. The previous counterfactuals remove all frictions associated with comprehension. In the perfect comprehension counterfactual we keep the estimated search costs but remove all comprehension frictions($\alpha_g^{\text{salience}} = 1$) for all types. The overall welfare gain from this counterfactual small: $+12\%$.

This result follows since the relatively high search costs still limit how many websites a consumer will acquire information about. While removing the comprehension costs increases the expected value of information, the mean search probability is only 2.1% and consumers search an average of 8.9 sites as the search cost remains prohibitively high for most (consumer, site) pairs. The small number of sites searched limits the scope for belief correction to translate into meaningful reallocation which results in the limited welfare gains.

This suggests both comprehension and search frictions must be addressed to unlock most of the welfare gains from information provision. At estimated comprehension ($\hat{\alpha}_g^{\text{salience}} \in [0.005, 0.130]$), the expected Value of Information falls below the search cost for essentially all consumers, producing zero endogenous search and zero welfare gain.

We can benchmark these estimates against our conjoint price coefficient. Full information is worth \$0.67/month per consumer, targeted disclosure \$0.47/month, and perfect comprehension \$0.11/month. Because the conjoint price coefficient exhibits substantial cross-consumer heterogeneity (SD $= 0.53$), these dollar figures should be interpreted as approximate magnitudes.

We can additionally benchmark these estimates against existing valuations in the privacy literature. Stated-preference studies generally find higher WTP: Winegar and Sunstein (2019) report median WTP of \$5/month to protect internet browsing data, and Savage and Waldman (2015) estimate WTP of \$1.19/month to avoid sharing browser history and \$1.75/month for location data on smartphones. Our full-information estimate of \$0.67/month is lower, but it measures a fundamentally different object—the welfare gain from *knowing* sites' privacy practices, not from *eliminating* them. Information enables reallocation toward less invasive alternatives but does not remove the underlying practices industry-wide. Revealed-preference estimates are scarcer and tend to find smaller magnitudes. Athey et al. (2025) document near-zero behavioral costs of privacy in a field experiment where students readily traded personal data for small incentives. More closely related, Tsai et al. (2011) find that consumers pay a \$0.60 premium per purchase for privacy-protective online merchants when privacy ratings are made salient—a transparency intervention analogous to our targeted-disclosure counterfactual (\$0.47/month).

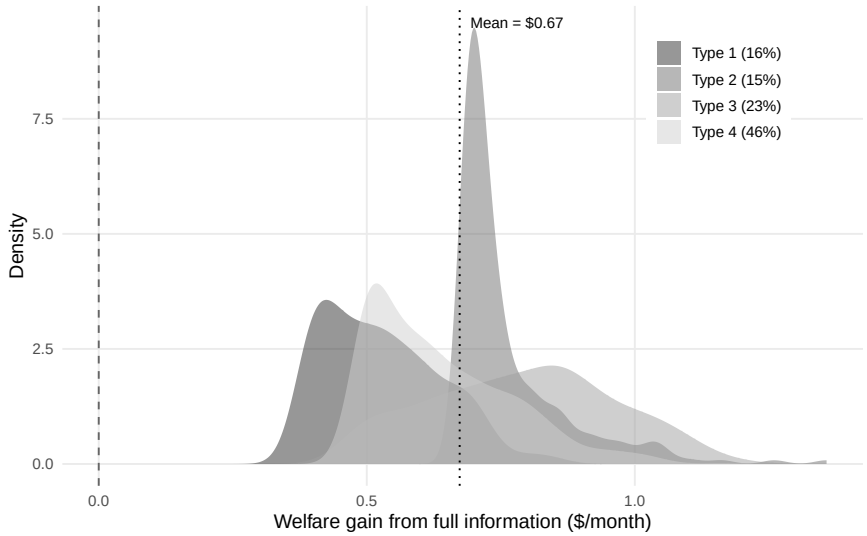
Figure 12 plots the consumer-level distribution of dollarized welfare gains under full information with mean gain of \$0.67/month (median \$0.67). The distribution exhibits substantial heterogeneity across types: Type 3 consumers gain the most (\$0.79/month on average), consistent with their pessimistic priors being corrected upward, while Type 1 gains the least (\$0.53/month).

10.3 GDPR-Style Bundled Regulation

We construct a bundled regulation counterfactual that directly restricts the privacy practices that websites can engage in. In order to make this counterfactual realistic we follow the practices explicitly regulated by the GDPR.

Of the 19 privacy attributes in our model, we exclude practices that the GDPR explicitly *promotes* (deletion rights, data anonymization, storage limits, automated-decision safeguards, policy-change notification), practices that are broadly permissible under the regulation (sharing with service providers under data-processor agreements per Article 28; sharing with law enforcement under legal obligation per Article 6(1)(c); collecting first-party browsing logs under legitimate in-

Figure 12: Distribution of Consumer-Level Welfare Gains from Full Information



Notes: Each density shows the distribution of ΔW_i under full information ($\theta_{ijk} = \mu^*_{jk}$ for all i, j, k), converted to \$/month via the conjoint price coefficient. Dotted vertical line marks the overall mean (\$0.67/month). Type shares: Type 1 (16%), Type 2 (15%), Type 3 (23%), Type 4 (46%).

terest), and personalization. We also exclude collecting biographical data, which is not a special category under the GDPR, and collecting financial data and sharing financial transactions, which are primarily governed by sector-specific regulation (PSD2) rather than the GDPR. Collecting social media data is excluded because it involves receipt rather than third-party dissemination of information. The remaining six practices map to three core GDPR mechanisms:

- *Special-category data* (Article 9): collecting sensitive personal information (18% prevalence), which requires explicit consent and may not rely on legitimate interest;
- *Purpose limitation and data minimization* (Articles 5–6): collecting location data (90%) and offsite browsing history (69%), which typically lack a legitimate-interest basis for commercial data aggregation;
- *Third-party sharing restrictions* (Articles 6, 13–14): sharing with advertisers (75%), partners (77%), and social media platforms (35%), all of which require explicit consent under GDPR and are frequently the subject of enforcement actions.

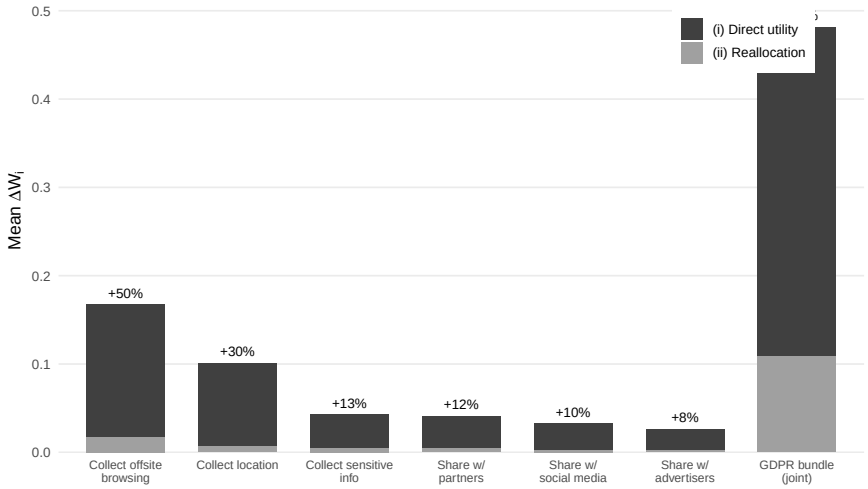
The GDPR bundle changes actual practices, so the welfare decomposition has two active channels: the direct utility gain from eliminating disliked practices at fixed browsing shares, and the reallocation gain from consumers re-optimizing their portfolio in response to the changed landscape.

The joint ban yields a welfare gain of $+143\%$ (\$1.22/month), with the direct channel contributing 77% and reallocation 23%. The gain is broadly uniform across types, reflecting the high

prevalence of the banned practices. This estimate is an upper bound: it assumes practices are eliminated outright with no adjustment to site quality or content provision. In practice, restricting data collection and sharing may reduce advertising revenue and personalization capabilities, potentially lowering the quality of ad-supported content. The magnitude of this supply-side response is beyond the scope of our demand-side model, but given the small estimated $\hat{\delta}$ coefficients—which imply negligible quality dependence on tracking infrastructure—the quality adjustment is likely modest.

Figure 13 decomposes the bundle by practice. Individually, the largest contributions come from banning offsite browsing collection ($+50\%$) and location tracking ($+30\%$). The sum of individual bans ($+121\%$) falls short of the joint bundle ($+143\%$), implying complementarity of 14.9%: eliminating multiple practices simultaneously yields additional reallocation gains beyond what individual bans achieve, as consumers shift toward sites that are clean on all dimensions.

Figure 13: GDPR Bundle Decomposition by Practice



Notes: Each bar shows the welfare gain from banning a single GDPR-related practice across all sites, decomposed into the direct utility channel (i) and the reallocation channel (ii). The rightmost bar shows the joint ban of all six practices. Percentages above each bar are relative to Train-corrected status quo welfare $\bar{W}^* = 0.337$.

Our model does not include an explicit opt-out decision, instead our GDPR counterfactual eliminates practices outright. In practice, GDPR operates through consent requirements—cookie banners, privacy dashboards—that impose time costs on consumers. Fradkin et al. (2024) estimate these costs directly using a field experiment that randomizes consent interfaces. They find that users spend an average of 7.3 seconds per consent banner interaction, implying a time cost of roughly \$4/week (at \$36/hour) when banners appear on every domain. Browser-level consent—a one-time choice replacing repeated site-by-site decisions—improves welfare by \$3.70/user-week

relative to optimal site-level banners, primarily by eliminating repeated decision costs. These time costs provide context for our estimates. Our GDPR bundle value of \$1.22/month (\$0.31/week) measures the *privacy gain* from eliminating harmful practices, abstracting from the consent mechanism. In our framework, the closest analog to the consent cost is the estimated search cost $\hat{c}_s = 4.92$—which similarly represents the friction of acquiring privacy information site-by-site. That this cost is prohibitively high for most consumers, producing near-zero voluntary search, parallels Fradkin et al.'s finding that site-level consent imposes costs large enough to distort choices away from consumers' true preferences.

11 Conclusion

This paper demonstrates that while participants value privacy and hold systematically biased beliefs about website data practices, the effectiveness of privacy information provision depends on reducing information complexity as well as reducing search costs. In a large-scale field experiment with nearly 1,600 participants, we find that making privacy policies salient increases search behavior by a factor of 17 but produces no improvement in knowledge about privacy practices, whereas providing simplified, personalized privacy information improves belief accuracy by 7 percentage points. Despite this learning, behavioral responses were modest: participants show limited changes on the intensive margin of website usage, with small but detectable shifts toward privacy-conscious sites on the extensive margin.

Using the resulting data and experimental variation, we estimate a structural model of browsing choices and privacy information acquisition. We use this to demonstrate that consumer's valuations of privacy are modest compared to their valuations of site quality. We evaluate two key provisions of privacy regulation: information provision and restriction of specific privacy attributes. We show that there are meaningful welfare gains from both – eliminating search and information frictions leads to an average gain of \$0.67 per consumer per month and GDPR-style attribute restrictions lead to an average gain of \$1.22 per consumer per month.

Our findings have important implications for privacy regulation. Disclosure requirements that mandate only notice, as in current US practice, are unlikely to meaningfully inform consumers, while simplified "nutrition label" approaches that reduce complexity improve knowledge and consumer welfare. Direct restrictions on data practices, as under the GDPR, would generate larger welfare gains but may be more difficult to implement due to potential costs on firms. This suggests that policies that reduce the search and comprehension costs of privacy information can deliver modest, but, meaningful consumer benefits in practice without as dramatic firm-side costs.

References

Acquisti, A. and R. Gross (2006). Imagined communities: Awareness, information sharing, and privacy on the facebook. In *International Workshop on Privacy Enhancing Technologies*, pp. 36–58. Springer.

Acquisti, A., L. K. John, and G. Loewenstein (2013). What is privacy worth? *The Journal of Legal Studies 42*(2), 249–274.

Acquisti, A., C. Taylor, and L. Wagman (2016). The economics of privacy. *Journal of Economic Literature 54*(2), 442–92.

Adjerid, I., A. Acquisti, and G. Loewenstein (2019). Choice architecture, framing, and cascaded privacy choices. *Management Science 65*(5), 2267–2290.

Adjerid, I., E. Peer, and A. Acquisti (2018). Beyond the privacy paradox: Objective versus relative risk in privacy decision making. *Management Information Systems Quarterly 42*(2), pp. 465–488.

Agte, P., C. Allende, A. Kapor, C. Neilson, and F. Ochoa (2024). Search and biased beliefs in education markets. Technical report.

Alcobendas, M., S. Kobayashi, and M. Shum (2026). The impact of privacy measures on online advertising markets. *Review of Economic Studies*.

Allcott, H., J. C. Castillo, M. Gentzkow, L. Musolff, and T. Salz (2025). Sources of market power in web search: Evidence from a field experiment. Technical report, National Bureau of Economic Research.

Allcott, H. and C. Knittel (2019). Are consumers poorly informed about fuel economy? evidence from two experiments. *American Economic Journal: Economic Policy 11*(1), 1–37.

Allenby, G. M., N. Hardt, and P. E. Rossi (2019). Chapter 3 - economic foundations of conjoint analysis. In J.-P. Dubé and P. E. Rossi (Eds.), *Handbook of the Economics of Marketing*, Volume 1, pp. 151–192. North-Holland.

Amos, R., G. Acar, E. Lucherini, M. Kshirsagar, A. Narayanan, and J. Mayer (2021). Privacy policies over time: Curation and analysis of a million-document dataset. In *Proceedings of the Web Conference 2021*, pp. 2165–2176.

Andersen, S., G. W. Harrison, M. I. Lau, and E. E. Rutström (2006). Elicitation using multiple price list formats. *Experimental Economics 9*, 383–405.

Aridor, G. (2025). Measuring substitution patterns in the attention economy: An experimental approach. *The RAND Journal of Economics 56*(3), 302–324.

Aridor, G., Y.-K. Che, B. Hollenbeck, M. Kaiser, and D. McCarthy (2025). Evaluating the impact of privacy regulation on e-commerce firms: Evidence from apple's app tracking transparency. *Management Science*.

Aridor, G., Y.-K. Che, and T. Salz (2023). The effect of privacy regulation on the data industry: empirical evidence from gdpr. *The RAND Journal of Economics 54*(4), 695–730.

Aridor, G., R. Jiménez-Durán, R. Levy, and L. Song (2026). Experiments with social media. In *Handbook of Experimental Methods in the Social Sciences*. Edward Elgar Publishing.

Athey, S., C. Catalini, A. Moehring, and C. Tucker (2025). The digital privacy paradox and choice architecture: Evidence from an experiment in fintech. *Management Information Systems Quarterly*.

Barahona, N., C. Otero, and S. Otero (2023). Equilibrium effects of food labeling policies. *Econometrica 91*(3), 839–868.

Beknazar-Yuzbashev, G., R. Jiménez Durán, J. McCrosky, and M. Stalinski (2022). Toxic content and user engagement on social media: Evidence from a field experiment. *Available at SSRN*.

Berry, S. T. (1994). Estimating discrete-choice models of product differentiation. *The RAND Journal of Economics*, 242–262.

Bian, B., X. Ma, and H. Tang (2025). The supply and demand for data privacy: Evidence from mobile apps. *Available at SSRN 3987541*.

Bian, B., M. Pagel, H. Tang, and D. Raval (2023). Consumer surveillance and financial fraud. Technical report, National Bureau of Economic Research.

Blake, T., S. Moshary, K. Sweeney, and S. Tadelis (2021). Price salience and product choice. *Marketing Science 40*(4), 619–636.

Bonhomme, S., T. Lamadon, and E. Manresa (2022). Discretizing unobserved heterogeneity. *Econometrica 90*(2), 625–643.

Brown, J., T. Hossain, and J. Morgan (2010). Shrouded Attributes and Information Suppression: Evidence from the Field. *The Quarterly Journal of Economics 125*(2), 859–876.

Brown, Z. Y. and J. Jeon (2024). Endogenous information and simplifying insurance choice. *Econometrica 92*(3), 881–911.

Brynjolfsson, E., A. Collis, and F. Eggers (2019). Using massive online choice experiments to measure changes in well-being. *Proceedings of the National Academy of Sciences 116*(15), 7250–7255.

Charness, G., U. Gneezy, and A. Imas (2013). Experimental methods: Eliciting risk preferences. *Journal of Economic Behavior & Organization 87*, 43–51.

Chrzan, K. (2010). Using partial profile choice experiments to handle large numbers of attributes. *International Journal of Market Research 52*, 827–840.

Collis, A., A. Moehring, A. Sen, and A. Acquisti (2023). Information frictions and heterogeneity in valuations of personal data.

Conjointly (n.d.). Practical tips for setting up conjoint studies: Nine tips for specifying attributes and levels. `https://conjointly.com/guides/tips-for-setting-up-conjoint-studies/#nine-tips-for-specifying-attributes-and-levels`. Accessed January 9, 2026.

Dave, C., C. Eckel, C. Johnson, and C. Rojas (2010). Eliciting risk preferences: When is simple better? *Journal of Risk and Uncertainty 41*(3), 219–243.

deHaan, E., Y. Song, C. Xie, and C. Zhu (2020). Disclosure obfuscation in mutual funds. *Available at SSRN*.

Demirer, M., D. J. J. Hernández, D. Li, and S. Peng (2024). Data, privacy laws and firm production: Evidence from the gdpr. Technical report, National Bureau of Economic Research.

Dranove, D. and G. Z. Jin (2010). Quality disclosure and certification: Theory and practice. *Journal of Economic Literature 48*(4), 935–63.

Farronato, C. (2025). Data as the new oil: Parallels, challenges, and regulatory implications. *Harvard University, Centre for Economic Policy Research and National Bureau of Economic Research*.

Fradkin, A., C. Farronato, and T. Lin (2024). Data sharing and website competition: The role of dark patterns. *Available at SSRN 4920040*.

Gerber, A. S. and D. P. Green (2012). *Field experiments: Design, analysis, and interpretation*. WW Norton.

Goldberg, S. G., G. A. Johnson, and S. K. Shriver (2024). Regulating privacy online: An economic evaluation of the gdpr. *American Economic Journal: Economic Policy 16*(1), 325–358.

Goldfarb, A. and V. F. Que (2023). The economics of digital privacy. *Annual Review of Economics 15*, 267–286.

Goldfarb, A. and C. Tucker (2012). Shifts in privacy concerns. *American Economic Review 102*(3), 349–353.

Goldfarb, A. and C. Tucker (2019). Digital economics. *Journal of Economic Literature 57*(1), 3–43.

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

Green, P. E., A. M. Krieger, and Y. Wind (2001). Thirty years of conjoint analysis: Reflections and prospects. *Interfaces 31*, S56–S73.

Haaland, I., C. Roth, S. Stantcheva, and J. Wohlfart (2025). Understanding economic behavior using open-ended survey data. *Journal of Economic Literature 63*(4), 1244–1280.

Haaland, I., C. Roth, and J. Wohlfart (2023). Designing information provision experiments. *Journal of Economic Literature 61*(1), 3–40.

Handel, B. R. and J. T. Kolstad (2015). Health insurance for "humans": Information frictions, plan choice, and consumer welfare. *American Economic Review 105*(8), 2449–2500.

Hoffman, M. D. and A. Gelman (2014). The no-u-turn sampler: Adaptively setting path lengths in hamiltonian monte carlo. *Journal of Machine Learning Research 15*, 1593–1623.

Janssen, R., R. Kesler, M. E. Kummer, and J. Waldfogel (2022). Gdpr and the lost generation of innovative apps. Technical report, National Bureau of Economic Research.

Jia, J., G. Z. Jin, and L. Wagman (2021). The short-run effects of the general data protection regulation on technology venture investment. *Marketing Science 40*(4), 661–684.

Jin, G. Z., M. Luca, and D. Martin (2022). Complex disclosure. *Management Science 68*(5), 3236–3261.

Johnson, G. A. (2024). Economic research on privacy regulation: Lessons from the gdpr and beyond. In *The Economics of Privacy*, pp. 97–126. University of Chicago Press.

Johnson, G. A., S. K. Shriver, and S. Du (2020). Consumer privacy choice in online advertising: Who opts out and at what cost to industry? *Marketing Science 39*(1), 33–51.

Johnson, G. A., S. K. Shriver, and S. G. Goldberg (2023). Privacy and market concentration: Intended and unintended consequences of the gdpr. *Management Science 69*(10), 5695–5721.

Levy, R. (2021). Social media, news consumption, and polarization: Evidence from a field experiment. *American economic review 111*(3), 831–870.

Lin, T. (2022). Valuing intrinsic and instrumental preferences for privacy. *Marketing Science 41*(4), 663–681.

List, J. A. and C. A. Gallet (2001). What experimental protocol influence disparities between actual and hypothetical stated values? *Environmental and Resource Economics 20*(3), 241–254.

Luca, M. and J. Smith (2015). Strategic disclosure: The case of business school rankings. *Journal of Economic Behavior & Organization 112*, 17–25.

McClain, C., M. Faverio, M. Anderson, and E. Park (2023). How Americans view data privacy. `https://www.pewresearch.org/internet/2023/10/18/how-americans-view-data-privacy/`.

Murphy, J. J., P. G. Allen, T. H. Stevens, and D. Weatherhead (2005). A meta-analysis of hypothetical bias in stated preference valuation. *Environmental and Resource Economics 30*(3), 313–325.

Norberg, P. A., D. R. Horne, and D. A. Horne (2007). The privacy paradox: Personal information disclosure intentions versus behaviors. *Journal of Consumer Affairs 41*(1), 100–126.

Peukert, C., S. Bechtold, M. Batikas, and T. Kretschmer (2022). Regulatory spillovers and data governance: Evidence from the gdpr. *Marketing Science 41*(4), 746–768.

Rossi, P. E., G. M. Allenby, and R. McCulloch (2005). *Bayesian Statistics and Marketing*. Chichester, UK: John Wiley & Sons, Ltd.

Savage, S. J. and D. M. Waldman (2015). Privacy tradeoffs in smartphone applications. *Economics Letters 137*, 171–175.

Sawtooth Software (n.d.). Cbc partial-profile designs. `https://sawtoothsoftware.com/help/lighthouse-studio/manual/cbc-partial-profile-designs.html`. Lighthouse Studio Manual.

Spiekermann, S., J. Grossklags, and B. Berendt (2001). E-privacy in 2nd generation e-commerce: privacy preferences versus actual behavior. In *Proceedings of the 3rd ACM conference on Electronic Commerce*, pp. 38–47.

Srinath, M., S. Sundareswara, P. Venkit, C. L. Giles, and S. Wilson (2023). Privacy lost and found: An investigation at scale of web privacy policy availability. In *Proceedings of the ACM Symposium on Document Engineering 2023*, DocEng '23, New York, NY, USA. Association for Computing Machinery.

Tang, C., Z. Liu, C. Ma, Z. Wu, Y. Li, W. Liu, D. Zhu, Q. Li, X. Li, T. Liu, and L. Fan (2023). Policygpt: Automated analysis of privacy policies with large language models.

Train, K. E. (2015). Welfare calculations in discrete choice models when anticipated and experienced attributes differ: A guide with examples. *Journal of Choice Modelling 16*, 15–22.

Tsai, J. Y., S. Egelman, L. Cranor, and A. Acquisti (2011). The effect of online privacy information on purchasing behavior: An experimental study. *Information Systems Research 22*(2), 254–268.

Wagner, I. (2022). Privacy policies across the ages: Content and readability of privacy policies 1996–2021.

Wernerfelt, N., A. Tuchman, B. T. Shapiro, and R. Moakler (2025). Estimating the value of offsite tracking data to advertisers: Evidence from meta. *Marketing Science 44*(2), 268–286.

Winegar, A. G. and C. R. Sunstein (2019). How much is data privacy worth? A preliminary investigation. *Journal of Consumer Policy 42*(3), 425–440.

A Data Collection Appendix

Table A.1: Privacy Policy Feature Descriptions

Feature	Field	Short Description	Long Description
use	law	Law Enforcement	Your user data may be shared with law enforcement.
use	ads	Advertising Partners	Your user data may be shared with advertising partners.
use	finance	Financial Transactions	Your user financial transaction information may be shared with other parties.
use	social-media	Social Media	Your user data may be shared with social media platforms.
use	partners	3rd Party Partners	Your user data may be shared with 3rd party partners (non-advertising business partners).
use	service	3rd Party Services	Your user data may be shared with 3rd party services (service providers or other businesses).
use	personalization	Website Personalization	You can choose to let the website personalize content for you, without sharing your data with partners.
collect	log	Browsing On-site	Your browsing behavior on and across the current website may be collected.
collect	offsite	Browsing Off-Site	Your browsing behavior off the current website may be collected.
collect	sensitive	Sensitive Personal Info	Your sensitive personal user information (religious affiliation, sexual orientation, political beliefs, etc.) may be collected.
collect	financial	Financial Data	Your financial information (purchase history, credit card information, etc.) may be collected if available.
collect	bio	Demographic Data	Your demographic data (gender, race, etc.) may be collected if available.
collect	location	Location	Your location data may be collected if available.
collect	social-media	Social Media Info	Your social media account information (and possibly associated behavior) may be collected, if available.
control	storage	Data is secured	The privacy policy specifies whether the collected data is properly secured.
control	anonymized	Data is anonymized	The privacy policy specifies whether the collected data is properly anonymized.
control	automated	Automated Deletion	Your data will be automatically deleted after a specified time period.
control	change	Notification of Change	You will be notified if there is a material change to the privacy policy.
control	deletion	Data Deletion	You have the ability to request the deletion of your data, but your data is not automatically deleted.

Chrome Extension: We discuss additional technical details regarding the Chrome Extension that we use in the study. This extension is developed to automatically collect information on data collection and provide information to users overlaid on their normal web experience.

The automatic data collection done by the Chrome Extension uses the same underlying detection mechanisms that power the popular open-source AdBlockPlus extension. Both AdBlockPlus and our extension use the continually updated rules from `easylist.to` that allow for detection of network requests to third party advertising and analytics firms as well as which HTML ele-

ments are advertisements. Our software extends the implementation from AdBlockPlus to enable directly reading metadata sent (e.g., cookies) via network requests. This provides us with data on the identity of 3rd party cookies and network requests. Beyond this, our extension also provides an implementation of robust time tracking across pages and websites that accounts for possible user idleness.

There are several important technical details that are relevant for compliance and data collection. First, our extension can uniquely identify each participant, as they are required to input their randomly assigned experiment ID when they enter into the study. Second, the extension logs an internal identifier that is maintained so long as the extension is installed. If the participant uninstalls the extension, then the timestamp, along with both identifiers, is recorded in our servers. Participants receive an automated email indicating that they have opted out of the study upon uninstalling. Thus, we can observe compliance with the experiment via the extension. Finally, all of our collected data (via the extension) is persisted and stored on our servers at an interval of one minute – ensuring we can continuously collect data as long as the extension is installed.

Privacy Policy Classification & Construction of Privacy Characteristics: To gauge the space of relevant privacy characteristics, we first ran an open response survey asking "What website characteristics do you consider privacy invasive? What website behaviors do you consider privacy enhancing?" Then, using sources such as `https://pirg.org/resources/how-to-read-a-privacy-policy/`, `https://usableprivacy.org/data` and consultation with legal experts we have constructed an exhaustive set of privacy relevant, internet service attributes. The attributes cover a variety of topics such as security, storage, sharing and selling of data, and rights. These sets of attributes are presented below in Table A.1.

We use the tracker data collected by the Chrome Extension to provide validation for our labeled privacy characteristics. Across our entire dataset, we observe nearly 30,000 unique third party domains that are shared data and nearly 1.5 million unique trackers. In order to ensure that we are able to accurately match these domains to their purpose, we rely on four external sources to classify them. The primary external source that we rely on comes from Third Party Web which pulls data from the HTTP Archive and provides classifications of the purpose for each request [29]. In addition, we supplement our classifications with data from Ghostery[30], WhoTracksMe[31], and Cloudflare Radar[32] to improve the coverage. The final classification is therefore assigned in a two-stage process. First, we query the sources in a descending priority of Third Party Web, Ghostery, WhoTracksMe, and finally Cloudflare Radar. Second, we align all final categories to the Third

[29] `https://www.thirdpartyweb.today/about`
[30] `https://github.com/ghostery/trackerdb`
[31] `https://github.com/whotracksme/whotracks.me`
[32] `https://radar.cloudflare.com`

Party Web taxonomy for consistency, as this source empirically classifies the most trackers.

Since our tracker data captures analytics data collection, we focus on the 'Browsing on-site data' measure from Table A.1. If we observe the presence of any of the analytics trackers then this reliably indicates that the website is engaged in this practice. Table A.2 provides the confusion matrix between the tracker and policy data for the 500 websites for which we have policy data and shows that there is general alignment between the two measures.

Table A.2: Confusion Matrix for Analytics Data

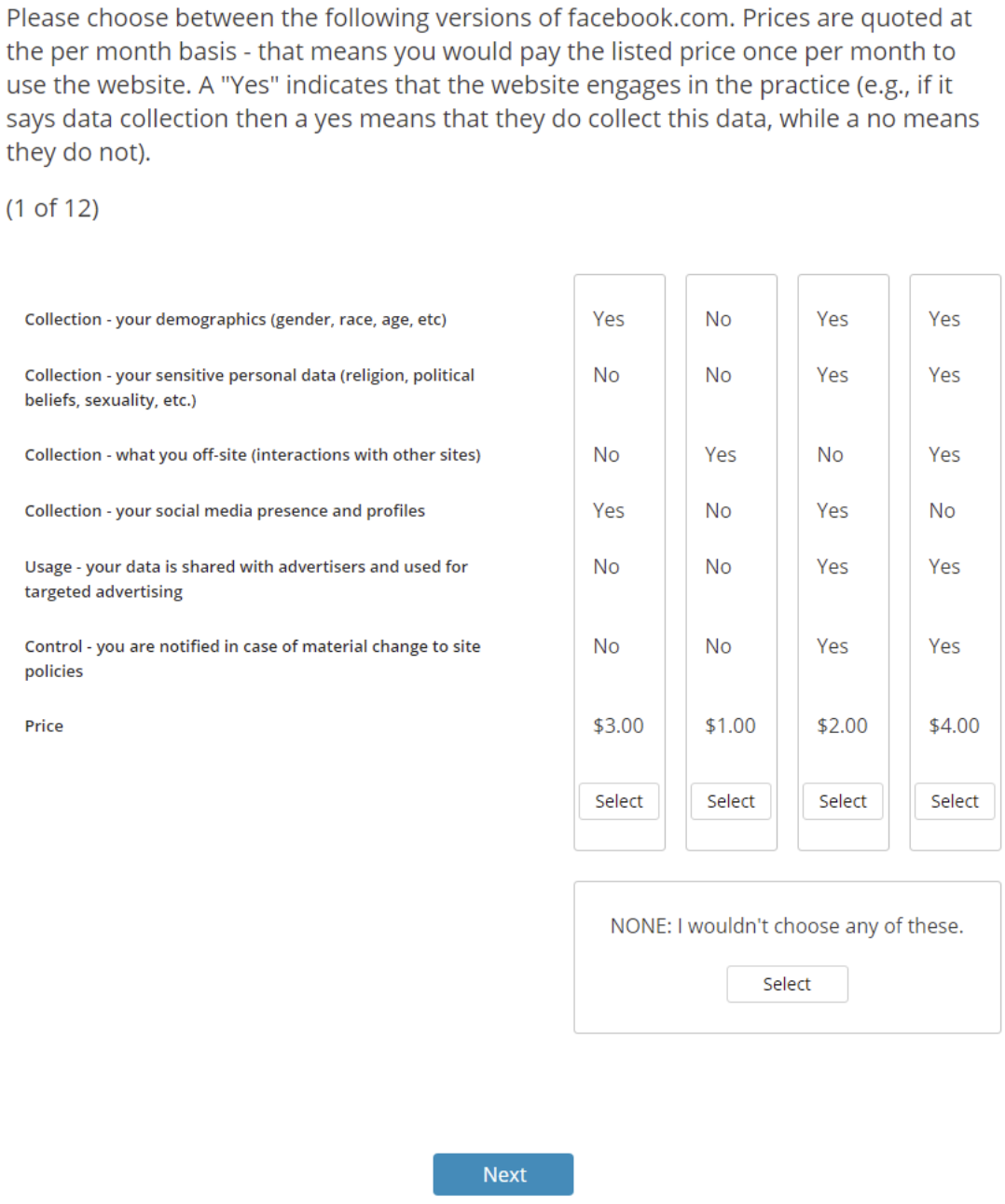
		Detect	
		Tracker: No	Tracker: Yes
Claim	Policy: No	1	11
	Policy: Yes	26	458

B Conjoint Details

In this section, we detail the conjoint method used to measure participant's stated preferences for privacy attributes. Conjoint experiments consist of a number of tasks. Each task presents a participant with a set of theoretical products which may vary across some set of characteristics, for example price, and asks the participant which product they would most like to purchase. Participants repeat this task multiple times, with product sets and characteristics varying across tasks. Thus, a conjoint experiment generates exogenous variation in product characteristics and choice sets, allowing for the estimation of preferences. As these choices tasks are entirely hypothetical, estimated preferences are reflective of stated-preferences. Yet, conjoint presents a systemized way of eliciting preferences that has been shown to contribute to more reliable as well as internally and externally valid estimates (Allenby et al., 2019; Green et al., 2001).

Our goal is to elicit preferences for the 19 privacy practices we measure. Our conjoint proceeds in two steps. First, participants are randomly assigned one of four website categories (e-commerce, news, entertainment, or social media) and asked "what is your favorite website within this category?" Participants then proceed to the conjoint. Each participant is presented with 12 choice tasks, with each choice task consisting of 4 products and a no purchase option. Products in the choice task consist of versions of the participant's favorite website, with each version varying in price *and* the privacy practices it implements. See Figure B.1 for an example choice task.

Figure B.1: Conjoint Details



Two aspects of our design should be noted. First, we implement our conjoint with a specific website in mind, as practitioners have long highlighted the necessity of specific and realistic choice settings for estimate consistency (Conjointly, nd). Second, the literature on conjoint has found that products with many attributes are likely to confuse participants, leading to poor quality data. To overcome this, we follow practitioners and implement a *partial-profile* conjoint design (Sawtooth Software, nd). In partial profile designs, a subset of attributes (in our case, six) are chosen at

random for each task and vary across products. Price is always included, and non-visible attributes are assumed to be held constant across products within a task. While partial profile designs require more data collection to estimate preferences, they have been shown to generate more reliable data (Chrzan, 2010).

To make sure participants understand the choice task, our survey includes multiple examples and attention checks to ensure participants (1) understand how products vary within a task and (2) recognize that non-visible attributes remain constant across all products within a choice task. Next, we estimate specify a hierarchical Bayesian utility model in order to estimate preferences from our conjoint data (Rossi et al., 2005).

B.1 Model and Estimation

Utility for respondent n, task t, alternative a is:

$$U_{n,t,a} = \begin{cases} x_{n,t,a}^{\top}\boldsymbol{\beta}_n + \alpha_{w(n)} + \boldsymbol{\gamma}^{\top} d_n + \varepsilon_{n,t,a} & a \in \{1, \ldots, A\} \text{ (in-market)} \\ 0 + \varepsilon_{n,t,A+1} & a = A+1 \text{ (outside)} \end{cases}$$

where $\varepsilon_{n,t,a} \overset{\text{iid}}{\sim} T1EV$, outside utility is fixed at 0 for identification. The utility function has three additive components beyond individual preferences:

1. $x_{n,t,a}^{\top}\boldsymbol{\beta}_n$: Individual-specific feature preferences (hierarchical)
2. $\alpha_{w(n)}$: Website-specific intercept (random effect)
3. $\boldsymbol{\gamma}^{\top} d_n$: Demographic utility shifter (fixed effect)

The website random intercepts (α_w) capture website-specific baseline attractiveness that is not explained by the observed features. Some websites may be inherently more engaging or trustworthy, leading respondents to be more likely to choose any alternative when viewing that website's privacy policy. In addition, we include demographic utility shifters ($\boldsymbol{\gamma}$) which are a *utility level shifters* that apply equally to all in-market alternatives. Demographics do not moderate the effects of privacy attributes.

Individual preferences $\boldsymbol{\beta}_n \in \mathbb{R}^F$ follow a three-level hierarchy:

Level 1 (Individual):

$$\boldsymbol{\beta}_n \sim \mathcal{N}(\boldsymbol{\mu}_{c(n)}, \boldsymbol{\Sigma}_{c(n)})$$

We use non-centered parameterization to improve sampling efficiency:

$$\boldsymbol{\beta}_n = \boldsymbol{\mu}_{c(n)} + \mathbf{L}_{\Sigma_{c(n)}} z_n, \quad z_n \sim \mathcal{N}(\mathbf{0}, \mathbf{I}_F)$$

Level 2 (Website Category): Covariance decomposition:

$$\boldsymbol{\Sigma}_c = \text{diag}(\boldsymbol{\tau}_c)\, \boldsymbol{\Omega}\, \text{diag}(\boldsymbol{\tau}_c), \quad \mathbf{L}_{\Sigma_c} = \text{diag}(\boldsymbol{\tau}_c)\mathbf{L}_\Omega$$

where:

$$\begin{aligned} \boldsymbol{\mu}_c &\sim \mathcal{N}(\mathbf{0}, 5^2\mathbf{I}_F) \quad \text{(category mean preferences)} \\ \tau_{c,f} &\sim \text{Lognormal}(0, \sigma_\tau^2) \quad \text{(within-category heterogeneity)} \end{aligned}$$

with $\sigma_\tau = 0.75$ for privacy and website features ($f = 1, \ldots, 20$), and $\sigma_\tau = 0.5$ for price ($f = 21$).

Level 3 (Population): Shared correlation structure:

$$\mathbf{L}_\Omega \sim \text{LKJcorr}(5), \quad \boldsymbol{\Omega} = \mathbf{L}_\Omega \mathbf{L}_\Omega^\top$$

In addition to preference parameters, our model includes website random effects for the individual's preferred website and demographic fixed effects. We complete our model by specifying their hierarchical structure. The website effects follow a hierarchical structure with sum-to-zero constraint:

$$\begin{aligned} \alpha_w^{\text{raw}} &\sim \mathcal{N}(0, 1) \quad \forall w \in \{1, \ldots, W\} \\ \sigma_w &\sim \text{Lognormal}(0, 0.5^2) \\ \alpha_w &= \sigma_w \cdot (\alpha_w^{\text{raw}} - \bar{\alpha}^{\text{raw}}) \end{aligned}$$

where $\bar{\alpha}^{\text{raw}} = \frac{1}{W}\sum_{w=1}^{W} \alpha_w^{\text{raw}}$. This parameterization ensures $\sum_w \alpha_w = 0$ for identification. Finally, demographic effects have a prior distribution of:

$$\boldsymbol{\gamma} \sim \mathcal{N}(\mathbf{0}, \mathbf{I}_D)$$

this weakly informative prior regularizes toward zero effect, appropriate given small sample sizes in some demographic groups.

Choice probability. After integrating out T1EV errors, we obtain the multinomial logit:

$$P(y_{n,t} = a \mid \boldsymbol{\beta}_n, \alpha_{w(n)}, \boldsymbol{\gamma}) = \frac{\exp(V_{n,t,a})}{\sum_{k=1}^{A+1} \exp(V_{n,t,k})}$$

where:

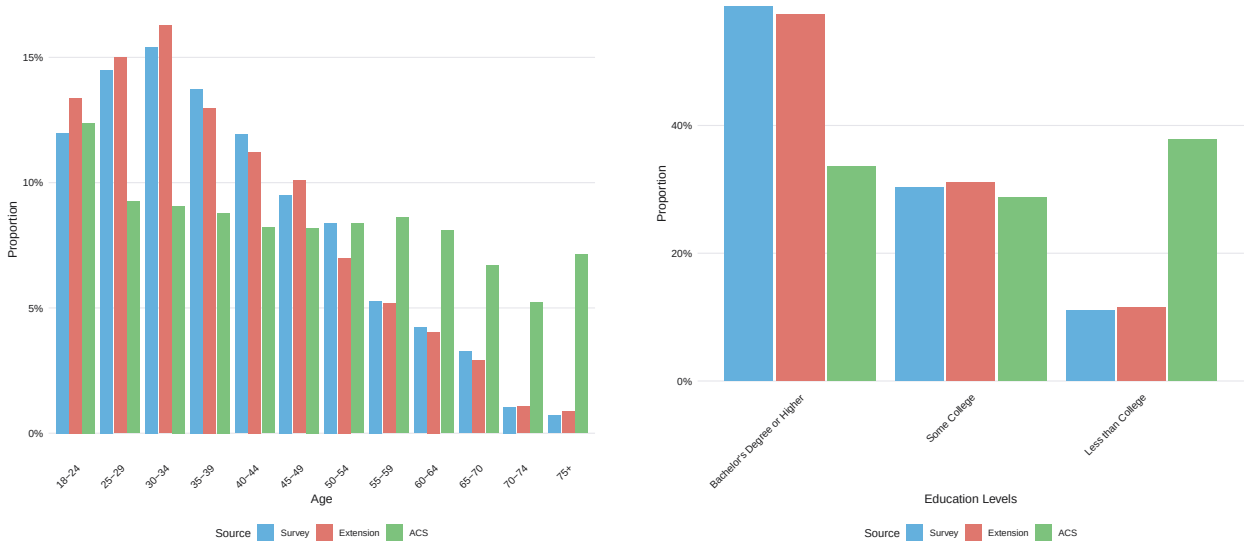
$$V_{n,t,a} = \begin{cases} x_{n,t,a}^{\top}\boldsymbol{\beta}_n + \alpha_{w(n)} + \boldsymbol{\gamma}^{\top} d_n & a \leq A \\ 0 & a = A + 1 \end{cases}$$

We use the No-U-Turn Sampler (NUTS) with within-chain parallelization via `reduce_sum` (Hoffman and Gelman, 2014) to estimate our model. We run four parallel chains, with 2000 warmup iterations and 3000 sampling iterations per chain.

C Additional Figures and Tables

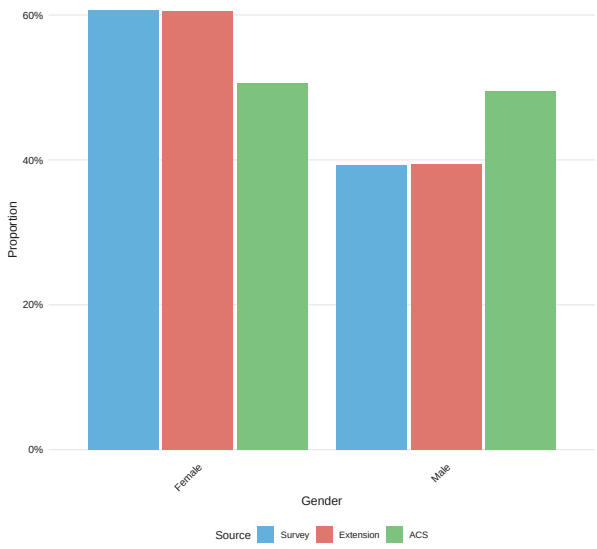
C.1 Figures on Experimental Sample

Figure C.1: Demographic Composition of Experimental Sample

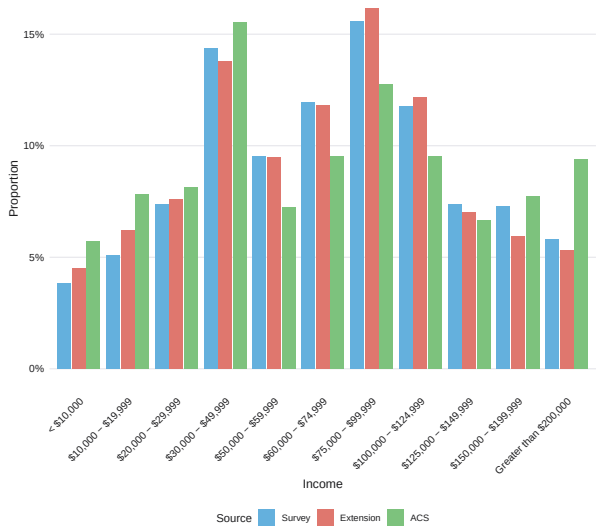


(a) Age

(b) Education



(c) Gender



(d) Income

NOTES: These figures compare the demographics of the survey sample (blue), extension sample (red), and the 2021 ACS (green) according to age (panel a), education (panel b), gender (panel c), and income (panel d). The survey sample includes all participants who completed the baseline survey. The extension sample conditions on participants who installed the browser extension and were randomized into the experiment.

Table C.1: Experimental Balance Tables

	Control Mean	Saliency Mean	Information Mean	p (S vs C)	p (I vs C)
Total Hours on Sites with Privacy Info	17.906	17.747	18.085	0.918	0.909
Daily Hours on Sites with Privacy Info	1.150	1.148	1.155	0.985	0.963
Total Hours Spent Online	28.311	28.476	28.526	0.936	0.916
Male	0.378	0.385	0.402	0.801	0.429
College Graduate	0.568	0.556	0.585	0.711	0.559
Age	38.410	38.605	39.411	0.812	0.226
Income	80586.374	79399.548	80932.890	0.708	0.915
No Income Disclosure	0.015	0.013	0.015	0.795	0.997
Is White	0.761	0.767	0.764	0.829	0.929
Privacy Attitude Index	0.580	0.579	0.575	0.875	0.559
Privacy Knowledge	0.914	0.910	0.908	0.829	0.755
PEW Extensive Margin	0.648	0.671	0.657	0.438	0.780
WTA	19.000	19.200	19.051	0.698	0.924
Mean Beliefs (Use)	52.509	52.093	52.093	0.660	0.663
Mean Beliefs (Collect)	60.741	61.016	59.622	0.802	0.305
Mean Beliefs (Control)	45.277	45.653	46.541	0.759	0.315
Mean Beliefs (Quality)	61.607	61.983	61.304	0.764	0.819
Conjoint category (News)	0.271	0.254	0.244	0.531	0.318
Conjoint category (Entertainment)	0.246	0.250	0.235	0.887	0.655
Conjoint category (E-commerce)	0.246	0.259	0.257	0.622	0.685
Conjoint category (Social)	0.237	0.237	0.265	nan	0.297

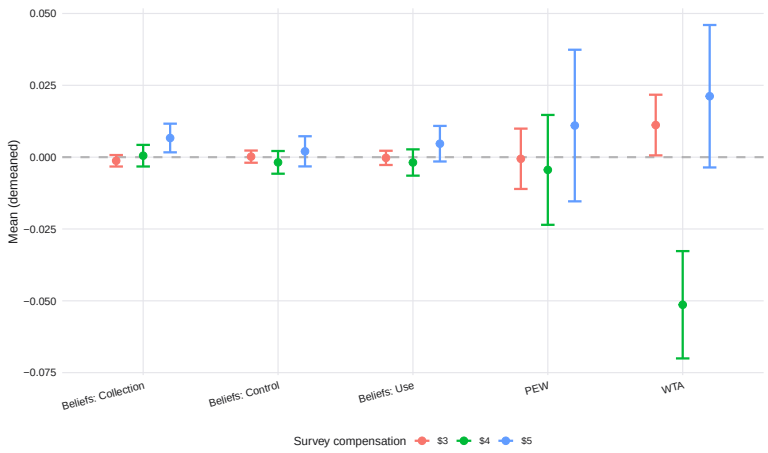
NOTES: This table presents the experimental balance table computed at the time of randomization. The columns denote the mean in the control, saliency, and information groups respectively with the final two columns denoting the p-value between the saliency and control as well as the information and control, respectively. The rows are as follows: total hours on sites with privacy information during baseline, daily hours on sites with privacy information during baseline (can differ due to differences in entry period), total hours spent online during baseline, male fraction, fraction with at least a college degree, age, income, choosing to not disclose income, fraction of participants that are white, privacy attitude index (combining PEW questions on privacy attitudes), privacy knowledge question regarding the ability of websites to track consumers across time without an email address, PEW question on likelihood of stopping to use a website based on privacy concerns, extension WTA, mean beliefs for prevalence of use/collect/control/quality attributes, and the fraction whose conjoint category is news/entertainment/e-commerce/social.

Table C.2: Top Websites used by Participants

	Domain	Daily Hours (per Active User)	Number of Users	Daily Hours (over all users)
1	Youtube	0.76	1030	0.52
2	Google	0.34	1487	0.34
3	Facebook	0.26	734	0.13
4	Amazon	0.18	1024	0.12
5	Reddit	0.19	936	0.12
6	Netflix	0.58	206	0.08
7	Chatgpt	0.18	520	0.06
8	Twitch	0.53	149	0.05
9	X	0.20	380	0.05
10	Hulu	0.49	137	0.04
11	Yahoo	0.15	423	0.04
12	Walmart	0.14	397	0.04
13	Microsoft	0.06	630	0.02
14	eBay	0.10	321	0.02
15	LinkedIn	0.08	399	0.02

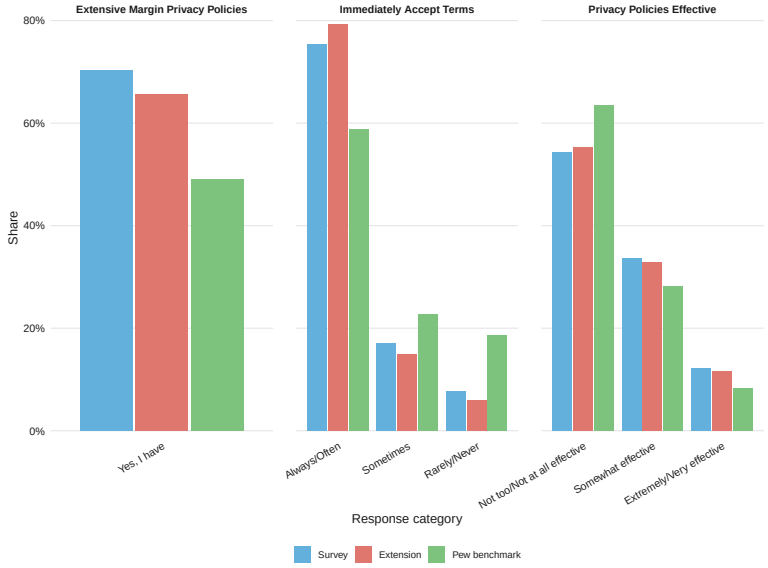
NOTES: This table presents the top 15 websites used in our sample with the domain, the daily hours per active user, the number of users and the daily hours including non-users. All statistics are computed using the baseline period.

Figure C.2: Privacy Preferences and Beliefs across (Exogenous) Survey Payment



NOTES: This figure presents a comparison across the random survey offers presented to the experimental participants ($3, $4, $5). From left to right, we compare degree of belief misspecification for collection attributes, control attributes, use attributes, an index of the PEW questions, and the extension WTA. Bars represent 95% confidence intervals.

Figure C.3: PEW Comparison



NOTES: This figure compares the PEW values across the (i) sample of participants who completed the survey (denoted survey sample, blue), (ii) sample of participants who were in the extension subsample (red), and (iii) the values from PEW (green). The question in the left panel measures the fraction of participants who have ever stopped using a website due to privacy concerns. The question in the center panel measures how often participants self-report clicking "agree" right away without reading the contents of the privacy terms. The question in the right panel measures how effective the participants believe privacy policies are as a way for companies to communicate how they are using people's data. The content of the questions can be found in `https://www.pewresearch.org/wp-content/uploads/sites/20/2023/10/PI_2023.10.18_Data-Privacy_TOPLINE-1.pdf`.

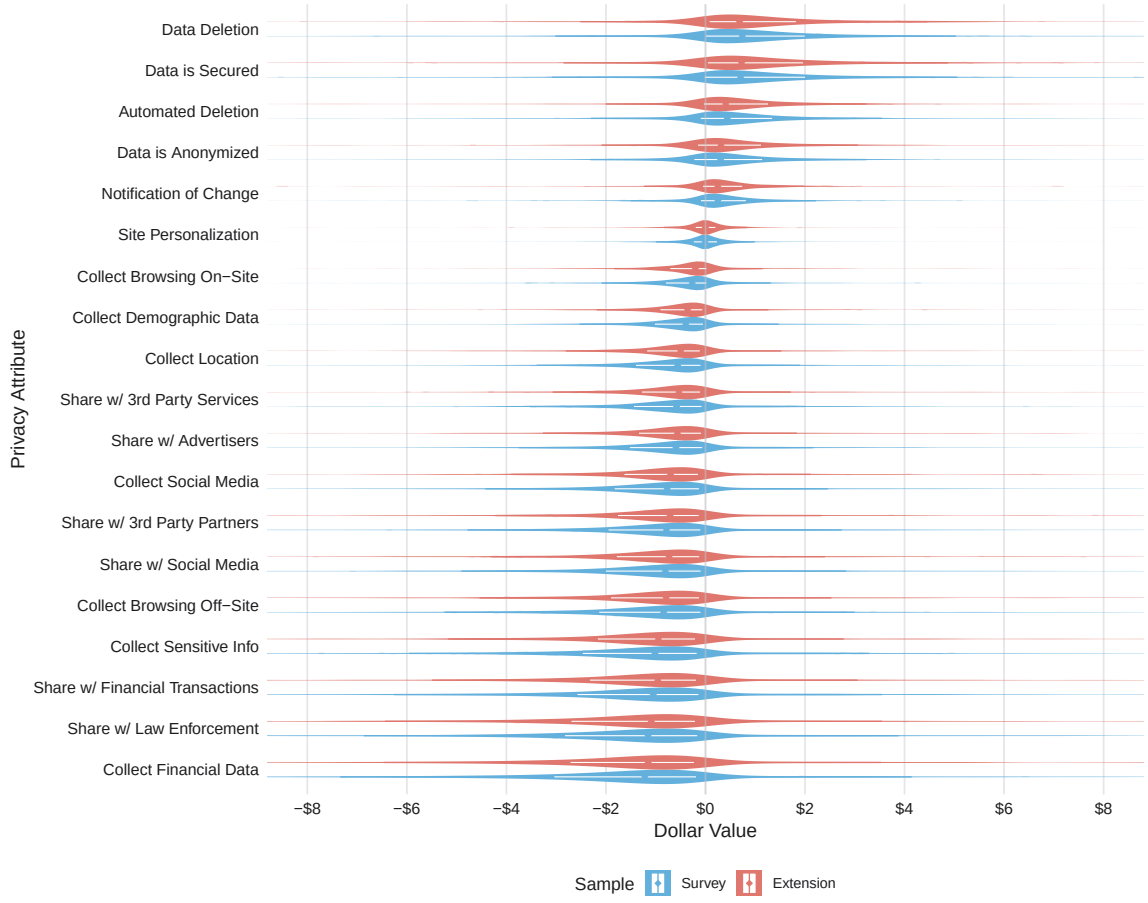
Table C.3: Changes in Experimental Behavior

Model:	Disrupted Browsing (1)	Evaded Extension/Chrome (2)
Constant	0.0281***	0.0208***
	(0.0091)	(0.0056)
Saliency Treatment	0.0963***	0.0171**
	(0.0129)	(0.0080)
Information Treatment	0.1132***	0.0239***
	(0.0127)	(0.0079)

IID standard-errors in parentheses
*Signif. Codes: ***: 0.01, **: 0.05, *: 0.1*

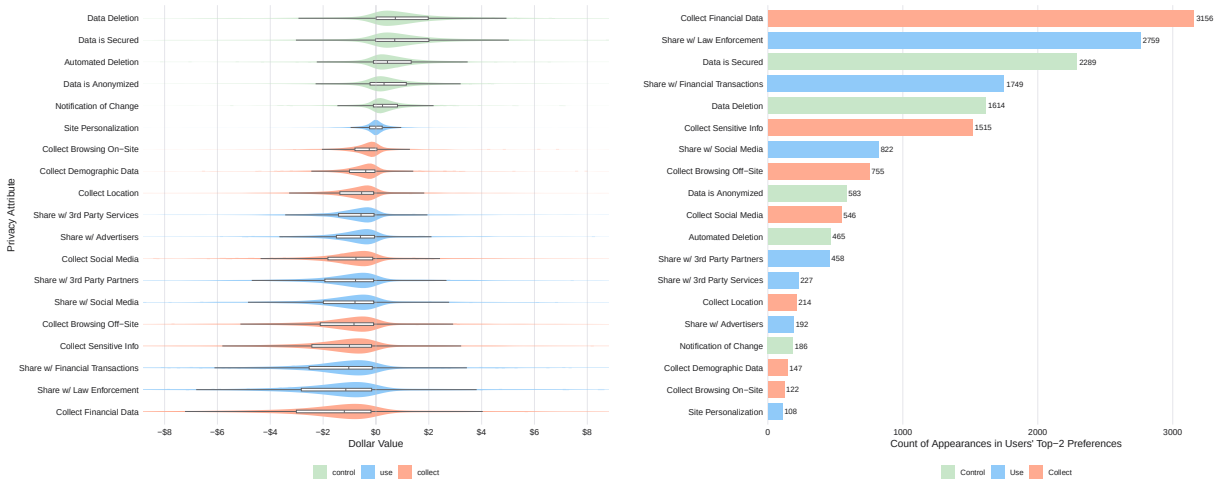
NOTES: This table presents the differences in self-reported changes in behavior due to the extension installation. We estimate OLS regressing the dependent variable on treatment status with clustered standard errors reported Column (1) reports differences to the following question: "How much did the experiment disturb your normal online browsing?" on a scale from 0 (Not at all) to 1 (Always) in increments of 0.25. Column (2) reports differences to the following question: "Did you take any actions to evade the browser extension? (e.g. browse more on your phone, or on a different browser)" on a scale from 0 (Not at all) to 1 (Always) in increments of 0.25. Robust standard errors are reported.

Figure C.4: Privacy Valuations between Extension and Survey Sample



NOTES: This figure presents a comparison of the estimated monthly willingness to pay for the different privacy attributes between the survey sample (blue) and extension sample (red).

Figure C.5: Privacy Valuations and Most Valued Information (Full Sample)



(a) WTP for Privacy Attributes

(b) Most Valued Information

NOTES: Both panels use the full conjoint sample, including all respondents who completed the baseline survey. Panel (a) presents the estimated monthly willingness to pay for different privacy attributes, computed using the estimated part-worth for each attribute and the price coefficient. Green rows denote "Control" attributes, blue rows denote "Use" attributes, and orange rows denote "Collect" attributes. Panel (b) shows the distribution of top-2 personalized attributes. Information value is computed per Equation (3). The extension sample version is in Figure 6.

Table C.4: Belief Misspecification and Privacy Preference Intensity

Dependent Variable:	Part-Worth (Aligned)		
	Discrete	Linear	Over/Under
Model:	(1)	(2)	(3)
Variables			
Underestimate	0.0159***		
	(0.0022)		
Overestimate	-0.0047*		
	(0.0025)		
Misspecification (pp)		-0.0001**	
		(5.05×10^{-5})	
Overestimate (pp)			9.29×10^{-5}
			(7.53×10^{-5})
Underestimate (pp)			-0.0003***
			(5.55×10^{-5})
Fixed-effects			
Attribute FE	Yes	Yes	Yes
Participant FE	Yes	Yes	Yes
Fit statistics			
Observations	141,094	141,094	141,094
R^2	0.59353	0.59335	0.59342
Within R^2	0.00057	0.00011	0.00029

Clustered (Participant FE) standard-errors in parentheses
*Signif. Codes: ***: 0.01, **: 0.05, *: 0.1*

NOTES: This table estimates the relationship between belief misspecification and privacy preference intensity. The dependent variable is the aligned part-worth, where positive values indicate stronger preferences to avoid the privacy practice. Misspecification is defined as the participant's belief (scaled 0–100) minus the ground truth prevalence (0–100). Column (1) uses discrete indicators for underestimation and overestimation relative to the rounded ground truth, with correct beliefs as the omitted category. Column (2) uses continuous misspecification in percentage points. Column (3) separates misspecification into overestimation and underestimation components. All specifications include attribute and participant fixed effects with standard errors clustered at the participant level.

Table C.5: Demographic Heterogeneity in Belief Misspecification and Preferences

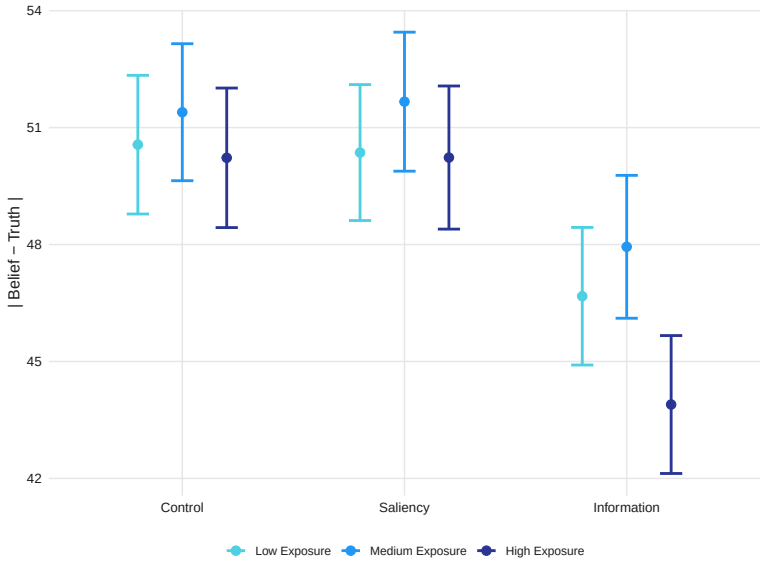
Dependent Variables:	—Misspecification— (pp)	Part-Worth (Aligned) Preferences
Model:	(1)	(2)
Variables		
Age 30–44	-0.0951	0.0018
	(0.1791)	(0.0074)
Age 45–59	-0.0086	0.0024
	(0.2045)	(0.0087)
Age 60+	0.3283	0.0229**
	(0.2677)	(0.0116)
Bachelor's Degree or Above	-0.9529***	0.0132**
	(0.1500)	(0.0064)
Household Income $\geq$ \$75k	-0.2298	-0.0071
	(0.1482)	(0.0063)
Female	0.0742	0.0566***
	(0.1437)	(0.0063)
Other/Prefer Not to Answer (Gender)	-0.8945	0.1083***
	(0.5895)	(0.0281)
Fixed-effects		
Attribute FE	Yes	Yes
Fit statistics		
Observations	139,232	138,130
R^2	0.21708	0.23605
Within R^2	0.00071	0.00628

Clustered (Participant FE) standard-errors in parentheses
*Signif. Codes: ***: 0.01, **: 0.05, *: 0.1*

NOTES: This table estimates demographic heterogeneity in belief misspecification and privacy preferences. Column (1) regresses absolute belief misspecification (in percentage points) on demographic covariates. Column (2) regresses the aligned part-worth on the same covariates. The omitted categories are age 18–29, no bachelor's degree, household income below \$75k, and male. All specifications include attribute fixed effects with standard errors clustered at the participant level.

C.2 Figures for Belief Updating and Search

Figure C.6: Heterogeneity based on (Ex Ante) Exposure - Tercile



NOTES: This figure presents the differences in beliefs about the most preferred privacy attributes in the endline survey compared to the ground truth. For each experimental group, we partition each (website, participant) tuple into terciles based on the time spent in the baseline period. We then plot the mean and associated 95% confidence interval for the estimate for each tercile and experimental group.

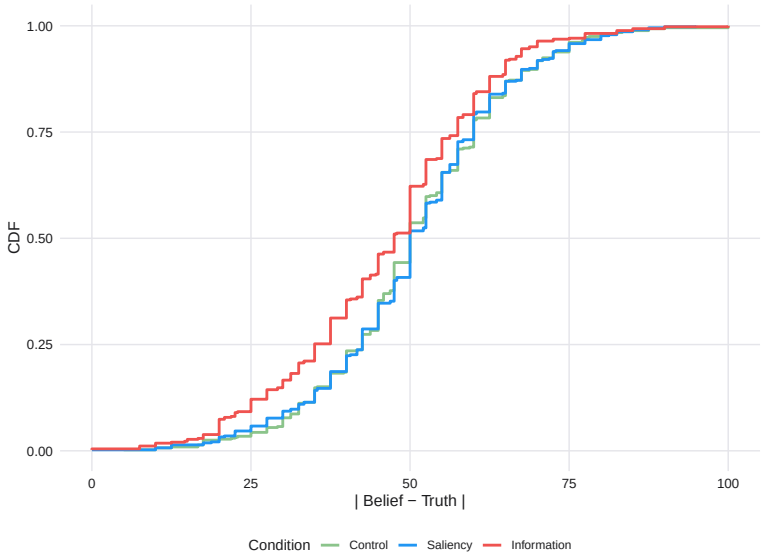
Table C.6: Belief Correctness on Top Sites – Random Information

Model:	|Belief − Truth| (1)	Correctness (Weak) (2)	Correctness (Strict) (3)
Saliency Treatment	2.065	-0.0454**	-0.0290
	(1.440)	(0.0226)	(0.0187)
Information Treatment	-2.671*	0.0395*	0.0417**
	(1.425)	(0.0216)	(0.0185)
Fixed-effects			
Website FE	Yes	Yes	Yes
Block FE	Yes	Yes	Yes

Clustered participant standard-errors in parentheses
*Signif. Codes: ***: 0.01, **: 0.05, *: 0.1*

NOTES: OLS estimates from Equation (3). Includes website, attribute, and randomization block fixed effects. Column (1): Dependent variable is $|\text{Belief}_{ijk} - \text{Truth}_{jk}|$, where beliefs are scaled to $\{0, 25, 50, 75, 100\}$ and truth is binary $\{0, 100\}$. Column (2): Binary indicator equal to 1 if response is correct, excluding neutral responses of 50. Column (3): Strict correctness, equal to 1 if $(\text{Belief} > 50) = (\text{Truth} = 100)$. The beliefs are the likelihood of the website engaging in the specified privacy practice for the top 5 most visited sites during the intervention and a random attribute that is not in the top 2 most informative attributes according to Equation (3).

Figure C.7: CDF of Belief Correctness



NOTES: This presents the CDF of belief correctness of the participant's belief regarding the website engaging in the specified privacy practice for the top 5 most visited sites during the intervention and that is in the top 2 most informative attributes according to Equation (3). For each participant we take the average across the two attributes and then compute the CDF across participants within each treatment group.

Table C.7: Treatment Effect on Cookie Banner Interactions

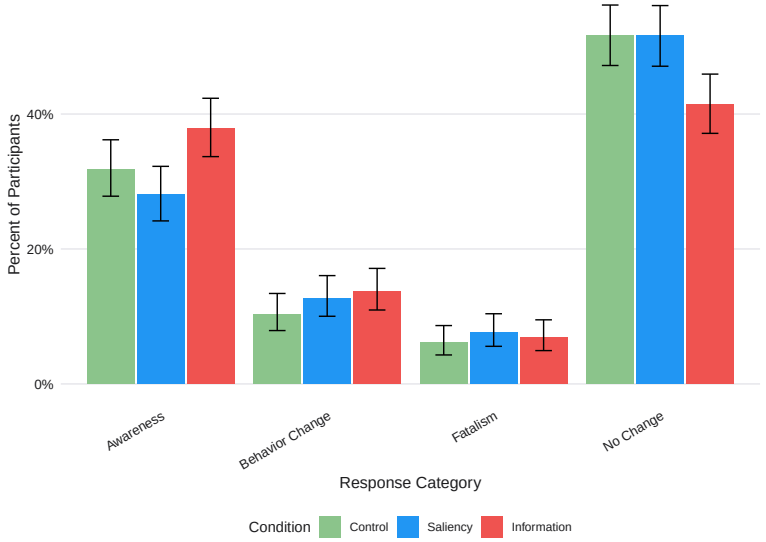
	Cookie Banner Total Interactions	Cookie Banner Distinct Domains
Model:	(1)	(2)
Saliency Treatment × Post	0.4158	0.1703*
	(0.4910)	(0.0881)
Information Treatment × Post	0.1106	0.0564
	(0.5082)	(0.0906)
Fixed-effects		
Weeks FE	Yes	Yes
Participant FE	Yes	Yes
Block FE	Yes	Yes

Clustered participant standard-errors in parentheses
*Signif. Codes: ***: 0.01, **: 0.05, *: 0.1*

NOTES: This table reports the estimated average treatment effects on engagement with a cookie banner. Engagements with a cookie banner were logged in the analytics system of our extension when a participant would mouse over the cookie banner. Column (1) presents the estimates using the dependent variable of the total number of cookie banner interactions and Column (2) presents the estimates using the dependent variable of the unique number of domains with at least one cookie banner interaction. The data is a balanced panel at the (site, participant) level across weeks. The difference-in-difference equation (4) is estimated for this dependent variable and estimates are clustered at the participant level.

C.3 Results on Changes in Behavior

Figure C.8: Self-reported Changes in Behavior



NOTES: Distribution of qualitatively coded free-form endline responses across four categories: Awareness (heightened awareness of privacy practices), Behavior Change (changes in site choice or data sharing), Fatalism (resignation that data collection is inevitable), and No Change. Responses were coded by three MTurk raters with researcher tiebreaking. The information group shows elevated awareness and behavior change relative to control; the saliency group shows patterns similar to or worse than control, with some participants expressing fatalism about privacy policy complexity.

Table C.8: Data Sharing Purposes

Consequence	Participants	Percentage	Category
Loss of data control	888	62.20	Negative
Personalized content	863	60.40	Positive
More annoying ads	713	49.90	Negative
Identity theft	711	49.80	Negative
Hacking	545	38.20	Negative
Better search results	469	32.80	Positive
Easier sharing/logins	411	28.80	Positive
Cyberstalking	280	19.60	Negative
Improved services	280	19.60	Positive
Coupons/discounts	223	15.60	Positive
More interesting ads	199	13.90	Positive

NOTES: Participants were asked to select the 3-5 most important consequences of data sharing online. The table presents the consequence, the number of participants who included it in their top 3-5 most important consequences and the overall percentage of participants that represents, followed by whether it is a positive or negative consequence.

Table C.9: Treatment Effect on Self-Reported Data Sharing Behavior

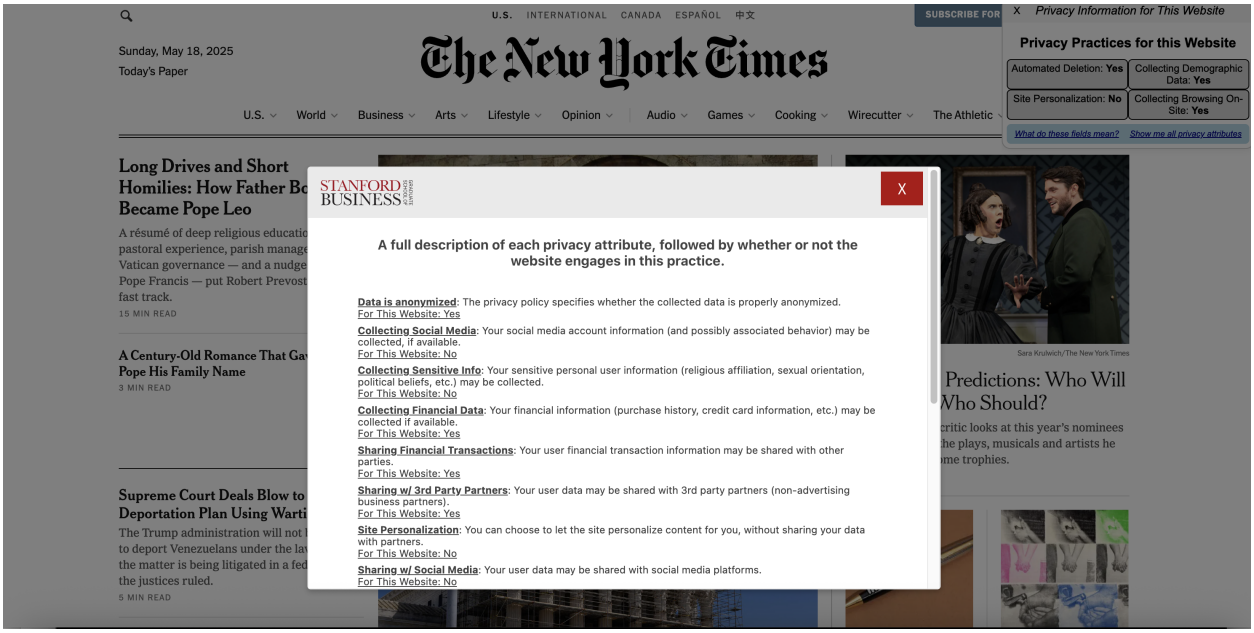
	Deleted Cookies	Opted out	Withheld data	Changed privacy settings
Model:	(1)	(2)	(3)	(4)
Information Treatment	0.018	0.069*	0.032	0.017
	(0.039)	(0.040)	(0.039)	(0.038)
Saliency Treatment	0.008	0.030	0.040	0.034
	(0.039)	(0.040)	(0.040)	(0.038)
Observations	995	995	995	995
R^2	0.344	0.363	0.360	0.369

Clustered participant standard-errors in parentheses
*Signif. Codes: ***: 0.01, **: 0.05, *: 0.1*

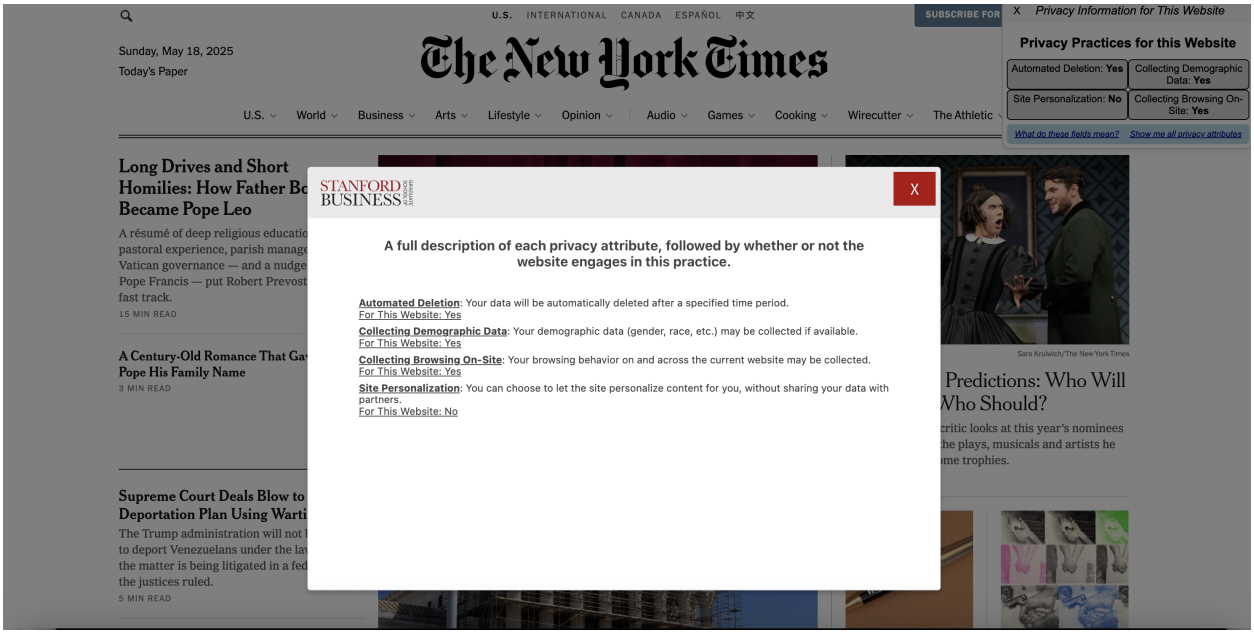
NOTES: This table reports the estimated average treatment effects on self-reported changes in data sharing behavior from the endline survey. The first column measures whether a participant self-reports deleting cookies on at least one website or at least once during the intervention period. The second column measures whether a participant self-reports explicitly opting out data sharing with at least one website. The third column measures whether a participant self-reports deciding not to share data with at least one website. The fourth column measures whether a participant self-reports changing privacy settings for at least one website. The data is only available for participants who went through the second wave of the study. We estimate OLS regressing the binary dependent variable on treatment status with clustered standard errors reported.

C.4 Experimental Design Figures

Figure C.9: Additional Information Dialogs



(a) Full Information



(b) Attribute Information

D Structural Model: Robustness

D.1 Model Fit on Experimental Moments

We evaluate the structural model's ability to reproduce the key experimental treatment effects it was estimated to match. Table D.1 compares model-predicted moments against their empirical counterparts across four classes of variation.

Table D.1: Model Fit on Experimental Moments

Moment	Metric	Observed	Predicted	Ratio
Site browsing shares	Pearson correlation	—	—	0.52
	Spearman rank correlation	—	—	0.81
Search rate (saliency arm)	Mean	0.171	0.157	0.92
Search rate (control arm)	Mean	≈ 0	≈ 0	—
Cookie deletion DiD	Mean $\Delta \log s_{ijt}$	-0.367	-0.306	0.83
Popup DiD (raw)	Mean $\Delta \log s_{ijt}$	$+0.012$	-0.063	-5.21
Popup DiD (salience-adj.)	Mean $\tilde{\delta}^{\text{popup}}$	—	—	—

Notes: "Ratio" is predicted/observed for scalar moments and the correlation coefficient for distributional moments. Site browsing shares compare model-predicted mean shares (under prior beliefs) against empirical pre-treatment shares across 489 sites. Search rates are the fraction of consumers searching at least one site's privacy policy. Cookie deletion DiD uses the early-deletion group (day 43) with the late group as control. Popup DiD (raw) is the unconditional mean change in log relative browsing share at the popup-treated site; the salience-adjusted version subtracts the Treatment 1 (popup-only) mean by type and is used in estimation (see text).

The model fits search rates and cookie deletion well (ratios of 0.83–0.92), and site-level browsing shares show strong rank-order agreement (Spearman 0.81) despite lower level correlation (Pearson 0.52), consistent with unobserved heterogeneity in site preferences beyond $\bar{Q}_j$.

Salience-differenced popup moments. The raw popup DiD conflates two channels: an *information* effect (belief updating changes expected utility) and a *salience* effect (receiving any popup increases top-of-mind awareness of the site). The structural model predicts only the information channel. Because the experiment includes a popup-only arm (Treatment 1: popup without privacy information) alongside the popup-plus-info arm (Treatment 2), we can difference out the salience channel. Specifically, we replace the dependent variable in the Block A moment conditions with $\tilde{\delta}_{ij}^{\text{popup}} = \delta_{ij,\,T=2}^{\text{popup}} - \bar{\delta}_{g,\,T=1}^{\text{popup}}$, where $\bar{\delta}_{g,\,T=1}^{\text{popup}}$ is the mean popup DiD for Treatment 1 consumers of type g. The Treatment 1 mean captures the type-specific attention effect of receiving a popup; subtracting it isolates the incremental effect of the privacy information that the model's belief-updating mechanism is designed to predict. The moment conditions $E[\tilde{\delta}_{ij}^{\text{popup}} - \hat{\delta}_{ij}^{\text{popup}}(\theta) \times Z_k \mid g] = 0$ retain

the same instrument $Z_k = (\mu^*_{jk} - \theta^0_{ijk}) \cdot \mathbf{1}_{\text{exposed}}$ and the same cross-attribute identifying variation, but the residual mean is now centered on the information effect rather than the composite treatment effect.

D.2 Alpha–Beta Decomposition Sensitivity

The popup DiD identifies the product $\hat{\alpha}^{\text{popup}}_g \times \hat{\beta}_{ik}$: the fraction of truth learned times the marginal utility of attribute k. The decomposition into separate $\hat{\alpha}$ and $\hat{\beta}$ relies on the conjoint moments (Block F), which pin down $\hat{\beta}^{\text{conj}}_{gk}$ up to the scaling factor $\hat{\xi}$. This section examines how sensitive the welfare counterfactuals are to the alpha–beta split by rescaling: $\tilde{\alpha}_g = f \cdot \hat{\alpha}_g$ and $\tilde{\beta}_{gk} = \hat{\beta}_{gk}/f$, which preserves the identified product while varying the decomposition.

Table D.2: Welfare Sensitivity to Alpha–Beta Decomposition

Factor f	Full Info dW	FI \$/month	GDPR dW	GDPR \$/month
0.50	0.2770	\$0.70	0.6219	\$1.57
0.75	0.3333	\$0.84	0.5471	\$1.38
1.00	0.3418	\$0.86	0.4814	\$1.22
1.25	0.3265	\$0.83	0.4254	\$1.08
1.50	0.3030	\$0.77	0.3790	\$0.96
2.00	0.2549	\$0.64	0.3091	\$0.78
3.00	0.1852	\$0.47	0.2240	\$0.57

Notes: Each row rescales $\hat{\alpha}_g \to f \cdot \hat{\alpha}_g$ and $\hat{\beta}_{gk} \to \hat{\beta}_{gk}/f$, preserving the $\alpha \times \beta$ product identified from popup DiD. $f = 1$ is the baseline. Higher f means consumers learn more per popup but have smaller privacy preferences; lower f means they learn less but care more per attribute. The Full Info column reports perceived-utility (naive) welfare gains; the Train-corrected baseline ($f = 1$) is 0.267 (\$0.67/month). GDPR welfare is unaffected by the Train correction since both sides use truth beliefs.

The key asymmetry is intuitive. The GDPR bundle operates through the direct channel—banning practices changes μ^*_{jk}, so welfare depends on $|\hat{\beta}_{gk}|$ (how much consumers dislike each practice) but not on $\hat{\alpha}$ (how much they learn). Scaling f down (larger $|\hat{\beta}|$) therefore increases GDPR welfare: from \$1.22/month at baseline to \$1.57/month at $f = 0.5$, a 29% increase. Scaling f up (smaller $|\hat{\beta}|$, higher $\hat{\alpha}$) reduces GDPR welfare: \$0.96/month at $f = 1.5$ and \$0.78/month at $f = 2.0$. Full information is non-monotone in f: it peaks near the baseline and declines in both directions, reflecting the tension between the belief-update magnitude (favored by large $\hat{\alpha}$, i.e., high f) and the utility response to those updates (favored by large $|\hat{\beta}|$, i.e., low f). Across the full range $f \in [0.5, 2.0]$, the GDPR estimate spans \$0.78–\$1.57/month and full information varies by a similar factor, confirming that while the product identification is sharp, the welfare magnitudes for regulation counterfactuals are moderately sensitive to the decomposition.

D.3 Type 3 Exclusion

Type 3 consumers (23% of the sample, $n = 342$) are estimated with near-zero popup comprehension ($\hat{\alpha}_3^{\text{popup}} = 0.005$), raising the concern that their welfare gains may be driven primarily by functional form extrapolation rather than empirically identified preference parameters. We assess this by recomputing each counterfactual excluding Type 3 consumers.

Table D.3: Welfare With and Without Type 3 Consumers

Counterfactual	All ($N = 1{,}489$)	Excl. Type 3	Type 3 Only	% Change
Full Information	0.2664	0.2527	0.3125	-5.2%
GDPR Bundle	0.4814	0.4778	0.4935	-0.8%
Targeted Disclosure	0.1842	0.1678	0.2394	-8.9%
Cookie Ban	-0.1246	-0.1221	-0.1330	$+2.0\%$

Notes: "% Change" is the change in mean ΔW from dropping Type 3 consumers. Negative values mean Type 3 was amplifying the effect (its mean exceeds the non-Type 3 mean); positive values mean Type 3 was dampening it. Type 3 is 23% of the sample ($n = 342$).

The GDPR bundle is essentially unchanged (-0.8%), confirming that Type 3 consumers are not disproportionately driving the headline regulation counterfactual. Their per-capita GDPR welfare gain (0.4935) is close to the sample mean (0.4814) because the GDPR bundle operates through the direct channel (banning practices), which depends on $\hat{\beta}_{gk}$ and site prevalence rather than comprehension $\hat{\alpha}$. Type 3's $\hat{\beta}$ values are broadly similar to other types—slightly larger in magnitude for most attributes, with the exception of personalization (the only attribute where the sign differs across types). Information counterfactuals show larger sensitivity: excluding Type 3 reduces full information by 5.2% and targeted disclosure by 8.9%. This is expected: Type 3 benefits disproportionately from information provision because its near-zero baseline comprehension means the gap between prior and truth is largest, generating the biggest reallocation gains. The cookie ban is nearly invariant to Type 3 exclusion ($+2\%$), as it operates through the $\hat{\delta}$ parameters that do not interact with type-specific comprehension.

D.4 Treatment Exposure Design

A potential concern for identification is that the information treatment may have been targeted—revealing attributes where the consumer's belief gap is largest—which would introduce selection into the popup DiD moments. We examine the exposure design using the 476 information-group consumers with recorded exposed attribute sets.

The modal number of exposed attributes is 17 out of 19 (near-complete disclosure), with a range of 14 to 18. No consumer receives information on all 19 attributes; the most commonly excluded

attributes are collecting financial data ($k = 2$) and sharing with law enforcement ($k = 14$). Importantly, the treatment is mildly *anti-targeted*: exposed attributes have a smaller mean belief gap ($|\mu^*_{jk} - \theta^0_{ijk}| = 0.453$) than non-exposed attributes (0.482), with the difference highly significant ($p < 10^{-14}$, paired t-test across consumer–site pairs). This pattern arises mechanically because the excluded attributes tend to be low-prevalence practices with large belief gaps (consumers tend to correctly believe most sites do not collect financial data).

To assess whether this anti-targeting biases the popup DiD estimates, we split popup events by the median belief gap of exposed attributes and test for differential effects. The mean popup DiD is statistically indistinguishable across high- and low-gap events ($p = 0.67$), indicating that the near-complete exposure and mild anti-targeting do not generate detectable heterogeneity in treatment effects. We also find no significant difference between the popup-only group (Treatment 1) and the popup-plus-info group (Treatment 2) in the raw popup DiD, consistent with the information treatment operating through belief updating rather than through differential salience.

E Survey Materials

The following provides a condensed description of the survey instruments used in the study. The study involved a short screener survey administered via Qualtrics to confirm eligibility, a longer baseline survey administered via Sawtooth Software that elicited privacy attitudes, beliefs, conjoint preferences, willingness-to-accept for the browsing tracker, and demographics, and an endline survey administered via Qualtrics that measured post-treatment beliefs, experience, and self-reported behavioral changes.

E.1 Screener Survey

The screener survey was distributed to participants recruited via Prolific and CloudResearch. It began with a consent page describing the study as examining browsing habits online and beliefs about data that firms collect online. Participants were informed that a follow-up survey would take 10–15 minutes, and that eligible respondents could have the opportunity to earn up to an additional $40 in a later stage of the study. The consent part of screener survey is shown in Figure E.1.

The screener collected a reCAPTCHA verification, a personal email address for follow-up contact, the respondent's Prolific or Connect ID, an email confirmation field, and whether Google Chrome was the respondent's main browser. Only Chrome users were eligible for the follow-up study. Eligible respondents were informed that they would receive a follow-up invitation by email within five minutes and that the follow-up survey had to be completed within one day.

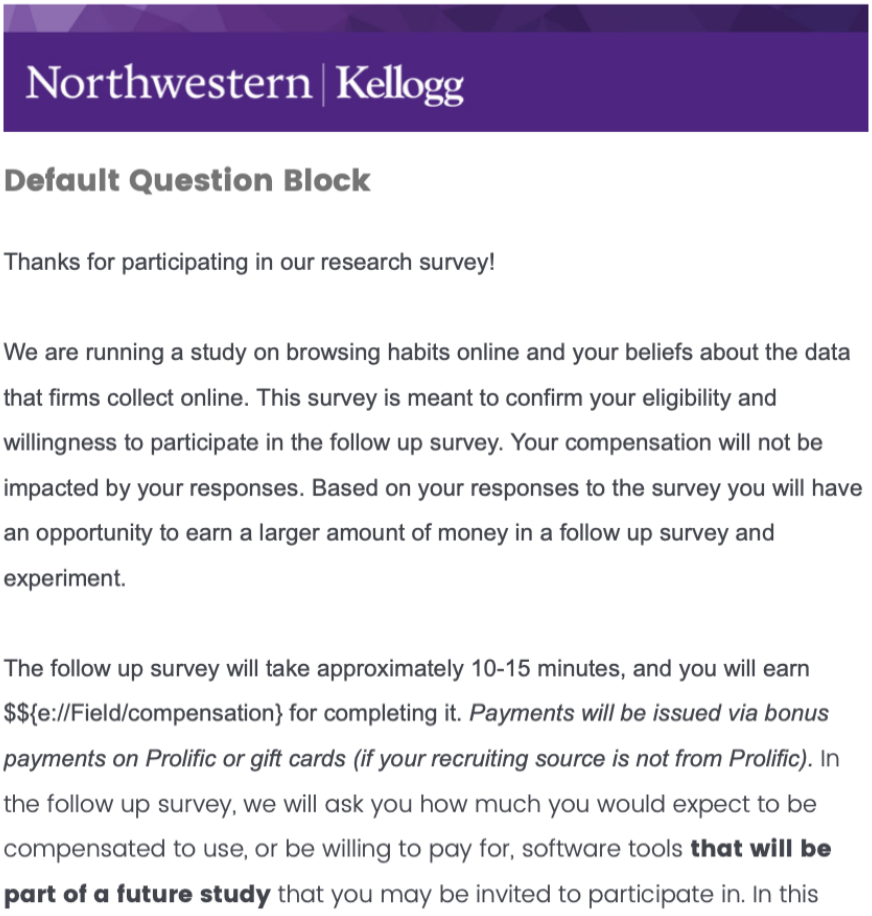


Figure E.1: Screener Survey: Consent Form

E.2 Baseline Survey

The baseline survey was administered via Sawtooth Software. It began with a consent form and collected the participant's email address and Connect/Prolific ID. It then proceeded through several blocks described below.

E.2.1 Privacy Attitudes

The survey began with several privacy-attitude questions, including items adapted from Pew Research Center surveys on privacy:

1. **Ad blocker usage.** "Do you use an ad blocker?" (Yes / No)
2. **Privacy policy effectiveness.** "How effective do you think privacy policies are as a way for companies to communicate how they are using people's data?" (Extremely effective / Very effective / Somewhat effective / Not too effective / Not at all effective)
3. **Privacy policy reading (PEW Q1).** "Thinking about privacy policies you might come across online or on your smartphone, how often do you click 'agree' right away, without reading what these policies say?" (Always or almost always / Often / Sometimes / Rarely / Never)
4. **Privacy policy effectiveness, repeated (PEW Q2).** Same as item 2, asked again after the reading-behavior question.
5. **Stopped using service (PEW Q3).** "Have you ever stopped using a digital device, website, or app because you were worried about how your personal information was being used?" (Yes, I have / No, I have not)
6. **Data collection awareness (PEW Q4).** "Let us know if you think the following statement is correct: When I go to a website, it can collect information about my online behaviors even if I don't register using my name or email address." (True / False / I don't know)
7. **Attention check.** "To ensure you are a real person, please select the vegetable from the list below." (Egg / Salmon / Broccoli / Cheeseburger / Pizza / Milk)

The survey also asked participants to select the 3 to 5 most important consequences of websites sharing their data online. The response options included both negative consequences (loss of data control, more annoying ads, identity theft, hacking, cyberstalking) and positive consequences (personalized content, better search results, easier sharing and logins, improved services, coupons and discounts, more interesting ads).

E.2.2 Beliefs About Data Practices

This block asked participants to state their beliefs about data practices of a randomly selected top-500 website, using a 5-point scale (0% / 25% / 50% / 75% / 100%).

Data collection beliefs. "How likely do you think this site is to **collect** the following data?"

- Browsing behavior on the site
- Demographic data (gender, race, income, etc.)
- Sensitive personal information (religious affiliation, sexual orientation, political beliefs, etc.)
- Financial data (purchase history, credit info, etc.)
- Browsing behavior on *other* sites
- Location data
- Social media profiles and presence

Data usage beliefs. "How likely do you think this site is to **use** your data in the following way?"

- Share data only after being anonymized (protecting user identity)
- Share with social media websites
- Share with financial providers
- Share with advertisers
- Share with law enforcement or government
- Share with service providers (security, analytics, etc.)
- Share with non-service providers (partner websites, data brokers, etc.)
- Use data to personalize your experience (such as search results or content)

Data control beliefs. "How likely do you think this site is to **allow you to control** the following data practices?"

- Notification of a material change to policy
- Automatic deletion of data after a specified time period
- Ability to delete data upon request
- Secure and anonymized data storage

Site quality beliefs. "How likely do you think this site exhibits the following characteristics?"

- High quality content
- Obstructive or annoying advertising
- Easy to use and navigate

E.2.3 Category-Level Beliefs

Participants were asked to rate how likely a randomly selected top site from each of four categories was to engage in each of: many data **collection** practices, many data **usage** practices, and many data **control** practices. The four categories were:

- E-commerce (Amazon, Walmart, Target, etc.)
- Entertainment (Netflix, YouTube, etc.)
- Social Media (Facebook, Reddit, X, etc.)
- News (Fox, CNN, New York Times, Wall Street Journal, etc.)

Responses were on a dropdown scale.

Participants also ranked a subset of ten site features, drawn from the collection, usage, control, and quality categories described above, in order of importance (1 = most important, 10 = least important).

E.2.4 Conjoint Analysis

Favorite website and usage. As the first step of the conjoint module, each participant was randomly assigned to one of four website categories (E-commerce, Entertainment, Social Media, or News) and asked to name their favorite website in that category (in standard format, i.e. domain.com). Participants also reported their usage frequency for this website (Once or twice a month / Every week or so / Multiple times per week / Every day or almost every day / Many hours (or multiple times) per day). The stated favorite website was used as the reference product throughout the conjoint exercise. Throughout this module and elsewhere in the survey, "[Category]" and "[Website]" refer to these participant-specific values.

Introduction and example. The conjoint section then presented participants with a detailed introduction explaining the choice task. Participants were told they would see four hypothetical versions of their favorite website, each with different privacy features and a monthly price, and should choose which version they would want to use. They were instructed to assume that all

versions were identical on all data practices *except* for those explicitly shown. This assumption is central to interpreting the partial-profile design, in which only a randomly selected subset of attributes varies within each choice task while all non-displayed attributes are held constant across alternatives.

Training and comprehension checks. Before the main conjoint tasks, participants were shown an example choice table with 8 attributes (3 collection, 1 usage, 1 control, 2 quality, and price) and 4 website versions. To ensure that participants understood both how to read the choice table and the assumption underlying the partial-profile design (that non-displayed attributes are held constant across alternatives), they answered two comprehension questions:

1. "What attribute does version 1 of [Website] have, that no other version does?"
2. "Which of these versions of [Website] engage in the collection of your financial data?"

Main conjoint tasks. Participants completed 12 conjoint choice tasks. Each task presented four website versions and a "NONE: I wouldn't choose any of these" outside option. Each version was characterized by six randomly selected attributes from the set below, plus price ($1–$5/month). Each attribute took a value of "Yes" (the website engages in the practice) or "No" (it does not). The full set of conjoint attributes is:

- *Collection (7):* what you browse onsite; your demographics (gender, race, age, etc.); your sensitive personal data (religion, political beliefs, sexuality, etc.); your financial information (purchase history, credit card info, etc.); what you do off-site (interactions with other sites); your location; your social media presence and profiles.
- *Usage (8):* data shared with social media; shared with financial providers; shared with advertisers and used for targeted advertising; shared with law enforcement; shared with service partners (security, analytics, etc.); shared with non-service partners (partner sites, data brokers, etc.); only shared after being anonymized; used to personalize your experience.
- *Control (4):* notified in case of material change to site policies; data automatically deleted after a time period; ability to delete your data; data stored securely and anonymized.
- *Quality (2):* high quality content; obstructive or annoying advertising.
- *Price:* $1, $2, $3, $4, or $5 per month.

An additional attention check ("Please select the answer '2' below") was inserted after the 4th conjoint task.

E.2.5 Demographics

Standard demographic questions were asked:

1. **Age.** Open-ended (in years).
2. **Education.** 10 levels from "Grade school" to "Doctorate degree."
3. **Gender.** Male / Female / Prefer not to answer.
4. **Household income (2024).** 12 brackets from "Less than $10,000" to "Greater than $200,000," plus "Prefer not to answer."
5. **Race.** Check all that apply: White / Black or African American / American Indian or Alaska Native / Asian / Native Hawaiian or Other Pacific Islander / Prefer not to answer.

E.2.6 Willingness-to-Accept (WTA) Elicitation

The final block used an incentive-compatible mechanism to elicit participants' minimum willingness-to-accept for installing a browsing tracker extension on their computer for up to 2 months. Participants were told: "The software tracks the identity of the websites you visit, how long you spend there and how much data is sent to 3rd parties. Beyond this, the software does not observe what you do on a website."

Mechanism. Participants were presented with a series of offers ($0, $1, $5, $10, $15, $20, $25, $30, $35, $40, or "Other") and asked to select a cutoff point representing their minimum acceptable payment. One offer was then randomly selected; if the selected offer was at or above the participant's cutoff, they were invited to the second stage and compensated at the selected offer amount. If the selected offer was below the cutoff, they were not invited. Participants were told that truthful reporting was optimal under this procedure.

Comprehension checks. Before stating their actual WTA, participants answered three comprehension questions about a hypothetical response in which the cutoff was set at $20:

1. "What would happen if the chosen offer was $30?" (Correct answer: the participant would be invited and compensated $30.)
2. "What would happen if the chosen offer was $15?" (Correct answer: the participant would *not* be invited.)
3. "What would happen if the chosen offer was $20?" (Correct answer: the participant would be invited and compensated $20.)

Participants who answered incorrectly were shown the instructions again and given a second attempt. Those who failed the comprehension checks were not eligible to be invited to the second stage.

E.3 Endline Survey

The endline survey was administered via Qualtrics after the intervention period. Its primary purpose was to measure what participants learned during the study and whether the experiment changed their privacy-related beliefs or behavior.

E.3.1 Post-Treatment Beliefs

The first substantive block confirmed study completion, extension uninstallation, and validated the respondent via reCAPTCHA. The survey then elicited beliefs about the privacy practices of the respondent's most visited websites during the intervention period. For each of the participant's top two personalized privacy attributes (selected using Equation (3) in the main text), participants rated how likely each of their five most visited sites was to engage in that practice, using the same 5-point scale (0%/25%/50%/75%/100%) as the baseline. A third question repeated this exercise for a randomly selected attribute outside the participant's top two, providing a control for attribute-specific learning.

E.3.2 Experience and Reactions

The next block asked about participants' experience in the study:

- "Did you find the privacy information helpful?" (five-category response scale including "I did not receive information")
- "How often did the privacy information surprise you?" (same scale)
- "How much did the experiment disturb your normal online browsing?" (5-point scale from "Not at all" to "Always")
- "Did you take any actions to evade the browser extension? (e.g. browse more on your phone, or on a different browser)" (same scale)

Participants were also asked to list their three most valuable websites.

E.3.3 Privacy Knowledge and Attitudes

The endline re-administered the same privacy-knowledge and privacy-attitude items from the baseline (Section E.2.1), including awareness of passive data collection, whether the respondent had

stopped using a service due to privacy concerns, frequency of agreeing to policies without reading, and perceived effectiveness of privacy policies. In addition, the endline included questions on online profiling: participants were told that companies combine data from purchasing histories, browsing behavior, and public records to create detailed profiles, and were asked how much they had previously heard about this practice and how many companies they believed used such profiles.

This block also re-elicited broader privacy beliefs about the prevalence of collection, usage, control, and quality practices using the same attribute set, response scale, and category-level questions as in the baseline (Sections E.2.2 and E.2.3). The survey also included a closely related question on the consequences of data sharing, paralleling the corresponding baseline item. Attention checks were included.

E.3.4 Risk Preferences

The survey elicited risk preferences using an incentivized lottery task following Charness et al. (2013) and Dave et al. (2010). Participants selected one of six 50/50 payoff pairs:

1. ($28, $28) (safe option)
2. ($24, $36)
3. ($20, $44)
4. ($16, $52)
5. ($12, $60)
6. ($2, $70) (riskiest option)

A ten-sided die determined which payoff was realized, and five participants were randomly selected to receive the resulting bonus payment. The chosen gamble indexes the participant's risk tolerance, with higher-numbered gambles corresponding to lower risk aversion.

E.3.5 Self-Reported Behavioral Changes

The final block measured self-reported behavioral responses. For each of the following practices, respondents indicated whether they had made changes as a result of the study (Yes, on several websites / Yes, on one website / No):

- Deleted cookies
- Opted out of third-party data sharing
- Withheld data due to privacy concerns

- Changed privacy settings

A final open-ended question asked: “Has this experiment changed your view on privacy online? Have you changed any online behaviors in response to this experiment?”

Note: The original survey instruments contain repeated instructions, validation logic, piped text, and additional attention checks beyond those described here. This appendix summarizes the main substantive blocks. Full survey instruments are available upon request.