DISCUSSION PAPER SERIES

DP15072

PRIVACY PROTECTION, SECURITY, AND CONSUMER RETENTION

Bruno Jullien, Yassine Lefouili and Michael Riordan

INDUSTRIAL ORGANIZATION

ISSN 0265-8003

PRIVACY PROTECTION, SECURITY, AND CONSUMER RETENTION

Bruno Jullien, Yassine Lefouili and Michael Riordan

Discussion Paper DP15072
Published 20 July 2020
Submitted 18 July 2020

Centre for Economic Policy Research
33 Great Sutton Street, London EC1V 0DX, UK
Tel: +44 (0)20 7183 8801
www.cepr.org

PRIVACY PROTECTION, SECURITY, AND CONSUMER RETENTION

Abstract

A website monetizes information it collects about its customers by charging third parties for targeted access to them. Allowing for third parties who are well-intentioned, a nuisance, or even malicious, the resulting consumer experiences might be good, bad, or neutral. As consumers learn from experience, the website especially risks losing those customers who su¤er a bad experience. Customer retention thus motivates the website to be cautious about monetization, or to spend resources to screen third parties. We study the website?s equilibrium privacy policy, its welfare properties, competition in the market for information, and the impact of regulations improving transparency and consumer control.

JEL Classification: D83, L15, L51

Keywords: Privacy Policy, Consumer Retention, Personal Data, regulation

Bruno Jullien - bruno.jullien@tse-fr.eu
Toulouse School of Eonomics and CEPR

Yassine Lefouili - yassine.lefouili@tse-fr.eu
Toulouse School of Eonomics

Michael Riordan - mhr21@columbia.edu
Columbia University

Acknowledgements
?We are grateful to Simon Anderson, Dirk Bergemann, Yeon-Koo Che, Jacques Crémer, Gabrielle Demange, Hanna Ha- laburda, Oliver Hart, Christian Hellwig, Justin Johnson, Louis Kaplow, Martin Peitz, Georgios Petropoulos, Mike Powell, Andy Skrzypacz, Kathryn Spier, Roland Strausz, Curtis Taylor, Jean Tirole, Hal Varian, Alex White, and our audiences at Harvard, Columbia, Duke, PSE, CREST, TSE, the University of Melbourne, the University of Tel Aviv, the University of Los Andes, the University of Montpellier, the MACCI Summer Institute (Schloss Gracht), the 9th Bi-Annual TSE Postal Economics Conference (Toulouse), the Workshop on the Digital Economy (Paris), EARIE 2016 (Lisbon), the Annual Meeting of the Toulouse Network for Information Technology 2016 (Seattle), the APIOC 2016 (Melbourne), the Northwestern-Toulouse IO Workshop 2017 (Evanston), the 2017 HKUST Workshop on Industrial Organization (Hong-Kong), the Workshop on In- stitutions, Individual Behavior and Economic Outcomes (Alghero), CRESSE 2017 (Crete), the Gilbert Center for Applied Economics Conference (Berkeley), the Workshop on the Economics of Platforms 2017 (Barcelona), the 10th TOI Workshop on Industrial Organization (Maitencillo), the UIBE Workhop on IO and Competition Policy (Beijing), ESSET 2018 (Gerzensee), the Luohan Conference on Privacy and Data Governance (Hangzhou), and the 4th Conference of the French Law and Economics Association (Rennes) for useful comments and discussions. The nancial support of the European Research Council (ERC) under the European Union?s Horizon 2020 research and innovation programme (grant agreement No 670494) is gratefully acknowledged. Jullien and Lefouili acknowledge funding from the Agence Nationale de la Recherche under grant ANR-17- EURE-0010 (Investissements d?Avenir program) and the nancial support of the TSE Digital Center, whose sponsors are listed on https://www.tse-fr.eu/digital?tabs=1.

Privacy Protection, Security, and Consumer Retention*

Bruno Jullien† Yassine Lefouili‡ Michael H. Riordan§

June 1, 2020

Abstract

A website monetizes information it collects about its customers by charging third parties for targeted access to them. Allowing for third parties who are well-intentioned, a nuisance, or even malicious, the resulting consumer experiences might be good, bad, or neutral. As consumers learn from experience, the website especially risks losing those customers who suffer a bad experience. Customer retention thus motivates the website to be cautious about monetization, or to spend resources to screen third parties. We study the website's equilibrium privacy policy, its welfare properties, competition in the market for information, and the impact of regulations improving transparency and consumer control.

Keywords: Privacy Policy, Consumer Retention, Personal Data, Regulation.

JEL Classification: D83, L15, L51.

*We are grateful to Simon Anderson, Dirk Bergemann, Yeon-Koo Che, Jacques Crémer, Gabrielle Demange, Hanna Halaburda, Oliver Hart, Christian Hellwig, Justin Johnson, Louis Kaplow, Martin Peitz, Georgios Petropoulos, Mike Powell, Andy Skrzypacz, Kathryn Spier, Roland Strausz, Curtis Taylor, Jean Tirole, Hal Varian, Alex White, and our audiences at Harvard, Columbia, Duke, PSE, CREST, TSE, the University of Melbourne, the University of Tel Aviv, the University of Los Andes, the University of Montpellier, the MACCI Summer Institute (Schloss Gracht), the 9th Bi-Annual TSE Postal Economics Conference (Toulouse), the Workshop on the Digital Economy (Paris), EARIE 2016 (Lisbon), the Annual Meeting of the Toulouse Network for Information Technology 2016 (Seattle), the APIOC 2016 (Melbourne), the Northwestern-Toulouse IO Workshop 2017 (Evanston), the 2017 HKUST Workshop on Industrial Organization (Hong-Kong), the Workshop on Institutions, Individual Behavior and Economic Outcomes (Alghero), CRESSE 2017 (Crete), the Gilbert Center for Applied Economics Conference (Berkeley), the Workshop on the Economics of Platforms 2017 (Barcelona), the 10th TOI Workshop on Industrial Organization (Maitencillo), the UIBE Workhop on IO and Competition Policy (Beijing), ESSET 2018 (Gerzensee), the Luohan Conference on Privacy and Data Governance (Hangzhou), and the 4th Conference of the French Law and Economics Association (Rennes) for useful comments and discussions. The financial support of the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation programme (grant agreement No 670494) is gratefully acknowledged. Jullien and Lefouili acknowledge funding from the Agence Nationale de la Recherche under grant ANR-17-EURE-0010 (Investissements d'Avenir program) and the financial support of the TSE Digital Center, whose sponsors are listed on https://www.tse-fr.eu/digital?tabs=1.

†Toulouse School of Economics, CNRS, Toulouse, France. E-mail: bruno.jullien@tse-fr.eu.

‡Toulouse School of Economics, University of Toulouse Capitole, Toulouse, France. E-mail: yassine.lefouili@tse-fr.eu.

§Columbia University, New-York, USA. E-mail: mhr21@columbia.edu.

1 Introduction

The internet has transformed how consumers shop. Sophisticated online platforms enable advertisers to target relevant ads to consumers, but, absent reliable safeguards, might also be a vehicle for cybercrime or unwanted targeted advertising. Malicious advertising (or "malvertising") is particularly nefarious because it degrades the fundamental business model for advertising-supported websites. Successful malvertising places deceptive ads that make disingenuous offers, install malicious code (e.g. ransomware), or phish for personal information (e.g. passwords).[1] Such risks discourage online commerce, or encourage protective measures that reduce website revenue (e.g. ad blockers).

Consider the following incident highlighted in an early U.S. Senate report on privacy hazards of online advertising:[2]

"In September 2009, the New York Times sold advertising space on its website using both third-party ad networks and direct sales. An advertiser claiming to represent the Internet telephony company Vonage contacted the New York Times offering to purchase advertising space on NYTimes.com. Vonage had previously run advertisements through the New York Times, so the newspaper allowed a third-party vendor it was unfamiliar with to actually deliver the ad. For several weeks, the advertiser submitted wholly legitimate-looking advertisements, which the New York Times ran without incident. Then, at the beginning of a weekend, the advertiser replaced the Vonage advertisements with an ad proclaiming that the user's computer was not safe, and that the user should purchase fake antivirus software to protect her computer. That fake antivirus software, once placed on a user's computer, could steal personal data and extort money from consumers hoping to make the virus go away."

This relatively simple case of malvertising illustrates several issues. First, even mainstream websites can expose consumers to significant cyber risks. Consumers who visited the Times website and clicked on the fake ad downloaded malicious code that potentially exposed them to data theft and extortion. In the case of a "drive-by download", however, the consumer doesn't even have to click on the ad to become infected by malware.[3] Second, a website can better protect its consumers by exercising precaution in its dealing with third parties. The Times acknowledged that it might have prevented the attack by vetting the third-party who delivered the fake Vonage ad, and promised that in the future it would

[1] See RiskIQ (2016) and U.S. Senate Committee on Homeland Security and Governmental Affairs (2014).

[2] U.S. Senate Committee on Homeland Security and Governmental Affairs (2014).

[3] See, for example, Danny Palmer, "This malvertising campaign infected PCs with ransomware without users even clicking a link," AZNEt, June 21, 2017, https://www.zdnet.com/article/this-malvertising-campaign-infected-pcs-with-ransomware-without-users-even-clicking-a-link/

not allow its advertisers to use unfamiliar third-party vendors.[4] While the Times incident involved direct sales advertising for which proper vetting might be relatively easy, malvertising is especially difficult to stop for targeted ads served by advertising networks, as these typically involve numerous third parties over which the website has little or no oversight, and target particular consumers whom the website may fail to identify before it is too late.[5] Third, consumers may be unsure of their vulnerability to cyberattacks. The malware attack on the Times only threatened computers running Windows.[6] More generally, however, malware can exploit many different software vulnerabilities about which consumers may have limited or no awareness. Finally, concerns about cyber risks might reduce website traffic. In its own reporting on the malvertising attack, the Times acknowledged that such events can damage a website's reputation and "make its visitors nervous."[7]

Motivated by such issues, we develop a theory of privacy protection for an environment in which consumers learn from experience about their utility of visiting a website, which depends both on the consumer's value of website content and on the consumer's vulnerability to intrusions. Under these conditions a website's privacy policy affects consumer retention by altering consumer experience and thus consumer learning. In our model, a website offers a free service and earns revenue from banner advertising (or another activity that doesn't compromise privacy). The website also collects information about its customers that it can use to charge third parties for targeted access to them (or profit from some form of marketing that raises privacy concerns). Such transactions with third parties could be beneficial to consumers, for example, by enabling targeted advertising that informs consumers of desirable products, or intrusive, for example, by increasing exposure to spam, phishing, or malware.[8] Those customers experiencing intrusions become more pessimistic about their overall utility from a return visit to the website, and this learning mechanism gives the website an incentive for a privacy policy that limits third-party transactions in order to protect at least partially its customers from intrusion and thereby improve customer retention.

More precisely, we study a simple two-period model. In the first period, a population of consumers enjoy a free service provided by the website.[9] The service is an experience good,

[4]David Gallagher , "Times site was victim of a malicious ad swap," New York Times, September 14, 2009, https://bits.blogs.nytimes.com/2009/09/14/times-site-was-victim-of-a-malicious-ad-swap/

[5]Ashlee Vance, "Times Web Ads Show Security Breach," New York Times, September 14, 2009, https://www.nytimes.com/2009/09/15/technology/internet/15adco.html.

[6]Riva Richmond, "What to do if you saw an 'antivirus' pop-up ad," New York Times, September 14, 2007, https://gadgetwise.blogs.nytimes.com/2009/09/14/what-to-do-if-you-saw-an-antivirus-pop-up-ad/.

[7]Op. cit., Gallagher.

[8]For simplicity, we assume banner ads pose no risk to consumers. What is important for our purposes is that targeted ads poses a greater risk of a malicious intrusion. This is a natural assumption because, as noted above, banners ads are much easier to vet.

[9]This baseline model can be interpreted as examining a particular cohort in an overlapping generation

for which the consumers have heterogeneous values in the second period. The website also sells a matching service to third parties that provides targeted access to consumers. The website's privacy policy is a choice of "precaution", determining the probability that the website does not sell the matching service to an interested third party in the first period. Consumers do not directly observe the website's choice of precaution, and instead form equilibrium beliefs. Matching with a third party results in a consumer experience that may be good, bad, or neutral; a neutral experience is the same as if there is no third-party match. Consumers are unsure of their vulnerability, defined as the probability of a bad experience. In the first period, consumers have identical prior beliefs about vulnerability, and, in the second period, consumers use Bayes Rule to update their beliefs about vulnerability based on their first-period experiences. The consumers optimize whether to return to the website, given their realized utility value of the website service and their posterior beliefs of vulnerability. An equilibrium is a profit-maximizing level of precaution and consumer posterior beliefs (determining their willingness to make a return visit) that are mutually consistent.

Equilibrium is well behaved in our baseline model with a single website or, equivalently, with single-homing consumers. Website precaution in the first period is decreasing in the first-period value of third-party sales relative to the second-period value of retaining customers. In a full-precaution regime, this relative value is sufficiently low that the website does not offer to match third parties with its customers. Conversely, in a no-precaution regime, the relative value is sufficiently high that the firm sells the matching service to all interested third parties. There is also an intermediate partial-precaution region, in which precaution is decreasing in the relative value. The website provides no precaution in the second period, because there is no future relationship with returning consumers.[10]

We then extend our analysis to the case of multiple websites serving multi-homing consumers, thus providing a novel model of competition in the market for information. We find that there always exists an equilibrium in which all websites exert no precaution but there may also exist an equilibrium that is qualitatively similar to the one obtained in our baseline model. We also show that competition reduces the price of the matching service and that there is less precaution than in the baseline model if the market for information is a major source of revenues for websites.

A website's equilibrium incentive for precaution is at best only imperfectly aligned with consumer welfare. This is not surprising because consumers cannot verify website precau-

model that crudely distinguishes between "young" and "old" consumers.

[10]Thus the two-period model captures starkly the intuitive idea that the website has a greater incentive to protect patrons with a higher customer lifetime value (CLV), which is the profit attributed to the entire future relationship.

tion. Consumer short-term (i.e. first period) welfare increases with precaution if their expected utility from third-party matching ("match utility") is negative, while long-term (i.e. second period) consumer welfare also decreases with precaution due to less informative learning about vulnerability. We show in our baseline model that the website would choose less precaution than the equilibrium level if it could commit to first-period precaution, assuming that no precaution in the second period remains profit-maximizing. The website, however, may find it optimal to commit to positive second-period precaution if that sufficiently improves consumer retention.

Robust welfare-improving regulations are not readily apparent. We examine the effects of policies improving transparency and consumer control such as those included in the European Union General Data Protection Regulation (GDPR) and the California Privacy Act. We show that a transparency policy that requires *ex post* disclosure of matching service sales leads to less precaution, which would benefit consumers if their expected utility from third-party matching is positive but would have mixed effects otherwise. We also examine a transparency policy that requires a website to commit *ex ante* to a minimum level of precaution. Under conditions implying the website always sells targeted access to returning customers, such policy is inconsequential in our baseline model with a single website because the website wants to commit to less, rather than more, precaution. The same holds in the setting with multiple websites provided the value of the first period matching service is sufficiently low. We then consider an opt-out rule allowing customers to refuse permission for the website to use their personal information for third-party matching. In the most interesting scenario, in which consumers opt out in the second period if and only if they have a bad experience in the first period, allowing firms to offer opt-out is always consumer surplus improving in the setting with a single website, but may be detrimental to consumers in an environment with multiple websites serving multi-homing consumers. When opt-out is not offered on a voluntary basis, a mandatory opt-out policy leads to more precaution. An opt out-rule necessarily improves consumer welfare in the second period by revealed preferences, but greater precaution in the first period is beneficial only if, given prior beliefs about vulnerability, expected match utility is negative.

We study three extensions of our model. These extensions add positive insights about website incentives for privacy protection, but do not reverse our normative conclusion about the difficulty of designing robust welfare-improving regulations. First we allow for costly verification that third-party uses of personal information are benign, enabling the website to prevent bad consumer experiences. The website's strategy is then given by the level of precaution and level of verification. We characterize the equilibrium when the website cannot

commit to its strategy, and show in particular that the equilibrium level of verification is non-monotonic in the value of third-party matching. We also show that verification and precaution are substitutes for the website, so that the ability to verify third-parties would reduce the level of precaution. The other two extensions relax some of the simplifying assumptions made in our baseline model. One of them assumes that consumers not only learn about the probability of a bad experience but also about the probability of a good one, and the other one considers an elastic demand in the first period.

Related literature. The economics of privacy literature echoes various themes from the broader information economics literature (Acquisti, Taylor and Wagman, 2016). For example, the disclosure of personal information can improve the allocation of goods and services via targeted advertising or price discrimination, while secrecy potentially leads to market failure due to adverse selection or costly signaling. We contribute to the literature by developing a neglected theme: website privacy policy influences how consumers learn about their tastes for a product attribute. In our model, consumers care both about their direct utility from website services, and their match utility from third party sales. Website privacy protection in essence is a product attribute, the value of which consumers learn imperfectly from experience. The website chooses privacy protection with the aim of influencing consumer beliefs, but, as is typical of signal-jamming models, consumers in equilibrium see through these incentives, and correctly predict the firm's actions.[11]

Our work is related to the paper by Spiegel (2013) on a software producer's choice between charging consumers for the software and offering for free a bundle of the software and ads. In his model, however, consumers are perfectly informed about the utility they derive from an impression and the firm's strategy is observable to them. In a similar vein, O'Brien and Smith (2014) investigate firms' private incentives to offer customers privacy and compare them to the socially optimal incentives. They, however, assume that sellers can commit to privacy policies while we suppose that they cannot (in the absence of privacy regulation). Moreover, there is no scope for learning in their model.

[11] To illustrate signal-jamming incentives for product quality, consider a firm selling an experience good for which a positive experience requires both a high-quality product and a discerning consumer. More precisely, a consumer has a positive experience with probability $q\theta$, where $q \in \{0, 1\}$ is a characteristic of the product and $\theta \in \{0, 1\}$ is a characteristic of the consumer. In response to a positive experience, the consumer forms a posterior belief $r_G = 1$ of being a discerning type; otherwise, the consumer's posterior belief is $r_N = 0$. Thus, even though quality is unobservable, the firm has an incentive for high quality in order to convince a discerning consumer to make a repeat purchase. Our model of equilibrium privacy provision follows a similar logic. The website invests in privacy protection to influence consumers' beliefs about the utility of returning to the website. See Judd and Riordan (1994) and Board and Meyer-ter-Vehn (2013) for more elaborate signal-jamming models of product quality. Early models of signal jamming include Riordan (1985), Fudenberg and Tirole (1986), and Holmström (1999).

Our paper also contributes to the growing literature on markets for information.[12] This literature has mainly focused on the design of information by a monopolist data seller (e.g. Admati and Pfeiderer, 1986, 1990; Bergemann et al., 2018), the optimal policy of a monopolist selling data to competing firms (e.g. Bounie et al., 2018; Belleflamme et al., 2019; Montes et al., 2019), and information sharing among competing firms (e.g. Vives, 1988; Raith, 1996). In contrast, we focus on the interaction between a website's decision to sell data and consumer retention, and provide a model of competition in the market for personal information.

Finally, the security economics literature studies positive externalities from security investments in networks with interdependent risks (e.g. Acemoglu, Malekian and Ozdaklar, 2016). Toh (2018) analyzes a website's *ex ante* incentive to invest in security when consumers gradually learn about a website's security level, and demonstrates an underinvestment problem. In contrast, our analysis examines the implications for consumer security of a website's *ex post* incentive to protect consumer privacy when consumers learn about their vulnerability to attacks, and demonstrates how a website might provide excessive privacy protection.

The rest of the paper is organized as follows. Section 2 lays out the baseline model. Section 3 presents the equilibrium analysis in the baseline (single-homing) scenario and extends it to the multihoming case. Section 4 provides a welfare analysis. Section 5 analyzes the effects of transparency and opt-out policies. Section 6 addresses the extensions. Section 7 concludes. All proofs are relegated to the Appendix.

2 Baseline Model

Consider a website offering a service to a unit-mass population of consumers for two periods: period 1 and period 2. A consumer derives a utility u from the service, but this utility is unknown at the beginning of period 1. Instead, this utility is perceived to be distributed independently in the population according to a cumulative distribution function $F(u)$ with mean u_1 and support $\mathbb{R}$. Furthermore, the mean utility is large enough that all consumers choose to participate in period 1. Each consumer learns her u upon consuming the service, and this knowledge informs her participation decision in period 2.

In this section we consider a scenario in which there is a single website, which we refer to as single-homing, but we later extend our setup to the case where consumers visit multiple websites, which we refer to as multi-homing. The single-homing scenario describes a

[12] See Bergemann and Bonatti (2019) for an excellent survey.

situation in which a website gains *unique* information about consumers and sells a matching service to interested third parties at the monopoly price. The multi-homing scenario describes a situation where multiple websites get the *same* information about consumers and compete to sell a homogeneous matching service. In order to focus on competition in the market for information, we will abstract from competition for consumers.

The website offers the service to consumers for free, and earns revenue from advertising. To fix ideas, suppose that the website has a single advertising slot per customer, that can be used in one of two ways. If it is used for a non-targeted ad, it generates a revenue $a > 0$ per customer. If instead the ad is targeted using information collected when the customer visits the website, it generates a revenue $a + v_t$ (per customer) in period $t = 1, 2$. This information, for example, might come from tracking consumers' online activity with a cookie, or it might be personal information that consumers disclose to the website (such as an email or mailing address). Information may be transferred to the advertiser (directly or through a data broker), or processed internally by the website as part of an ad display service. Recognizing that there are several interpretations of our model, we use the short-hand "selling information" or "information sale" to mean a transaction with a third party, using consumer information collected by the website, that involves a payment to the website.[13]

Our model is based on the premise that targeting allows specialized advertisers to reach consumers that value most their products but also raises the risk of unwanted or malicious targeted ads. There is evidence that targeted advertising may induce negative reactions by consumers (White et al., 2008; Goldfarb and Tucker, 2011), and the rise of malvertising can be partly explained by the increase in the collection of personal data and the resulting targeting opportunities (see e.g. Dwyer and Kanguri, 2017).

We capture this dual nature of targeting by assuming that it can result in three possible consumer experiences, which impact consumer utility. With probability λ, the targeted ad results in a good experience (G) that generates a positive utility $U_G > 0$. With probability θ, the ad is a nuisance or worse, causing a disutility from a bad experience (B), with negative utility $U_B < 0$. We allow that disutility to be small, as in the case of mildly irritating spam, or large, as in the case of ransomware. In all other events, the ad is inconsequential, and the consumer has a neutral experience (N) with no utility consequences, $U_N = 0$, which we assume to be also the case when the ad is not targeted.[14] Hence, a neutral experience occurs with probability 1 in the absence of an information sale, and with probability $1 - \lambda - \theta$ if

[13] Bergemann and Bonatti (2019) distinguish direct and indirect sales of information depending on whether the data seller provides information or access to a consumer.

[14] We simplify the exposition by assuming that untargeted banner ads generate a neutral experience for consumers. However, we could allow for all types of experiences as long as targeting increases both the probability of a good experience and the probability of bad one.

the website sells customer information.

For simplicity, we assume the following specific targeted advertising scenario. A unit-mass of specialized third-party advertisers arrive each period, and each consumer in a period is of interest to exactly one of these advertisers. In order to target its consumer of interest with an ad, the advertiser is willing to pay a fee $a + v_t$ to secure the cooperation of the website.[15]

Critically, consumers are unsure of their expected utility from third-party advertising. We model this by assuming that, while λ is a known parameter, θ is an unknown characteristic of the consumer-website pair.[16] Consumers may be highly vulnerable to a bad experience, i.e. $\theta = \theta_h$, or weakly vulnerable, i.e. $\theta = \theta_l < \theta_h$. We denote by r_1 the *ex ante* probability of weak vulnerability, assumed to be the same for all consumers, and $\theta_1 = r_1\theta_l + (1 - r_1)\,\theta_h$ the *ex ante* vulnerability of the consumer.[17]

Vulnerability can be interpreted in several ways. One interpretation is mistargeted advertising: θ is the probability of a nuisance ad that generates a small disutility for the consumer, and $1 - \lambda - \theta$ is the probability of innocuous poorly targeted ads. Another interpretation is malvertising: θ is the probability of abusive or criminal use of customer information that imposes a large utility loss, e.g. a denial-of-service attack or identity theft. Finally, some consumers might have a poor understanding of how well they are protected against aggressive intrusions, for instance because of a superior antivirus or firewall, so that intrusion by malicious third parties is more likely to fail.

A website visit thus is an experience good. During period 1, a consumer observes u, and also experiences a good (U_G), bad (U_B), or a neutral (0) utility increment from third party ads. The consumer learns about her θ from the realized experience. At the end of period 1, a consumer knows her value of the website service, and revises her beliefs about vulnerability. We will denote by r the updated probability that the type is l. The same situation repeats in period 2 for returning consumers, except that the value of targeting for third parties is v_2.

In our baseline model, the privacy policy of the website determines the probability $X \in$

[15] This scenario can be interpreted as a special case of the one considered in Bergemann and Bonatti (2015) where a "database" maps consumer-advertisers pairs into a "match value" for the advertiser, which is equal to v_t on the diagonal of the corresponding matrix and 0 elsewhere. A "query" by a third-party is a request to identify the unique consumer with match value v_t.

[16] We make this assumption to emphasize the role of learning from a bad experience. However, we discuss in an extension the implications of allowing λ to be unknown as well.

[17] While it is simpler to think about θ as a consumer's charasteristic, our model encompasses also the case where θ is related to the website's characteristics. In this latter case we assume away word of mouth and market learning.

$[0, 1]$ that customer information is not sold in period 1[18] – we consider later an extension with a more active scrutiny by a website who can invest in verification to reduce the likelihood of a bad match. For instance, the website could sell access to a database the content of which depends on its privacy policy. Each advertiser might find the database useful, or not, for the purpose of targeting customers. Thus, the design of the database determines the probability that an advertiser seeking access to a given consumer finds him and buys the targeting service.[19] We will refer to X as (the level of) precaution, and say we have *full precaution* when $X = 1$, *no precaution* if $X = 0$, and *partial precaution* if $0 < X < 1$.[20]

Website privacy policy is unobserved by consumers. In equilibrium, consumers update beliefs about their vulnerability using Bayes Rule and taking as given the website's privacy policy, and the website chooses a profit-maximizing policy given consumer beliefs. Equilibrium and its welfare and policy implications are analyzed next. Later, we consider a richer privacy policy in which the website also can incur a cost to verify third party use of customer information prior to its sale.

3 Equilibrium precaution

3.1 Single-homing

Website privacy policy and consumer beliefs are determined jointly in equilibrium. A website has an incentive for privacy protection only if the long-run value of retaining a customer (V_2) is sufficiently high relative to the short-rung gain from exposing the customer to third parties (v_1). The probability of retention ($Q(r)$) depends on consumers' beliefs about the utility of returning to the website, which depends on consumers' expectations of privacy. Furthermore, the distribution of beliefs in the consumer population, and therefore average retention, depends on the website's privacy policy. In equilibrium, consumers correctly anticipate privacy, and the website correctly anticipates how privacy affects retention.

[18] More specifically, a privacy policy that restricts what consumer information a website collects and how it is used might reduce the ability of the website to match consumers and interested third parties.

[19] Alternatively, we can intepret $X = 0$ (resp. $X = 1$) as meaning that the website always (resp. never) sells customer information to a third party and interpret $X \in (0, 1)$ as a mixed strategy. Note that the latter can be "purified" by introducing into the model a vanishingly small amount of incomplete information about the value of personal information to third parties. See, for instance, Bagwell and Wolinsky (2002).

[20] Note that, in deriving equilibrium, we can fully characterize the second period with the retention value V_2 and the retention rate $Q(r)$, suggesting our model admits alternative interpretations. For example, setting $\lambda = 0$ and $v_2 = 0$, we could interpret $v_1 X$ as an investment to protect consumer data, $p_B(X)$ as the probability of a security breach, $-U_B$ as possible consumer harm, r as the consumer's beliefs about her vulnerability, and V_2 as the average profit from serving consumers. See Toh (2018) for a more elaborate model of security investment with a similarly structured equilibrium.

Our two-period model gives specific content to this notion of equilibrium. It is immediate that the website sells information to all interested third parties at price v_2 in period 2, as there is no further interaction with the customer. Hence the value of retention is $V_2 = \delta^F (a + v_2)$ where δ^F is the firm's discount factor. The retention probability for a given posterior belief r is

$$Q(r) \equiv 1 - F(-M(r)),$$

where

$$M(r) \equiv \lambda U_G + (r\theta_l + (1-r)\theta_h) U_B$$

is the expected benefit from third party matching when the probability of low vulnerability is r. Posterior beliefs, denoted (r_G, r_B, r_N), are the updated probabilities that $\theta = \theta_l$ at the beginning of period 2 after the events G, B, and N are observed by the consumer. For the most part, we focus on scenarios with $M(r_1) \geq 0$, meaning consumers in period 1 have no objection to third party matching. Heterogeneity of posterior beliefs, however, allows that consumers may disagree on the desirability of third party matching in period 2.

Precaution determines the distribution of consumer beliefs at the beginning of period 2. The probability of a good experience (event G) and the probability of a bad experience (event B) are respectively $p_G(X) = \lambda(1-X)$ and $p_B(X) = \theta_1(1-X)$. Both are decreasing in precaution; it follows that the probability of a neutral experience (event N), $p_N(X) = 1 - p_G(X) - p_B(X)$, is increasing in precaution.

Consumer beliefs at the beginning of period 2 are formed using Bayes Rule, conditioning on realized experience and taking precaution X as given. Posterior beliefs after events G and B are respectively $r_G = r_1$ and $r_B = (\theta_l/\theta_1) r_1 < r_1$. Notice that r_G and r_B are independent of X, and therefore can be treated as parameters.[21] This is because these events occur only if information is sold, and the website's strategy does not affect the conditional probabilities of these events. Event N, however, can occur whether or not information is sold. Consequently, a consumer's posterior belief that $\theta = \theta_l$ after a neutral experience depends on anticipated precaution:

$$r_N = \phi(X) \equiv \frac{1 - (\lambda + \theta_l)(1-X)}{1 - (\lambda + \theta_1)(1-X)} r_1, \tag{1}$$

Of course, posterior and prior beliefs must be consistent:

$$p_G(X) r_G + p_N(X) r_N + p_B(X) r_B = r_1.$$

[21] r_G and r_B are not defined by Bayes rule under full privacy (i.e. $X = 1$). We assume that their values remain equal to r_1 and $(\theta_l/\theta_1) r_1$, respectively, in this case.

A neutral experience is good news in our model in the sense that, for all levels of precaution less than full, it yields the most optimistic posterior belief about vulnerability: $r_B < r_1 < r_N$. The intuition behind this result is as follows. Provided $X < 1$, the consumer reasons that a neutral experience could have resulted from the possibility that a third party did gain access to the consumer but the consumer had a neutral experience due to low vulnerability. Thus the consumer becomes more optimistic after a neutral experience. Moreover, $\phi(X)$ decreases in precaution because, by reducing exposure to third parties, higher precaution reduces the likelihood that a neutral experience results from low consumer vulnerability rather than from information not being sold. When there is full precaution, however, the consumer is never exposed to third parties and a neutral experience conveys no information: $\phi(1) = r_1$.

We are now in a position to characterize equilibrium and provide some comparative static results. Given the posterior beliefs, the website has an incentive for full (no) precaution if and only if

$$L(r_N) \equiv v_1 + [\lambda Q(r_G) + \theta_1 Q(r_B) + (1 - \lambda - \theta_1) Q(r_N)] V_2 \leq (\geq)\ Q(r_N) V_2.$$

The intuition behind the equilibrium is then as follows. Selling customer information in period 1 yields extra revenue v_1, but raises both the probability of a good experience by λ and the probability of a bad experience by θ_1, which reduces the retention probability by $\lambda [Q(r_N) - Q(r_G)] + \theta_1 [Q(r_N) - Q(r_B)]$, sacrificing future revenue proportionally. In terms of second-period payoff, the total gain from being cautious and not selling information is then

$$P(r_N) = V_2 \{\theta_1 [Q(r_N) - Q(r_B)] + \lambda [Q(r_N) - Q(r_G)]\}.$$

The optimal strategy for the website is thus defined by the following "best response" correspondence:

$$X^{br}(r_N) \in \arg \max_{X \in [0,1]} \{a + L(r_N) + XP(r_N) - Xv_1\}, \tag{2}$$

where the term between brackets is the website's expected profit. The best response correspondence optimizes the trade-off between avoiding bad experience and selling information. Notice that the gain from precaution decreases with $r_N \in [r_1, \phi(0)]$ and lies in the interval $\left[\psi^f V_2 - v_1, \psi^n V_2 - v_1\right]$ where

$$\psi^f \equiv (\theta_1 + \lambda) Q(r_1) - \theta_1 Q(r_B) - \lambda Q(r_G) < \psi^n \equiv (\theta_1 + \lambda) Q(\phi(0)) - \theta_1 Q(r_B) - \lambda Q(r_G). \tag{3}$$

If $\psi^f V_2 - v_1 < 0 < \psi^n V_2 - v_1$, the level of precaution $X^{br}(r_N)$ jumps from full precaution to

no precaution when the posterior belief induced by neutral experience crosses a threshold $r^M \in (r_1, \phi(0))$. This threshold is defined (uniquely) as the solution of

$$P\left(r^M\right) = v_1. \tag{4}$$

At $r_N = r^M$, the website is indifferent between all levels of precaution. An equilibrium is a level of precaution X^* and a consumer belief r_N^* such that $r_N^* = \phi(X^*)$ and $X^* = X^{br}(r_N^*)$.

Proposition 1 *A unique equilibrium exists under single-homing. Equilibrium precaution is non-increasing in the relative value of information v_1/V_2 and the website chooses full precaution if $v_1/V_2 \leq \psi^f$, partial precaution ($r_N^* = r^M$) if $v_1/V_2 \in \left(\psi^f, \psi^n\right)$ and no precaution if $v_1/V_2 \geq \psi^n$.*

This result implies that v_1/V_2 is a negative indicator for equilibrium precaution. Rewriting this indicator as $\left[\delta^F\left(a/v_1 + v_2/v_1\right)\right]^{-1}$, we see that equilibrium precaution increases with the relative share of income not raising privacy concerns and with the growth rate of the value of information over time. Therefore, everything held equal, we expect more precaution by an e-seller relying extensively on merchandising than by a social network relying extensively on monetization of personal information. Similarly, the website exerts more precaution if a long history record is more valued than a short history record. Furthermore, equilibrium precaution is non-decreasing in the sensitivity of retention to beliefs about vulnerability – measured by the slope of $Q(r)$ for $r \geq r_B$, and in the sensitivity of beliefs to experience – measured by the (absolute value of the) slope of $\phi(X)$ for $X \in [0,1]$.[22]

Example 1 *To illustrate the above comparative statics and provide comparative statics with respect to other parameters of the model, assume that $\theta_l = 0$ (hence $r_B = 0$), u is distributed uniformly over an interval $[\underline{u}, \bar{u}]$ with density α on its support and the retention rate is always interior (i.e. $\underline{u} < -M(1) < -M(0) < \bar{u}$). In this case, we obtain*

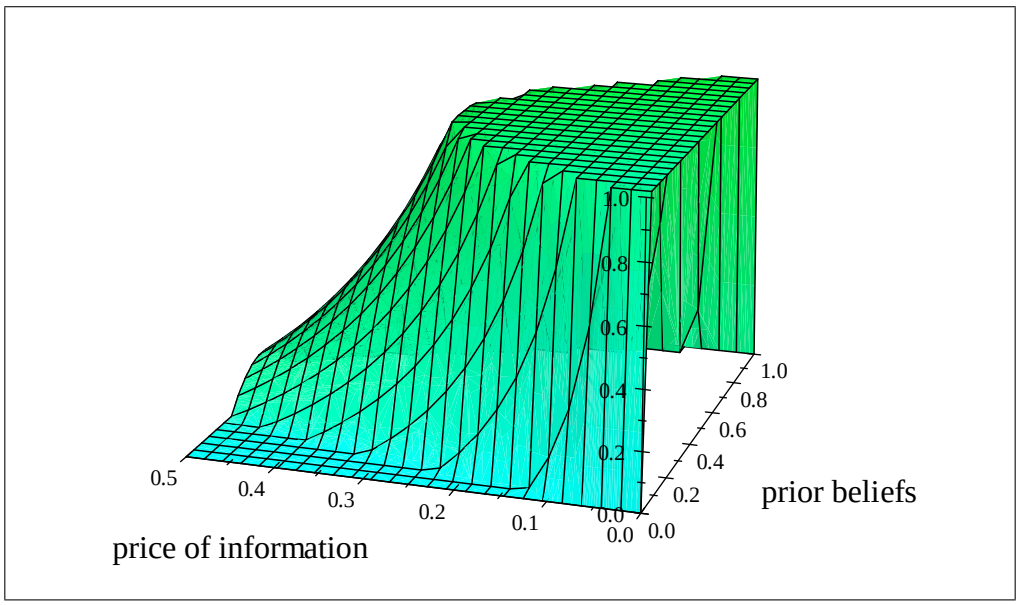
$$\psi^f = \alpha\theta_h^2(1-r_1)r_1\left|U_B\right|; \qquad \psi^n = \frac{\alpha\theta_h^2(1-r_1)r_1\left|U_B\right|}{1-\left[\lambda+(1-r_1)\theta_h\right]}$$

and

$$X^* = 1 - \frac{1-\alpha\theta_h^2(1-r_1)r_1\left|U_B\right|\frac{V_2}{v_1}}{\lambda+(1-r_1)\theta_h} \quad \textit{for} \quad \frac{v_1}{V_2} \in \left[\psi^f, \psi^n\right].$$

[22] Notice that increasing $Q'(r)$ for all $r \geq r_B$ raises the height of $\Delta_B(r) - \Delta_G(r)$ for all $r \geq r_1$, while increasing $|\phi'(X)|$ for all X increases the height of $\phi(X)$, except at $X = 1$ since $\phi(1) = r_1$.

The level of precaution does not depend on U_G as it does not affect the sensitivity of retention to consumer beliefs. It is increasing in $|U_B|$, λ, θ_h, and r_1: precaution is higher when matches are less likely to generate a neutral experience and bad matches are more detrimental to consumers. The effect of the prior belief r_1 is non-monotonic. For extreme levels of prior beliefs, the threshold ψ^n becomes very small as the posterior beliefs do not react to information and $\phi(0)$ is close to r_1. Hence, there is no precaution if consumers are very confident they are not vulnerable, or if they strongly believe that they are vulnerable. Precaution emerges only for intermediate prior beliefs and in this case is decrease in r_1 for a large prior r_1 but increasing for a small prior and a small price of information v_1 (i.e. close to $\psi^f V_2$).[23] *The next graph plots the equilibrium level of precaution X as a function of v_1/V_2 and r_1 for $\alpha = 1$, $|U_B| = 10$ and $\lambda = \theta_h = 1/3$:*



3.2 Multi-homing

Let us now consider $K \geq 2$ websites facing a unit-mass population of consumers in periods 1 and 2. Each website is as in the previous single-homing setting, except that consumers derive a utility u^i from visiting website i that follows a joint distribution $F_K\left(u^1, ...u^K\right)$, which we assume to be symmetric with marginal cumulative distribution $F(u^i)$. It is well understood that competition for consumers affects firms' incentives to provide access to personal data (see e.g. Casadesus-Masanell and Hervas-Drane, 2015). Here we wish to highlight the way

[23] Formally $\frac{\partial X^*}{\partial r_1} = \frac{\theta_h}{(\lambda+(1-r_1)\theta_h)^2}\left(\alpha\theta_h^2(1-r_1)r_1\left|U_B\right|\frac{V_2}{v_1}-1\right)+\frac{1-2r_1}{\lambda+(1-r_1)\theta_h}\alpha\theta_h^2\left|U_B\right|\frac{V_2}{v_1}$.

competition in the market for data shapes those incentives. For this reason we assume that websites are not competitors on the consumer side. Moreover, we assume that websites are *ex ante* symmetric so that consumers multi-home and are active on all websites in period 1.[24] This implies that all websites have access to all customer information and can potentially sell it to each third party.

In a setting with multiple websites, two issues arise: attribution and inference. The first relates to the fact that a consumer having a bad experience may not know which website is responsible for the sale of her personal information. Inference refers to the way a consumer revises her beliefs on each website after a given experience, which depends in particular on the correlation of the vulnerability across websites. Both imperfect attribution and correlation induce a "collective reputation" problem as the sale of personal information by one website affects all websites. To highlight this aspect we assume that consumers do not observe if information is sold or not, and by whom, and that the parameter θ is the same for all websites. We show below that even under these extreme assumptions the market may provide some privacy protection.[25]

Given our maintained assumption that there is a one-to-one match between consumers and third parties, each consumer faces the same potential outcomes as before: she can either have *one* good experience (G), *one* bad experience (B), or a neutral experience (N). The consumer then revises her beliefs about her vulnerability to bad experiences, observes the realization of $\left(u^1, ...u^K\right)$, and decides whether to return to the websites.

A new feature here is that the probability of a non-neutral experience accounts for the fact that many websites can sell information. Thus, if x is a symmetric equilibrium probability that a website does not sell customer information, the total probability that a third party interested in buying such information does not acquire it (aggregate precaution) is $X = x^K$. Notice that we assume independent probabilities that a third-party transacts with a website.[26] With this adjustment in the determination of X, the behavior of consumers is unchanged with respect to the single-homing scenario and, in particular, the equilibrium posterior beliefs for events G, B and N are respectively r_G, r_B and $r_N = \phi(X)$.

We model competition by assuming that all websites decide independently and simultaneously on x and the price p for personal information. We assume for simplicity that all

[24] The number K of websites could vary across consumers, provided it is observable and privacy policy is personalized.

[25] Our conclusion would extend to the case where vulnerability is website-specific, with θ^i denoting vulnerability on website i, if either the consumer cannot attribute perfectly a bad or good experience to a website or the parameters θ^i are positively correlated.

[26] This is the case if x is interpreted as a mixed strategy, or if websites collect different relevant pieces of information and third-parties need only one piece of relevant information.

websites observe when information is sold but do not observe consumers' experiences. On the market for information, multi-homing affects the selling prices in both periods. Let us first consider period 2. A consumer with posterior belief r returns to a given website i with probability $Q_K(r)$. Assuming zero probability of a tie $u^i = u^j$, this probability is:[27]

$$Q_K(r) \equiv \Pr\left\{u^i \geq 0; \sum_{j=1,K} \max\left(u^j, 0\right) \geq -M(r)\right\} + \Pr\left\{0 > u^i \geq -M(r); u^i > \max_{j \neq i} u^j\right\}.$$

Note that the retention rate under multi-homing is smaller than the retention rate under single-homing, i.e. $Q_K(r) < Q(r)$, if and only if the expected utility from third-party matching $M(r)$ is positive.

The website obtains profit a on each retained consumer and competes with other websites for the sale of personal information. If information is not sold in period 1, then the equilibrium expected revenue of a website in period 2 from selling information on returning customers is $v_2 \hat{Q}_K(r_N)$, where (again assuming zero probability of a tie):[28]

$$\hat{Q}_K(r) \equiv \Pr\left\{u^i \geq -M(r); \min\left(u^i, 0\right) > \max_{1 \leq j \leq K;\, j \neq i} u^j\right\}$$

is the probability of unique customer retention, i.e. the probability that a customer returns only to website i. Obviously, $\hat{Q}_K(r) \leq Q(r)$. The expression of the equilibrium profit can be obtained as follows. Each website where the consumer is still active knows that the consumer had a neutral experience but does not know on which other websites she returned.[29] The equilibrium symmetric strategy of the pricing game (on the market for information) in period 2 is then a mixed strategy on an interval (a standard undercutting argument shows that there are neither mass points nor holes). The website earns the same expected profit for all prices in the interval. Moreover, as the upper bound of that interval wins only if there is no other bid (which happens with probability $\hat{Q}_K(r_N)$), it has to be v_2. This yields the expected payoff $v_2 \hat{Q}_K(r_N)$. Thus, if information is not sold by any website in period 1, the period 2 discounted expected revenue from returning consumers is

$$\delta^F \left[a Q_K(r_N) + v_2 \hat{Q}_K(r_N)\right].$$

[27] In the case of perfect correlation, i.e. $u^i = u^j$ for all i and j, we have $Q_K(r) = \Pr\left\{u^i \geq \max\{0; -M(r)/K\}\right\} + \frac{1}{K}\Pr\left\{0 > u^i \geq -M(r)\right\}$.

[28] In the case of perfect correlation, we have $\hat{Q}_K(r) = \frac{1}{K}\Pr\left\{0 > u^i \geq -M(r)\right\}$, which is equal to zero when $M(r)$ is negative.

[29] Alternatively, we could assume that websites track consumers and know where they return. In this case, the equilibrium price of information in the second period is v_2 if there is a monopoly and 0 if the consumer returns to two or more websites. A website's expected payoff would then be the same.

If information is sold in period 1, then the equilibrium period 2 expected revenue is similar but accounts for uncertainty on the posterior belief r. Viewed from period 1, it is given by

$$L_K(r_N) \equiv \mathbb{E}\left\{\delta^F\left[aQ_K(r) + v_2\hat{Q}_K(r)\right] \mid \text{ information is sold in period 1}\right\}.$$

Let us now turn to competition on the market for information in period 1. At this stage of the game, the strategy of a website can be summarized by the choice of x and a probability distribution over prices $p \leq v_1$ with cumulative distribution function $H(p)$.[30] We now characterize a symmetric equilibrium of the game. Let us first derive the website's optimal strategy for a given consumer belief r_N, assuming that all other websites follow the strategy $(x, H(.))$. For this purpose, define

$$\begin{aligned} P_K(r_N) \equiv\ & \delta^F a\left\{\lambda\left[Q_K(r_N) - Q_K(r_G)\right] + \theta_1\left[Q_K(r_N) - Q_K(r_B)\right]\right\} \\ & + \delta^F v_2\left\{\lambda\left[\hat{Q}_K(r_N) - \hat{Q}_K(r_G)\right] + \theta_1\left[\hat{Q}_K(r_N) - \hat{Q}_K(r_B)\right]\right\}, \end{aligned}$$

which is the expected increase in payoff from retaining a consumer when no information about the consumer is sold to third parties. Given that $Q_K(r_N)$ and $\hat{Q}_K(r_N)$ are both increasing in r_N, $P_K(r_N)$ is positive and non-decreasing in r_N. Then, the expected payoff of a website choosing x_i and $p_i \leq v_1$ is

$$a + L_K(r_N) + x_i x^{K-1} P_K(r_N) + (1 - x_i)\left(x + (1-x)(1 - H(p_i))\right)^{K-1} p_i. \tag{5}$$

This expression can be interpreted as follows. Not selling customer information does not imply that access to that information by a third party will not occur because another website may sell the information. In that scenario, the payoff is $L_K(r_N)$. However, with probability x^{K-1}, all other websites also do not sell the information and the payoff is then higher by $P_K(r_N)$. Finally, information is sold by the website if the third party is willing to buy and either no other website sells relevant information or the website's price is the lowest price. Therefore, the comparison of the payoff from selling information at a price p below v_1 and the payoff from not selling information boils down to comparing the expected revenue $\left(x + (1-x)(1 - H(p))\right)^{K-1} p$ with the gain from privacy $P_K(r_N)\, x^{K-1}$.

Equilibrium is not necessarily unique in a multi-homing environment. The following result shows that one of the possible outcomes is that all websites sell personal information to any interested third party and competition dissipates fully their profits from data sales.

[30] We could allow the price to be above v_1 without altering the analysis. In this case x would be replaced by $x' = x + 1 - H(p)$. Given that putting a mass above v_1 is equivalent to increasing x, we ignore this possibility. See, however, our discussion of transparency.

Proposition 2 *In the multi-homing scenario ($K \geq 2$), there always exists a no-precaution equilibrium where all websites quote a zero price for information, and information is always sold.*

Thus, as soon as there are at least two websites that can sell the same personal information, there is a risk of a total collapse in the provision of privacy.[31] The next proposition shows, however, that there also exists a symmetric equilibrium with a positive level of precaution, denoted X_K^*, if the value of information v_1 is not too high.

Proposition 3 *Consider the multi-homing scenario ($K \geq 2$). There exists a symmetric equilibrium with positive precaution (i.e., $X_K^* > 0$) if and only if $v_1 < P_K(\phi(0))$. It is uniquely defined as follows:*

- If $v_1 \leq P_K(\phi(1))$, then the websites provide full precaution (i.e., $X_K^* = 1$).

- If $P_K(\phi(1)) < v_1 < P_K(\phi(0))$, then the websites' level of precaution and price distribution are given by

$$\begin{aligned} v_1 &= P_K\left(\phi\left(X_K^*\right)\right); \\ H(p) &= \frac{1 - \left(X_K^*\right)^{\frac{1}{K}} \left(\frac{v_1}{p}\right)^{\frac{1}{K-1}}}{1 - \left(X_K^*\right)^{\frac{1}{K}}} \text{ for } p \in \left[\left(X_K^*\right)^{\frac{K-1}{K}} v_1, v_1\right]. \end{aligned}$$

Notice that the level of precaution would not be affected by the number of websites if it was commonly known that the consumer would return to one website only.[32] Indeed, given that the marginal distribution of any u^i is assumed to be F regardless of the number of websites, we would have $Q_K(r) = \hat{Q}_K(r) = Q(r) = 1 - F(-M(r))$ and, therefore, $X_K^* = X^*$. Hence, the effect of multi-homing on the equilibrium level of first-period precaution is related to the negative effect of second-period multi-homing on the website's returns to privacy.

From Proposition 3 it follows that $X_K^* < X^*$ if and only if $P_K(r_N^*) < P(r_N^*)$, i.e. if and only if a website's expected net gain from being cautious is smaller under multi-homing than under single-homing. The following proposition provides sufficient conditions under which this condition holds.

Proposition 4 *Consider the multi-homing scenario ($K \geq 2$).*

[31] However, we conjecture that if there is an arbitrarily small mass ε of single-homing consumers, the zero-price equilibrium exhibited in Proposition 2 exists only if $v_1 \geq \Delta_B(r_N) - \Delta_G(r_N)$.

[32] This is the case for instance if only one website generates utility $u^i > \underline{u}$ while all others generate a large negative utility $\underline{u}$.

(i) If consumer preferences are perfectly correlated across websites then equilibrium aggregate precaution is lower than in the single-homing scenario ($X_K^* \leq X^*$).

(ii) Alternatively, if $\Pr(u^j > 0 | u^i) > 0$ *for* $j \neq i$*, then equilibrium aggregate precaution is lower than in the single-homing scenario ($X_K^* \leq X^*$) if a/v_2 is small enough or $M(r_B) > 0$.*

Multi-homing tends to dissipate future profits due competition in the market for information. This suggests that websites' marginal gain from inducing more optimistic beliefs about vulnerability through higher precaution is lower under multi-homing than under single-homing. Proposition 4 identifies sufficient conditions under which this intuition holds. However, multi-homing has another effect stemming from participation externalities exerted by each website on the others. Under the assumption that websites do not compete in the market for consumers, participation by consumers at a website boosts participation at other websites if the expected matching value is negative. In this case, a website inducing more optimistic beliefs gets not only a direct benefit but also an indirect benefit resulting from higher consumer participation at other websites. Therefore, it seems possible in general that websites' gain from inducing more optimistic beliefs through higher precaution increases with multi-homing. However, this is not to the case for the following extension of Example 1 to multi-homing.

Example 2 *We now extend the setting considered in Example 1 to a scenario with $K = 2$. We assume that u^1 and u^2 are independently drawn from the same uniform distribution over $[\underline{u}, \bar{u}]$ with density α over its support and such that $\underline{u} < 0 < \bar{u}$. We suppose again that $\theta_l = 0$, hence that $r_B = 0$, and that all retention rates are interior (i.e., $\underline{u} < -M(1)) < -M(0) < \bar{u}$). We show in the Appendix that the unique retention rate is given by*

$$\hat{Q}_2(r) = \begin{cases} \alpha^2 \left(-\frac{1}{2} M(r)^2 - \underline{u} M(r) - \underline{u}\bar{u}\right) & \text{if} \quad M(r) \geq 0 \\ \alpha^2 \left(-\underline{u} M(r) - \underline{u}\bar{u}\right) & \text{if} \quad M(r) < 0 \end{cases}$$

and the total retention rate is

$$Q_2(r) = \alpha^2 \left(-\frac{1}{2} M(r)^2 - \underline{u} M(r)\right) + \alpha\bar{u}.$$

Assume now that $M(r_1) \geq 0 > M(0)$. We compute in the Appendix $P_2(r_N)$ using the expressions of $\hat{Q}_2(r)$ and $Q_2(r)$ provided above and establish that $P_2(r_N) < P(r_N)$. This implies that equilibrium aggregate precaution in the multihoming secnario with two websites is lower than equilibrium precaution in the single-homing scenario.

4 Welfare

Before turning to public policies intended to enhance privacy in the next section, we define our measures of welfare, and then discuss the divergence between social and equilibrium incentives for precaution.

Denote by $U(r_N, X)$ the expected utility of a consumer for given beliefs and a given level of aggregate precaution. It is useful to decompose U into a short-term component plus a long-term component:

$$U(r_N, X) \equiv U_1(X) + \delta^C U_2(r_N, X)$$

where $U_1(X) \equiv Ku_1 + (1-X)M(r_1)$ is the expected utility of consumers in period 1 with a prior belief r_1, $U_2(r_N, X)$ is the expected utility of consumers in period 2 with zero precaution in that period, i.e.

$$U_2(r_N, X) \equiv \mathbb{E}\left\{\max\left(\max_{J\subset\{1,..,K\}} \sum_{k\in J} u_k + M(r), 0\right) \mid X\right\},$$

and δ^C is consumers' discount factor. The marginal effect of precaution on short-term utility is equal to the loss of match utility $M(r_1)$, which in general can be either positive or negative. The following lemma, however, establishes that greater precaution always decreases long-term consumer utility inclusive of its effect on beliefs.

Lemma 1 *$U_2(\phi(X), X)$ is decreasing in X.*

It follows that consumer expected utility is decreasing in precaution if beliefs adjust to changes in precaution and $M(r_1)$ is not too negative.

Multi-homing has mixed effects on consumer welfare. Obviously, access to a greater number of websites directly increases the utility consumers get from website content. Moreover, under conditions of Proposition 4 or Example 2, equilibrium precaution is lower under multi-homing, which increases long-run utility. Short-term utility, however, increases only if the *ex ante* match utility is positive.

For a given belief r_N (and treating r_G and r_B as parameters) and a symmetric equilibrium with aggregate precaution X, a website makes an expected profit

$$\Pi_K(r_N, X) \equiv \underbrace{a + \left(X^{\frac{K-1}{K}} - X\right) v_1}_{\text{profit in period 1}} + \underbrace{L_K(r_N) + XP_K(r_N)}_{\text{discounted profit in period 2}}.$$

This expression follows from (5) by taking $x_i = X^{\frac{1}{K}}$ and $p_i = v_1$.[33] As retention rate Q_K and unique retention rate $\hat{Q}_K$ are both increasing in r_N, holding X constant, an improvement in consumer beliefs increases expected profit, i.e. $\partial \Pi_K (r_N, X) / \partial r_N > 0$. Moreover, in the single-homing scenario ($K = 1$), the website internalizes the effect of precaution on consumers' experience. Thus, for an interior equilibrium with $r_N = \phi (X^*)$, the overall effect of precaution reduces to its negative impact on retention through the change in consumer belief. However, when $K > 1$, there are countervailing positive effects resulting from the fact that a website selling information exerts a negative externality on the other websites, i.e. $\partial \Pi_K (\phi (X_K^*), X_K^*) / \partial X > 0$. Thus, in the single-homing case, the website would benefit from lower precaution, i.e. $d\Pi (\phi (X^*), X^*) / dX < 0$,[34] while in the multi-homing case the effect of lower precaution on websites' profits is ambiguous because of two opposite effects:

$$\frac{d\Pi_K (\phi (X_K^*), X_K^*)}{dX} = \underbrace{\phi' (X_K^*) \frac{\partial \Pi_K (\phi (X_K^*), X)}{\partial r_N}}_{<0} + \underbrace{\frac{\partial \Pi_K (\phi (X_K^*), X_K^*)}{\partial X}}_{>0}.$$

Multi-homing has mixed effects on websites' profits. The direct effect of multi-homing on websites' first-period profits is negative because another website may sell customer information, which is captured by $X^{\frac{K-1}{K}} < 1$. The effect of multi-homing on second-period profits is in general ambiguous. On the one hand, multi-homing makes the unique retention rate $\hat{Q}_K$ smaller, which tends to decrease the second-period profit conditional on information being sold in the first period, i.e. $L_K (r_N)$. It also makes the unique retention rate less sensitive to beliefs, which tends to decrease the return from precaution $P_K (r_N)$. On the other hand, multihoming increases the retention rate of consumers whose expected utility from third-party matching is negative but decreases the retention rate of those whose expected utility from third-party matching is positive, with ambiguous effects on $L_K (r_N)$. Finally, the impact of multi-homing on the sensitivity of total retention is in general ambiguous but under the conditions of Proposition 4, multi-homing leads to a decrease in the return from precaution $P_K (r_N)$. Note, however, that the effect of multi-homing is unambiguously negative if the expected utility from third-party matching is always positive, i.e. $M(r_B) > 0$.

Let us now consider the joint profit of websites and third parties. Define

$$\tilde{Q}_K (r) \equiv \Pr \left\{ \max_{J \subset \{1,..,K\}} \sum_{k \in J} u_k \geq -M (r) \right\},$$

[33] Recall that each website is indifferent between all prices in the support of the equilibrium price distribution.

[34] For notational consistency, we denote Π, instead of Π_1, the website's expected profit in the single-homing scenario.

which is the probability that a consumer returns to at least one website in the second period. Then, assuming that websites and third parties have the same discount factor, their discounted joint second-period profit if information is sold is

$$\tilde{L}_K(r_N) \equiv \mathbb{E}\left\{\delta^F\left[KaQ_K(r) + v_2\tilde{Q}(r)\right] \mid \text{ information is sold in period 1}\right\}.$$

Not selling information increases this profit by

$$\begin{aligned}\tilde{P}_K(r_N) \equiv\ & \delta^F Ka\left\{\lambda\left[Q_K(r_N) - Q_K(r_G)\right] + \theta_1\left[Q_K(r_N) - Q_K(r_B)\right]\right\} \\ & +\delta^F v_2\left\{\lambda\left[\tilde{Q}(r_N) - \tilde{Q}(r_G)\right] + \theta_1\left[\tilde{Q}(r_N) - \tilde{Q}(r_B)\right]\right\}.\end{aligned}$$

Hence, the joint profit of websites and third-parties is

$$\tilde{\Pi}_K(r_N, X) \equiv Ka + \tilde{L}_K(r_N) + X\tilde{P}_K(r_N) + (1-X)v_1$$

which is increasing in beliefs r_N and precaution X, because $\tilde{P}_K(r_N)$ is greater than $KP_K(r_N)$,[35] and therefore greater than $v_1 = P_K(r_N)$.

Let us now turn to total welfare. Expected total welfare for given beliefs and precaution is

$$W(r_N, X) \equiv \tilde{\Pi}_K(r_N, X) + U(r_N, X).$$

Clearly, there is a divergence between equilibrium and social incentives for precaution. First, websites ignore the direct effect of precaution on consumer expected utility. Second, websites do not internalize the effect of their actions on other websites and third parties. Third, because precaution is unobservable, websites cannot control consumer beliefs. Finally, it may be desirable to have some privacy protection in period 2 if $M(r) < 0$ for a sufficiently large fraction of consumers in the second period.

We may then distinguish two extreme cases depending on whether match utility is positive or negative. In the "optimistic" case, match utility is positive for all relevant beliefs. Then, with single-homing, equilibrium precaution in period 1 is necessarily excessive because both consumers and websites prefer less precaution and third parties are indifferent. Furthermore, if $M(r_B) > 0$, then zero precaution in period 2 is efficient. However, with multi-homing, websites may prefer more precaution to internalize competitive externalities.

[35] To see this, notice that $\tilde{Q}(r) = K\hat{Q}_K(r) + \Pr\left\{\sum_j \max\left(u^j, 0\right) \geq -M(r); u^i > 0 \text{ and } u^j > 0 \text{ for some } i \neq j\right\}$ where the second term is the probability of at least 2 websites being visited, which is increasing in r. Hence $\tilde{Q}'(r) > K\hat{Q}'_K(r)$ implying $\tilde{P}(r) > KP_K(r)$.

In the "pessimistic" case, expected match utility is negative for the relevant beliefs, and the social desirability of precaution is ambiguous for additional reasons. First, zero precaution in period 2 is inefficient. Second, consumers might prefer more precaution in period 1, depending on whether the short-term utility gain outweighs the long-term utility loss. In between these extremes, the welfare effects of raising precaution above the equilibrium level are ambiguous.

Which case is more salient depends on the nature of customer information and the perceived consumer risks. For example, disclosure of highly sensitive health information may be perceived to cause negative match utility. Alternatively, negative match utility depends on the extent to which information sharing exposes the consumer to malicious attacks. On the one hand, nuisance advertising or innocuous spam might cause only a minor annoyance (small $|U_B|$), that is overshadowed by mutually beneficial targeted advertising (relatively large $|U_G|$). On the other hand, even a small possibility of identity theft due to phishing or malware (very large $|U_B|$) could weigh heavily on match utility. Whether the optimistic case or the pessimistic case is more salient might ultimately depend on policymakers' perceptions of the extent to which online advertising platforms, or other forms of information sharing, increase consumers' risk of becoming victims of serious cybercrime.

5 Policy

5.1 Transparency policy

How should a policy governing privacy protection be conducted? One salient policy intervention is to enforce transparency regarding the website's collection and use of personal data. In our setup, transparency may relate to *ex post* disclosure of information sales or to *ex ante* commitment on information that may be sold. For instance, transparency in the European GDPR is of the latter type**.** A key issue is the extent to which a transparency policy makes credible an announced privacy policy.

***Ex post* transparency** Under *ex post* transparency, each website chooses precaution at the beginning of period t and discloses its value once trade with third parties has been completed. Assuming this is aggregate information, each consumer knows the probability to be matched with a third party by each website at the end of period t but not the realized matches. A direct implication is that in our model, all websites still want to trade with all third parties in the second period because *ex post* revelation of second-period precaution does not affect demand. A difference from the case in which precaution is unobservable

is that the website directly affects posterior beliefs with its choice of precaution. In the single-homing scenario, since profit is increasing in consumer belief r_N, it is immediate that the website will change X in the direction that raises r_N. This is driven by the fact that the website would benefit from consumers interpreting a neutral experience as a stronger signal about their low vulnerability.

Lemma 2 *(i) Consider the single-homing scenario ($K = 1$). An ex post transparency equilibrium precaution level $\tilde{X}^*$ satisfies:*

- $v_1 \leq P(\phi(1)) + \phi'(1)[L'(\phi(1)) + P'(\phi(1)]$ *for* $\tilde{X}^* = 1$,
- $v_1 = P\left(\phi\left(\tilde{X}^*\right)\right) + \phi'\left(\tilde{X}^*\right)\left[L'(\phi\left(\tilde{X}^*\right)) + \tilde{X}^* P'(\phi\left(\tilde{X}^*\right)\right]$ *for* $\tilde{X}^* \in (0,1)$.
- $v_1 \geq P(\phi(0)) + \phi'(0) L'_K(\phi(0))$ *for* $\tilde{X}^* = 0$.

Let us now consider the effect of a transparency policy under multi-homing, focusing on symmetric equilibria with precaution. The analysis of an *ex post* transparency equilibrium is similar to the baseline multi-homing case. A difference with the case of a non-transparent policy is that websites internalize the negative effect of precaution on consumers' posterior beliefs and retention rates.

Lemma 3 *Consider the multi-homing scenario ($K \geq 2$). An ex post transparency symmetric equilibrium with positive aggregate precaution $\tilde{X}_K^*$ satisfies:*

- $v_1 \leq P_K(\phi(1)) + \phi'(1)[L'_K(\phi(1)) + P'_K(\phi(1)]$ *for* $\tilde{X}_K^* = 1$,
- $v_1 = P_K\left(\phi\left(\tilde{X}_K^*\right)\right) + \phi'\left(\tilde{X}_K^*\right)\left[L'_K(\phi\left(\tilde{X}_K^*\right)) + \tilde{X}_K^* P'_K(\phi\left(\tilde{X}_K^*\right)\right]$ *for* $\tilde{X}_K^* \in (0,1)$

and

$$H(p) = \frac{1 - \left(\tilde{X}_K^*\right)^{\frac{1}{K}} \left(\frac{v_1}{p}\right)^{\frac{1}{K-1}}}{1 - (X_K^*)^{\frac{1}{K}}} \text{ for } p \in [\left(\tilde{X}_K^*\right)^{\frac{K-1}{K}} v_1, v_1].$$

We provide in the proof of Lemma 3 conditions for existence and uniqueness of a symmetric equilibrium with positive precaution, although we lack simple intuitive conditions because the private return to privacy under *ex post* transparency is not necessarily decreasing in precaution. We do, however, have a general result about the comparison of equilibrium precaution with and without *ex post* transparency.

Proposition 5 - *In the single-homing scenario ($K = 1$), the equilibrium precaution level in the absence of ex post transparency is (weakly) higher than the precaution level in any equilibrium under ex post transparency.*

- *In the multi-homing scenario ($K \geq 2$), if a symmetric equilibrium with positive aggregate precaution exists in the absence of ex post transparency, then this equilibrium features a (weakly) higher precaution than any symmetric equilibrium under ex post transparency.*

Ex post transparency regulation always increases consumer long-term utility by improving the learning process. Thus, when the expected value from a match for consumers is positive ($M(r_1) > 0$) or consumers are sufficiently patient (δ^C is large), consumers benefit from such a regulation. However, if $M(r_1) < 0$, then consumers are negatively affected by a weaker privacy policy in the first period, creating a tension between short-term and long-term effects on consumer utility.

***Ex ante* transparency** A policy enforcing *ex ante* transparency induces revelation of a website's precaution level in a given period before consumers decide on whether they visit the website in that period. This plausibly enables the website to credibly commit to greater precaution than the equilibrium level, i.e. enforces an announced lower bound on precaution. For example, the website might commit to more precaution by promising (transparently) to collect less data about consumers, which compromises the website's ability to match consumers and interested third parties. In this case, it may be profitable for a website to exert some precaution in the second period, when this boosts demand. The next lemma provides sufficient conditions for this not to occur under single-homing.

Lemma 4 *Consider the single-homing scenario ($K = 1$) and assume that $M(r_1) \geq 0$. Under ex ante transparency, the website exerts no precaution in the second period if either (i) $M(r_B)$ is not too negative, or (ii) θ_1 is small enough, or (iii) $uF'(u)$ is concave.*

We have shown that in the single-homing scenario the website wants to commit to less rather than more precaution in period 1 in order to improve customer retention after a neutral experience. But it is not clear how the website credibly commits to less precaution. For example, suppose the website announces that it will collect additional personal data about its customers, potentially improving matching with interested third parties. The website would still have the ability and the incentive to refuse to deal with interested third parties, as long as the refusal is unobservable to consumers and did not violate its announced privacy policy. In other words, a commitment to reduce precaution is not credible.[36] In this case, a policy enforcing *ex ante* transparency, without also sanctioning refusals to deal, would have no effect on equilibrium precaution, under the condition that no-precaution in the second period is optimal for the website, as stated in Lemma 4. The following proposition summarizes the above discussion.

[36] Formally, if the website were to announce $X < X^*$, the equilibrium would be in mixed strategies with the website refusing to sell with positive probability, such that the probability of a match is X^*.

Proposition 6 *Consider the single-homing scenario ($K = 1$). Under the conditions of Lemma 4, a regulation enforcing ex ante transparency causes the website to choose the same privacy policy as in the equilibrium with no transparency.*

Consider now the multi-homing scenario. As in the single homing case, committing to some precaution in the second period may reduce or increase the average retention rates. Moreover, *ex ante* transparency implies that a firm's precaution level is observed before price competition takes place. As a consequence, raising the level of precaution induces competing sellers of information to raise their prices. Due to this strategic effect, there is no simple extension of Lemma 4 to the multi-homing scenario.[37] As an illustration, no precaution cannot be an equilibrium if $a = 0$ and preferences are perfectly correlated, because in this case websites have no revenue from sales of information if none of them commits to some positive level of precaution. By contrast, when a is large relative to v_1, websites commit to no-precaution in the second period if refraining from selling information depresses expected demand. Nevertheless, we can provide conditions under which no-precaution in the second period is an equilibrium outcome.

Lemma 5 *Consider the multi-homing scenario with $K = 2$ and $M(r_1) \geq 0$. Under ex ante transparency, no-precaution remains an equilibrium outcome in the second period if v_2 is small enough and either $M(r_B)$ is not too negative or θ_1 is small enough.*

Consider the first period. Assuming that information is sold to all interested third parties in the second period, reducing the level of precaution may not be credible if the website can refuse to sell. In particular, starting from the equilibrium precaution level $x = (X_K^*)^{\frac{1}{K}}$, a website who deviates and announces $x_i < (X_K^*)^{\frac{1}{K}}$ would refuse to sell with probability y such that $y + (1-y)\, x_i = (X_K^*)^{\frac{1}{K}}$. The equilibrium distribution of prices and the market level of precaution would then be unaffected. Hence the equilibrium obtained under *ex post* transparency would not be credible under *ex ante* transparency, suggesting that equilibrium precaution is not affected.

However, with multiple websites competing on the market for information, a website will also care about the effect of its policy on the intensity of competition. In particular,

[37] Denoting $\hat{Q}(x_1, x_2)$ and $Q(x_1, x_2)$ the unique and total retention rates if the website's precaution is x_1 and the other website's precaution is x_2, it can be shown that with $K = 2$ websites, no precaution is an equilibrium if for all $x \in [0, 1]$:

$$(1-x)\left((1-x)\,\hat{Q}(0,x) + xQ(0,x)\right)\frac{Q(x,0)}{Q(0,x)} \leq \hat{Q}(x,0).$$

a commitment to a higher level of precaution than under a non-transparent policy, i.e. $x_i > (X_K^*)^{\frac{1}{K}}$, is credible and would induce competitors to price less aggressively. The reason is that firm i's data becomes less attractive, which allows competitors to raise prices without changing the probability of sale. If this effect is strong, which occurs when X_K^* is small enough, the websites would deviate from $(X_K^*)^{\frac{1}{K}}$ toward higher precaution. In this case, we show in Section 1 of the Online Appendix that there does not exist an equilibrium in which websites choose pure precaution strategies. However, we also show that if X_K^* is large enough, which holds if v_1 is small enough, then there exists a symmetric equilibrium in which all websites chooses $(X_K^*)^{\frac{1}{K}}$ in period 1.

Proposition 7 *Consider the multi-homing scenario. If v_1 is small enough and no-precaution remains an equilibrium outcome in the second period, then the symmetric equilibrium with no transparency is also an equilibrium under ex ante transparency.*

5.2 Opt-out

Another salient policy measure is to give consumers control rights over their personal information. Ideally, a consumer would like to choose which third parties can access her personal information, and for what purposes. However, contracts typically are incomplete do to private information and lack of verifiability. Here, we assume that whether information is sold is verifiable, but that the nature of the match (good, bad, or neutral) is not. We modify the previous models by allowing consumers to opt out, meaning they can prevent any sale of personal information i.e. full precaution prevails for those consumers.[38]

We assume u_1 is sufficiently high that consumers do not want to opt out in the first period, but a consumer might want to do so after revising beliefs about her vulnerability. Thus, at the end of the first period, a consumer has three options: stop their relationship with the website, return to the website and opt in (i.e. not prevent the website from selling her personal information), or return to the website and opt out.

A consumer's decision to opt out depends on her beliefs about match value. Let

$$\bar{r} \equiv \frac{\lambda U_G + \theta U_B}{(\theta_h - \theta_l)\, U_B}$$

denote the solution to $M(r) = 0$. Our leading case in this section is the one in which consumers opt out only after a bad experience, i.e. $r_1 \geq \bar{r} > r_B$ or, equivalently, $M(r_1) \geq$

[38] Bloch and Demange (2018) also analyze the effect of opt-out on a website's privacy policy (captured by it level of data exploitation), but assume that the website's policy is observable to consumers.

$0 > M(r_B)$.[39]

Since opting out leads to a match vale of $M(\bar{r}) = 0$ instead of $M(r_B) < 0$, the expected retention of consumers with a bad experience rises from $Q(r_B)$ to $Q(\bar{r})$. At the same time, under single homing, the website's value of a retain consumer falls from $a+v_2$ to a because the website no longer earns second-period revenue from third parties for these consumers. The same is true under multi-homing, except that the forgone expected revenue is less because of competition in the market for information. The consequences for equilibrium precaution and posterior beliefs are stated in the next two propositions, first for single-homing in the next proposition, and then for multi-homing.

Proposition 8 *Assume $M(r_1) \geq 0 > M(r_B)$. In the single-homing scenario, consumers' ability to opt out leads to weakly lower equilibrium precaution (lower X) and weakly more favorable equilibrium posterior beliefs (higher r_N) if $Q(\bar{r})a \geq Q(r_B)(a + v_2)$. Conversely, opt-out leads to higher precaution and less favorable posterior beliefs if $Q(\bar{r})a < Q(r_B)(a + v_2)$.*

Proposition 9 *Assume $M(r_1) \geq 0 > M(r_B)$. In the multi-homing scenario, consumers' ability to opt out leads to weakly lower equilibrium precaution (lower X) and weakly more favorable equilibrium posterior beliefs (higher r_N) if $Q_K(\bar{r})a \geq Q_K(r_B)a + \hat{Q}_K(r_B)v_2$. Conversely, opt-out leads to higher precaution and less favorable posterior beliefs if $Q_K(\bar{r})a < Q_K(r_B)a + \hat{Q}_K(r_B)v_2$.*

We distinguish two alternative policy regimes. In a voluntary regime, under single-homing, a website at its discretion can offer consumers the ability to opt out. Under multi-homing, websites can collectively commit to offering the ability to opt out consumers. Notice that, in the multi-homing case, a consumer opts out of all websites or none. There is in fact a strong externality governing websites' opt-out decisions. Indeed, the option offered by any given website is irrelevant if the consumer visits another website not offering the option. Rather than examine individual incentives of websites to offer the option, we focus on a form of collective self-regulation, i.e. the policy permitting websites collectively to commit to offer opt-out. In contrast, in a mandatory regime, websites have no choice but to allow consumers to opt out.

Voluntary regime Assume that websites voluntarily allow consumers the ability to opt-out if is weakly profitable under single-homing and collectively profitable under multi-homing. Suppose websites can offer the option at any point in the game. Then websites offer the option at the end of the first period if $Q(\bar{r})a \geq Q(r_B)(a + v_2)$ in the single homing case,

[39] Other cases in the single-homing scenario are considered in Section 2 of the Online Appendix.

and if $Q_K(\bar{r})a \geq Q_K(r_B)a + \hat{Q}_K(r_B)\, v_2$ in the multi-homing case. Furthermore, anticipating this, websites might just as well commit to the policy at the beginning of the first period, assuming that u_1 is sufficiently large that consumers do not opt-out initially.

If $Q_K(\bar{r})a < Q_K(r_B)a + \hat{Q}_K(r_B)\, v_2$, however, websites might have a collective interest in committing to opt-out at the beginning of the first period. To see this, suppose $[Q_K(r_B) - Q_K(\bar{r})]a - \hat{Q}_K(r_B)\, v_2 = \varepsilon \geq 0$. Starting at $\varepsilon = 0$, a small increase in ε has three effects on profits if consumers can opt-out. First, there is a direct negative effect of lost third-party revenue in the second period. Second, from Proposition 9, the resulting increase in first-period precaution has a negative effect on second period profits due to less favorable beliefs. Finally, greater precaution increases first-period profits by reducing competition in the market for information. In the single-homing case the effect of a small change in X on first-period profits is second-order, because the single website is choosing precaution optimally in equilibrium. This is not the case under multi-homing because each website enjoys a first-order benefit from greater precaution exercised by its rivals in the market for information.

Consequently, while there is a clear case for a voluntary opt-out policy under single homing, the same cannot be said for permitting a voluntary collective agreement to offer opt-out under multi-homing.

Proposition 10 *Assume $M(r_1) \geq 0 > M(r_B)$. A voluntary opt-out policy under single homing results in (weakly) less precaution and greater short- and long-run consumer utility compared to no opt-out. In contrast, under multi-homing, a collectively voluntary opt-out policy might similarly result in less precaution and greater consumer utility, but, under certain conditions, might also result in more precaution, less short-consumer utility, and an ambiguous effect on long-term utility.*

Mandatory regime A mandatory policy would enable consumers to opt-out even when it is unprofitable for websites. Thus, relative to a voluntary opt-out policy, a mandatory policy affects website conduct if and only if $Q(\bar{r})a < Q(r_B)(a + v_2)$ under single homing, and only if $Q_K(\bar{r})a < Q_K(r_B)a + \hat{Q}_K(r_B)\, v_2$ under multi-homing. Under these circumstances, according to Propositions 8 and 9, opt-out results in greater precaution, and according to Proposition 10, this is detrimental to consumers in the short run if $M(r_1) > 0$, and has mixed effects on consumers in the long-run.

Proposition 11 *Assume $M(r_1) \geq 0 > M(r_B)$. Compared to a voluntary opt-out policy, a mandatory opt-out policy results in more precaution, and lower profit and short-term consumer utility, while the effect on long-term consumer utility is ambiguous.*

The inevitable conclusion is that an unambiguous case for a mandatory opt-out policy is lacking under both single-homing and multi-homing.

6 Extensions

In this section we present three extensions of our model. For the sake of simplicity, we focus on the single-homing scenario ($K = 1$).

6.1 Verification of third party use of information

Assume that the (unique) website can verify third party use of information. For instance, a website may use Artificial Intelligence and human resources to screen out ads with offensive content, or may verify that requests to access customer information originate from known legitimate third parties. To capture this, we suppose that, by incurring a cost $C_v(Y)$, the website can generate a signal on the user experience that is informative with probability Y, referred to as the level of verification. A non-informative signal conveys no information while an informative signal allows to detect perfectly whether the third party will generate a bad experience or not.[40] Thus, an informative signal allows screening third parties generating bad experiences from those leading to good or neutral experiences.[41] For conciseness, we assume that $C_v(Y)$ is convex, $C_v'(0) = 0$ and $C_v'(1) = +\infty$.

The website's strategy now consists of a choice of a level of verification Y as well as the probability X that customer information is not sold in case verification fails. If verification succeeds then the website denies access to information to third parties generating a bad experience and grants other third parties full access to information. Therefore, we can characterize the website's strategy by a pair

$$(X, Y) \in [0, 1]^2.$$

Under a full precaution policy ($X = 1$), the consumer is immune to unwanted intrusions from the sale of personal data, and verification is a way to raise the value to the consumer

[40] We chose this specification of the verification technology because it yields a nice separation property between precaution and verification, thus making the analysis more transparent.

[41] As an illustration, suppose the website can incur a cost z, drawn from a distribution with an increasing continuous cdf $Z(.)$ over the support $\mathbb{R}_+$, to identify (with certainty) whether a match with a third party will generate a bad experience or not. It is straightforward that there must exist a critical level $\hat{z}$ (potentially zero) such that the website verifies the third party's use of information if $z < \hat{z}$. We can then denote by $Y = Z(\hat{z})$ the probability of verification and the cost of verification is $C_v(Y) = \int_0^{\hat{z}} z dZ(z)$.

of visiting the website. The variable Y then determines the benefit from allowing access to third parties that do not provide a bad experience. On the contrary, under a no precaution policy ($X = 0$), verification is the only way to avoid interactions with third parties that generate a bad experience and, therefore, determines the level of protection against them.

Let us now provide the probability of each type of first-period experience and determine how it depends on the level of precaution X and the level of verification Y. The probabilities of a good experience (event G) and a bad experience (event B) are given, respectively, by $p_G(X,Y) = \lambda[Y + (1-Y)(1-X)]$ and $p_B(X,Y) = \theta_1(1-Y)(1-X)$. Both probabilities decrease with X because a higher level of precaution leads to less sales of personal information to third parties. Moreover, the probability of a good experience is increasing in the level of verification unless there is no precaution, because more verification decreases the likelihood that a third party generating a good experience buys customer information. By contrast, the probability of a bad experience decreases with the level of verification unless there is full precaution. The probability of a neutral experience (event N) is then

$$p_N(X,Y) = 1 - Y\lambda - (1-Y)(\lambda + \theta_1)(1-X). \tag{6}$$

This probability increases with X as in the case with no verification. The effect of the level of verification Y on the probability of a neutral experience depends on the level of precaution X, as shown by $\partial p_N/\partial Y = \theta_1 - (\lambda + \theta_1)X$. The reason is that a higher level of verification has two (potential) opposite effects on the likelihood of a neutral experience. First, it affects it positively by increasing the probability that a third party generating a bad experience is denied access to customer information. Second, it affects it negatively by making it more likely that a third party generating a good experience gets access to customer information. The former effect dominates the latter if the level of precaution is low (so that third parties have an easy access to customer information).

The posterior beliefs after a good experience and a bad experience, r_G and r_B, are the same as in the baseline model, while the posterior belief after a neutral experience is now given by

$$r_N = \Phi(X,Y) \equiv \frac{1 - Y\lambda - (1-Y)(\lambda + \theta_l)(1-X)}{1 - Y\lambda - (1-Y)(\lambda + \theta_1)(1-X)} r_1. \tag{7}$$

A neutral experience is again good news in the sense that $r_B < r_G < r_N$ for any $X < 1$. The following lemma shows how the posterior belief after a neutral experience depends on the levels of precaution and verification.

Lemma 6 *i)* $\Phi(X,Y)$ *is decreasing in* X*, ii)* $\Phi(X,Y)$ *is decreasing in* Y *for any* $X < 1$*, iii)* $\Phi(0,0) = \phi(0)$ *and* $\Phi(1,Y) = r_1$ *for all* Y*.*

Notice that the range of possible beliefs is the same with or without verification. By reducing exposure to a third party generating a bad experience, verification makes the experience less informative about vulnerability and reduces the posterior r_N whenever $X < 1$. Under full precaution, no bad experience can occur and the posterior is not affected by a neutral or a good experience.

The analysis of the website's decision regarding the level of precaution is similar to the one in the baseline scenario (with no verification). Indeed, as precaution refers to the probability of selling information when verification fails, the website's optimal level of precaution does not depend on the level Y of verification. The equilibrium level of precaution when the website anticipates a posterior r_N is still $X^{br}(r_N)$, as given by equation (2).

We decompose the total gain of not selling between the part related to bad experiences and the rest. Toward this end, define

$$\Delta_G(r_N) \equiv \lambda V_2 [Q(r_G) - Q(r_N)] + (1-\theta_1) v_1,$$

as the website's gain from selling information to the third party when it generates a neutral or a good experience. Define also

$$\Delta_B(r_N) \equiv \theta_1 V_2 [Q(r_N) - Q(r_B)] - \theta_1 v_1$$

as the gain from avoiding selling information when it induces a bad experience. Then

$$P(r_N) - v_1 = \Delta_B(r_N) - \Delta_G(r_N).$$

implying that when $v_1 \in \left(\psi^f V_2, \psi^n V_2\right)$, the equilibrium level of precaution without verification r^M verifies $\Delta_B\left(r^M\right) = \Delta_G\left(r^M\right)$. Notice that, because $\Delta_G(r)$ decreases in r while $\Delta_B(r)$ increases in r, verification is a single-peaked function of r_N achieving a maximum at posterior belief r^M.

Consider now the verification decision. The trade-off faced by the website is different from the one underlying the precaution decision because verification allows to sell customer information only to third parties that generate a good or neutral experience (and would be used only for this purpose since $r_B < r_G$). For a given level of precaution X, verification raises the probability to sell customer information to a third party generating a good or neutral experience from $1-X$ to 1 and, therefore, yields a benefit $X\Delta_G(r_N)$ from selling customer information to such a third party more often. Verification also reduces the probability to sell customer information to a third party that generates a bad experience from $1-X$ to 0,

which leads to another expected benefit given by $(1-X)\,\Delta_B(r_N)$. The total benefit from verifying third parties' use of information is then the sum $X\Delta_G(r_N)+(1-X)\,\Delta_B(r_N)$. When evaluated at $X=X^{br}(r_N)$, the equilibrium condition (2) implies that this gain is equal to $\min(\Delta_B(r_N),\Delta_G(r_N))$. Thus, some verification occurs (i.e. $Y>0$) whenever this benefit is positive, and the website's optimal verification level $Y^{br}(r_N)$ is given by

$$C_v'\left(Y^{br}(r_N)\right)=\max\left\{\min\left(\Delta_G(r_N),\Delta_B(r_N)\right),0\right\}. \tag{8}$$

Thus, an equilibrium is characterized by X^{**}, Y^{**} and r_N^{**} that solve equations (2), (7) and (8). From equation (2), we can distinguish three scenarios:

- *Full precaution* occurs when $r_N^{**}=r_1$ and the level of verification is $C_v'(Y^{**})=\max\{\Delta_G(r_1),0\}$;
- *No precaution* requires a low posterior belief $r_N^{**}<r^M$ and a level of verification given by $C_v'(Y^{**})=\max\{\Delta_B(r_N^{**}),0\}$.
- *Partial precaution* is only possible if the posterior belief after a neutral experience satisfies $r_1<r_N^{**}=r^M<\phi(0)$, and the level of verification is given by $C_v'(Y^{**})=\max\left\{\Delta^M,0\right\}$ where $\Delta^M\equiv\Delta_G\left(r^M\right)=\Delta_B\left(r^M\right)$.

For v_1 close to 0, the gain $\Delta_G(r_N)$ can be made arbitrarily close to 0 or negative (because $r_G=r_1=\phi(1)$), while the gain $\Delta_B(r_N)$ remains strictly positive. Thus, for sufficiently small values of v_1, the website chooses a full precaution policy and no verification, so that customer information is never sold. Similarly, for sufficiently large values of v_1, the website chooses no precaution and no verification, so that customer information is sold to all interested third parties. The following equilibrium characterization shows that the results of Proposition 1 pertaining the level of precaution extends to a setting with verification:

Proposition 12 *A unique equilibrium exists.*

(i) Equilibrium precaution is non-increasing in the value of information v_1 and there exist a threshold $v^n\in\left(\psi^fV_2,\psi^nV_2\right)$ such that the website chooses full precaution if $v_1\leq\psi^fV_2$, partial precaution if $v_1\in\left(\psi^fV_2,v^n\right)$ and no precaution if $v_1\geq v^n$;

ii) There exists a threshold $\bar{v}>v^n$ such that the level of verification is positive if and only if $v_1<\bar{v}$. Equilibrium verification is increasing in v_1 in the full- and partial-precaution region and non-increasing in v_1 in the no-precaution region.

Verification always occurs under the full precaution regime because the benefit $\Delta_G(r_1)$ from selling information when it does not generate a bad experience is positive. It also occurs

under the no precaution regime if the benefit $\Delta_B(\phi(0))$ from avoiding sales of information leading to a bad experience is positive, which is the case when v_1 is not too large. Finally, in the partial precaution regime, the gain from verification Δ^M is always positive.

Let us now consider the way the website's equilibrium strategy depends on v_1. As in the case when there is no verification, the equilibrium level of precaution is non-increasing in v_1. When the value of information increases the equilibrium moves toward more frequent access of third parties to customer information, leading to a higher posterior belief r_N. Verification allows to restrict sales to third parties generating good or neutral experiences, which induces a short-term revenue loss that depends on the price v_1 but also on the level of precaution. Under the full precaution regime, raising v_1 makes verification more attractive as it generates more sales. In contrast, under the no precaution regime, raising v_1 makes verification less attractive as it reduces the probability to sell customer information. The proposition shows that the partial precaution regime is similar to the full precaution regime in this respect. Hence, we find a non-monotonic effect of the value of personal information v_1 on the level of verification.

The level of verification Y in our model can be viewed as a form of security investment that reduces the risk of a bad experience from θ to $(1-Y)\theta$. Introducing such a technology affects the learning process and thus the website's incentives to sell information to third parties that are not proved to generate good or neutral experiences. Since the introduction of the verification technology does not affect the website's incentive to exert precaution X for a given posterior belief r_N but depresses the posterior belief for a given level of precaution, it follows that the website exerts (weakly) less precaution when the probability of a bad experience is lower.

The introduction of the verification technology may have one of the following effects on the level of precaution and the posterior belief r_N.

i) First, if $X^* = 1$, they are not affected (this is because $\Phi(1, Y) = r_1$ is not affected by Y).

ii) Second, if $X^* = 0$ and some verification occurs, the level of precaution remains equal to zero and the posterior belief r_N declines.

iii) Finally, if there is partial precaution in the absence of a verification technology, i.e. $0 < X^* < 1$, there are two possible scenarios:

- In the first scenario, the level of precaution declines but remains positive and the posterior belief $r_N = r^M$ is unchanged – this happens when $\Phi\left(0, Y^{br}\left(r^M\right)\right) \geq r^M$;

- In the second scenario, the level of precaution falls from positive to $X^{**} = 0$ and the posterior belief declines to $r_N^{**} = \Phi(0, Y^{**}) < r^M$ – this happens when $\Phi\left(0, Y^{br}\left(r^M\right)\right) <$

r^M .

Notice that in all cases the level of precaution is either unaffected or lower than in the baseline model without verification. In this sense, precaution and verification are substitutes.

Proposition 13 *Verification reduces the equilibrium level of precaution. More generally, a reduction in the marginal cost of verification raises Y^{**} and reduces X^{**}.*

Welfare and policy implications Consider the effect of a uniform reduction in the marginal cost of verification that either makes verification profitable or makes it easier when it is already profitable. Such a reduction of the verification cost could result from technological advances or policy measures. For instance, the law may impose an obligation of transparency on data buyers or policy may promote public or private certification improving information on potential data buyers. From the above analysis it follows that any technology or policy change that would reduce the cost of verification would lead to less precaution, more verification and (weakly) lower posterior belief r_N.

Implementing a verification technology may raise or lower profit. Indeed, when r_N is not affected, the website benefits from having more instruments and a simple revealed preference argument shows that profit increases. However, this may be countervailed by a reduction in the posterior r_N that reduces long-term profits. Notice that such reduction of r_N can only occur if the availability of the verification technology results in no precaution.

Let us now consider the consumer welfare implications of verification. We show in the proof of the next proposition that verification reduces the equilibrium probability of a bad experience and raises the probability of a good experience. However by denying access to third-parties verified to generate a bad experience, it impedes the consumer's learning process. We thus reach the following conclusion.

Proposition 14 *The introduction of the verification technology raises short-term consumer utility and reduces long-term consumer utility.*

Introducing verification technology into the baseline model strengthens our perspective that simple regulations that reliably improve consumer welfare are elusive. Consider, for example, a case in which $X^{**} = 0$ and $Y^{**} > 0$; in equilibrium, the website verification partially screens out bad actors, but otherwise exercises no precaution. In this case, *ex post* transparency of the level of precaution would create an incentive for the website to reduce verification (due to larger posterior r_N). The consequence is to reduce short-term consumer welfare because consumers are more exposed to bad experiences in the first period,

but to increase long-term utility because consumer learn about their vulnerability from a neutral experience. Thus, in comparison to the baseline model, *ex post* transparency has an ambiguous effect on consumer welfare in the model with verification. Notice that an *ex ante* transparency policy that doesn't constrain verification would again be inconsequential. Similarly a mandatory opt-out policy reduces a website's incentive for verification, harming consumers in the short-run, while improving long-term consumer welfare. Also, opt-out and reduced verification could create incentives for precaution, further reducing short-term consumer utility if $M(r_1) > 0$, and with countervailing effects on long-term utility.

6.2 Learning from good news

We have emphasized the case where consumers learn about the probability of a bad event as we view it as particularly relevant. However, consumers may also be unsure about the likelihood of a good experience and thus learn from good experiences. This is the case when the parameter λ is an unknown characteristic of the consumer. To address this, we now assume that this parameter depends on the type of each consumer. A consumer of type l is characterized by probabilities of good and bad experiences (λ_l, θ_l), while a consumer of type h is characterized by (λ_h, θ_h). We denote by λ_1 and θ_1 the *ex ante* means of these parameters. When both λ and θ are unknown, the consumer expected match utility is

$$M(r) \equiv r\left(\lambda_l U_G + \theta_l U_B\right) + (1-r)\left(\lambda_h U_G + \theta_h U_B\right).$$

As before, we suppose that type l has higher expected match utility, which amounts to assuming that

$$(\lambda_h - \lambda_l)\left(\frac{U_G}{-U_B}\right) < \theta_h - \theta_l.$$

Under this assumption, $M(r)$ is increasing and, therefore, the expected second-period revenue of the website is increasing in the posterior belief r_N that the type is l.

The equilibrium analysis follows the same lines as before except that i) the posterior belief that the type is l after a good experience is $r_G = (\lambda_l/\lambda_1)\, r_1$, and (ii) the posterior belief that the type is l after a neutral experience is

$$r_N = \phi(X) \equiv \frac{1-(\lambda_l+\theta_l)(1-X)}{1-(\lambda_1+\theta_1)(1-X)} r_1. \tag{9}$$

Notice that both r_G and r_N may be greater or smaller that the initial belief r_1.

Consider first the scenario in which $\lambda_l \leq \lambda_h$ and $\lambda_l + \theta_l < \lambda_h + \theta_h$. In this case, $\phi(X)$ is decreasing and larger than r_G and r_B, as in the baseline model. Consequently, both the

equilibrium analysis and the welfare and policy analysis carry over with no (qualitative) changes.[42]

Suppose now that $\lambda_l > \lambda_h$ and $\lambda_l + \theta_l < \lambda_h + \theta_h$. The function $\phi(X)$ is still decreasing but it is now smaller than r_G. The equilibrium is still unique but the terms ψ^f and ψ^n can be negative, which expands the range of values v_1 for which the website exerts no precaution in equilibrium. In particular if $\psi^n < 0$, there is no precaution for any value of information v_1.

Consider finally the scenario $\lambda_l + \theta_l > \lambda_h + \theta_h$, i.e. the scenario in which the consumers benefiting most from matching are those having a neutral experience the least often. In this case, a neutral experience is bad news, i.e. $r_N < r_1$, and the function $\phi(X)$ is increasing. It is then straightforward to see that there will be three equilibria for an intermediate range of values v_1 (and a unique one out of this range), still characterized by $X^* \in \arg\max_{X\in[0,1]} X\left(P\left(\phi\left(X^*\right)\right) - v_1\right)$. In that range, both $X^* = 0$ and $X^* = 1$ are equilibria, and there is in addition an interior equilibrium with $0 < X^* < 1$ (that is unstable). The intuition is that in this case more (less) precaution anticipated by consumers makes precaution more (less) attractive to the website.

6.3 Elastic demand

We now assume that first-period demand is elastic and show how the equilibrium analysis extends to this case.

In the baseline setting where the privacy policy is not observable, the mass of consumers is endogenous and depends on the anticipated privacy policy. Suppose that in period 1, consumers observe an (imperfect) signal of their utility. Thus, the expected utility u_1 at the beginning of period 1 depends on the received signal and is randomly distributed within the population, according to some distribution F_0. Consumers who decide to visit the website in period 1 observe their utility u and then decide whether to visit the website in period 2. A consumer anticipating precaution level X^a visits the website in period 1 if and only if

$$u_1 + (1 - X^a) M(r_1) + \delta^C U_2(r_N, X^a, u_1) \geq \delta^C \max\{0, u_1 + M(r_1)\},$$

where $U_2(r_N, X^a, u_1) \equiv \mathbb{E}\left\{\max\left(u + M(r), 0\right) \mid X^a, u_1\right\}$. Let $\mathcal{U}(X^a)$ be the set of values of u_1 for which consumers visit the website in period 1.

The equilibrium analysis is the same as in Section 3.1 except that the cumulative dis-

[42] The only difference is that in the opt-out section, the threshold for opting out is now $\bar{r} \equiv \frac{\lambda_h U_G + \theta_h U_B}{(\lambda_h - \lambda_l) U_G + (\theta_h - \theta_l) U_B}$.

tribution function $F(u)$ has to be replaced with $F(u_1; X^a) = \Pr(u \leq u_1 \mid u_1 \in \mathcal{U}(X^a))$. Indeed, as demand is exogenous from the perspective of the website, its policy aims at maximizing the total (intertemporal) revenue per user, hence its profit for an average distribution $F(.; X^a)$. Let us define the functions $Q(r; X^a)$ and $P(r_N; X^a)$ as in Section 3.1 by replacing distribution F with distribution $F(.; X^a)$, and assume that there are continuous in X^a. An equilibrium exists and is characterized as follows:

- If $(\theta_1 + \lambda) Q(r_1; 1) - \theta_1 Q(r_B; 1) - \lambda Q(r_G; 1) \geq v_1/V_2$, then full precaution is an equilibrium and first-period demand is $\Pr(u_1 \in \mathcal{U}(1))$.
- If $(\theta_1 + \lambda) Q(\phi(0); 0) - \theta_1 Q(r_B; 0) - \lambda Q(r_G; 0) \leq v_1/V_2$, then no precaution is an equilibrium and first-period demand is $\Pr(u_1 \in \mathcal{U}(0))$.
- If none of the two conditions above holds, a partial precaution equilibrium exists such that $P(\phi(X^*); X^*) = v_1$ and first-period demand is $\Pr(u_1 \in \mathcal{U}(X^*))$.

Note that, in contrast to the baseline model with an inelastic first-period demand, multiple equilibria may exist because $P(\phi(X); X)$ need not be decreasing in X.

In terms of policy implications, note first that the conclusions regarding *ex post* transparency in the baseline model still hold (adjusting for the change in demand). However, an important difference with the baseline model is that under *ex ante* transparency, the website takes into account the effect of its privacy policy on first-period demand. Transparency may induce a larger level of precaution in the case where *ex ante* expected match values are negative enough, if the gain in terms of demand outweighs the costs of lower retention rate. However, if the expected match values are positive, the website would like to boost demand and retention by committing to lower precaution but cannot do so, implying again that *ex ante* transparency has no effect.

Consider now the impact of an opt-out policy. In the current setting, offering the ability to opt out to consumers has the additional effect of boosting first-period demand (for a given precaution level), which makes this more attractive for the website. In particular, the website will still offer voluntarily the ability to opt out if $Q(\bar{r})a \geq Q(r_B)(a + v_2)$. Given that under the voluntary regime opt-out is offered more often than in the baseline model, it is not clear whether an elastic first-period demand makes the mandatory regime more or less desirable.

7 Conclusion

Imperfect information creates incentives for a website to protect consumer privacy. Our model demonstrates this in a novel way by assuming that consumers who visit a website learn from experience about their vulnerability to intrusions due to the website sharing personal information with third parties, and that consumers who become pessimistic about their vulnerability are less likely to return to the website. In response, the website exercises precaution in dealing with third parties and verifies third party use of customer information, in order to profit from better consumer retention.

Our analysis shows how a website's incentive for privacy protection improves with the value of consumer retention relative to the revenue from sharing personal information, the sensitivity of consumer retention to consumer beliefs about vulnerability, and the sensitivity of consumer beliefs to experience. Greater privacy protection, however, is a mixed blessing for consumers, who, on the one hand, are better protected from intrusions, but, on the other hand, may be deprived of positive matches with third parties and are less informed about their vulnerability to third-party intrusions. Consequently, it is difficult for authorities to regulate privacy protection in a way that reliably improves consumer welfare. For example, policies that improve the transparency of privacy policies and give consumers more control over their personal information have either mixed or neutral effects on consumer welfare. Perhaps the best case for policy is to enforce a credible voluntary opt-out regime that enables websites to commit to allow consumers to choose their privacy regime.

There are many interesting directions for further research. One is to assume that consumers have some ability to protect themselves by concealing their identities when returning to a website, e.g. by endogenously removing cookies. Another is to allow websites to charge a subscription fee for continued access, possibly enabling them to better control their own incentives for privacy protection. Finally, studying alternative models of multi-homing by consumers and competition between websites may yield richer insights.

Appendix

Proof of Proposition 1. Clearly, since $P(r_N) - v_1$ decreases in v_1 from positive to negative values, full precaution must be an equilibrium for sufficiently small values of v_1/V_2; a necessary and sufficient condition is $P(\phi(1)) - v_1 \geq 0$ which yields the threshold ψ^f. Similarly, no precaution is an equilibrium if and only if v_1/V_2 is sufficiently large that $P(\phi(0)) - v_1 \leq 0$ which yields ψ^n. Thus, there is no pure strategy equilibrium if $P(\phi(0)) - v_1 > 0 > P(\phi(1)) - v_1$. In this range of v_1/V_2, there exists a unique belief that holds

the website indifferent about precaution, and this belief pins down equilibrium precaution: $P\left(r^M\right) = v_1$ and $\phi\left(X^*\right) = r^M$. Monotonicity follows from r^M increasing in v_1/V_2 and $\phi\left(X\right)$ decreasing.

Proof of Lemma 1. The function $v\left(r\right) \equiv \mathbb{E}\left\{\max\left(\max_{J\subset\{1,..,K\}} \sum_k u_k + M\left(r\right), 0\right)\right\}$ is convex in r. Moreover, if $r_N = \phi\left(X\right)$, an increase in X induces a mean-preserving contraction of the distribution of r; this follows from $\mathbb{E}\left\{r \mid X\right\} = r_1$, $p_G\left(X\right)$ and $p_N(X)$ both decreasing, and r_N decreasing in X. Therefore, from Rothschild and Stiglitz (1971), $\mathbb{E}\left\{v\left(r\right) \mid X\right\}$ is decreasing in X. The result follows because $U_K\left(\phi\left(X\right), X\right) = \mathbb{E}\left\{v\left(r\right) \mid X\right\}$.

Proof of Proposition 2. If all other websites quote $p = 0$, then the future payoff of a website is independent of the price it sets, and is equal to $L_K\left(r_N\right)$. Therefore, quoting $p = 0$ is a best reply.

Proof of Proposition 3. An equilibrium with full precaution ($x = X = 1$) induces $r_N = \phi\left(1\right)$ and exists if and only if $v_1 \leq P_K\left(\phi\left(1\right)\right)$. Consider now a symmetric equilibrium with $0 < x < 1$ and thus $X = x^K$, and $r_N = \phi\left(X\right)$. For any $p \leq v_1$ on the support of the equilibrium strategy we must have

$$p\left[x + \left(1 - x\right)\left(1 - H\left(p\right)\right)\right]^{K-1} = x^{K-1} p_{\max}$$

where $p_{\max}$ is the upper bound of the support. This leads to a cumulative distribution function

$$H\left(p\right) = \frac{1 - x\left(\frac{p_{\max}}{p}\right)^{\frac{1}{K-1}}}{1 - x} \text{ on an interval } \left[p_{\inf}, p_{\max}\right] \text{ with } p_{\max} \leq v_1$$

Notice that there cannot be a mass point because it could be undercut profitably. Moreover, we must have $p_{\max} = v_1$ because otherwise setting $p = v_1$ would strictly dominate setting $p = p_{\max}$. Thus, we have

$$H\left(p\right) = \frac{1 - x\left(\frac{v_1}{p}\right)^{\frac{1}{K-1}}}{1 - x} \text{ on the interval } \left[x^{K-1} v_1, v_1\right].$$

The equilibrium payoff is then

$$L_K\left(r_N\right) + x_i x^{K-1} P_K\left(r_N\right) + \left(1 - x_i\right) x^{K-1} v_1,$$

implying that an interior equilibrium verifies $v_1 = P_K\left(\phi\left(X_K^*\right)\right)$. Given that $P_K\left(\phi\left(X\right)\right)$ is decreasing in X, the solution to the equation $v_1 = P_K\left(\phi\left(X\right)\right)$ exists in $(0, 1)$ and is unique

when $P_K(\phi(1)) < v_1 < P_K(\phi(0))$. This implies that the equilibrium exists and is uniquely defined for this range of values of v_1. Thus, we have

$$H(p) = \frac{1 - (X_K^*)^{\frac{1}{K}} \left(\frac{v_1}{p}\right)^{\frac{1}{K-1}}}{1 - (X_K^*)^{\frac{1}{K}}}$$

which gives $p_{\inf} = (X_K^*)^{\frac{K-1}{K}} v_1$.

Proof of Proposition 4. (i) Let us consider the case of perfect correlation, i.e. $u^i = u$ for all i, where u is distributed according to F. When $M(r)$ is negative, the total retention rate and the unique retention rate are given by:

$$Q_K(r) = 1 - F\left(-\frac{M(r)}{K}\right) \text{ and } \hat{Q}_K(r) = 0.$$

If $M(r)$ is positive, the consumer visits in the second period only one website (assumed to be chosen randomly) whenever the expected utility from third-party matching is mildly negative. More specifically, this occurs when $0 > u > -M(r)$. So for positive $M(r)$ we have:

$$Q_K(r) = 1 - F(0) + \frac{F(0) - F(-M(r))}{K} \text{ and } \hat{Q}_K(r) = \frac{F(0) - F(-M(r))}{K}.$$

In both cases, $Q'_K(r)$ and $\hat{Q}'_K(r)$ are smaller than $Q'(r)$. This holds because $\frac{1}{K}F(-M(r)/K) < F(-M(r))$ if $M(r) < 0$ and $\frac{1}{K}F(-M(r)) < F(-M(r))$ if $M(r) > 0$. Consequently, the expected future gain from not selling is less sensitive to posteriors r than in the single-homing case, i.e. $P_K(r_N^*) < P(r_N^*)$, implying that there is less precaution with multi-homing for any a and v_2.

(ii) Assume now that $\Pr(u^j > 0 | u^i) > 0$ for $j \neq i$. Note that the unique retention rate can be written as

$$\hat{Q}_K(r) = \int_{-M(r)}^{+\infty} \Pr\left(\max_{j \neq i} u^j < \min\left(u^i, 0\right) | u^i\right) f\left(u^i\right) du^i$$

where $f = F'$, which implies that

$$\hat{Q}'_K(r) = \Pr\left(\max_{j \neq i} u^j < \min\left(-M(r), 0\right) | -M(r)\right) Q'(r) \leq Q'(r).$$

We have $\Pr\left(\max_{j\neq i} u^j < \min\left(u^i, 0\right)|u^i\right) < 1$ and, therefore, $\hat{Q}'_K\left(r\right) < Q'\left(r\right)$ for all r. Since $r_N^* \geq r_G = r_1 > r_B$, we conclude that if a/v_2 is small, $P_K\left(r_N^*\right) < P\left(r_N^*\right)$.

Suppose $M\left(r\right) > 0$, then

$$Q_K\left(r\right) = 1 - F\left(0\right) + \int_{-M(r)}^{0} \Pr\left(\max_{j\neq i} u^j < u^i|u^i\right) f\left(u^i\right) du^i$$

and

$$Q'_K\left(r\right) = M'\left(r\right) \Pr\left(\max_{j\neq i} u^j < -M\left(r\right)|u^i = -M\left(r\right)\right) f\left(-M\left(r\right)\right) \leq Q'\left(r\right).$$

Hence again, $P_K\left(r_N^*\right) < P\left(r_N^*\right)$.

However if $M\left(r\right) < 0$, then

$$Q_K\left(r\right) = Q\left(r\right) + \int_{0}^{-M(r)} \left(1 - H\left(-M\left(r\right) - u^i|u^i\right)\right) f\left(u^i\right) du^i.$$

where H is the cdf of $\sum_{j\neq i} \max\left(u^j, 0\right)$. Therefore

$$Q'_K\left(r\right) - Q'\left(r\right) = M'\left(r\right)\left(\int_{0}^{-M(r)} h\left(-M\left(r\right) - u^i|u^i\right) f\left(u^i\right) du^i - \left(1 - H\left(0| - M\left(r\right)\right) f\left(-M\left(r\right)\right)\right)\right)$$

may be either positive of negative.

Computations for Example 2. We have

$$\hat{Q}_2\left(r\right) = \Pr\left\{u^1 > -M\left(r\right),\ u^2 < 0,\ u^1 > u^2\right\}$$

which leads to

$$\hat{Q}_2(r) = \alpha^2 \int_{-M(r)}^{\bar{u}} \left(\min\left(0, u\right) - \underline{u}\right) du$$

where we use the double inequality $\bar{u} > \min\left(0, u^1\right) > \underline{u}$. Then, denoting $Q^{12}\left(r\right)$ the expected joint retention rate of both firms, we have

$$Q^{12}\left(r\right) = \Pr\left\{u^1 + u^2 > -M\left(r\right),\ u^2 > 0,\ u^1 > 0\right\} = \alpha^2 \int_{0}^{\bar{u}} \left(\bar{u} - \max\left(0, -u - M\left(r\right)\right)\right) du,$$

where we use $u > 0 \Rightarrow \bar{u} > \max\left(0, -M\left(r\right)\right) > \max\left(0, -u - M\left(r\right)\right) > 0 > \underline{u}$.

Summing up, we have

$$\begin{aligned} if \quad M(r) \quad &\geq \quad 0 : \hat{Q}_2(r) = \alpha^2 \int_{-M(r)}^{0} (u - \underline{u})\, du - \underline{u}\bar{u} = \alpha^2 \left(-\frac{1}{2} M(r)^2 - \underline{u} M(r) - \underline{u}\bar{u} \right) \\ if \quad M(r) \quad &\leq \quad 0 : \hat{Q}_2(r) = \alpha^2 \left(-\underline{u} M(r) - \underline{u}\bar{u} \right) \end{aligned}$$

and

$$\begin{aligned} if \quad M(r) \quad &\geq \quad 0 : Q^{12}(r) = \alpha^2 \bar{u}^2 \\ if \quad M(r) \quad &\leq \quad 0 : Q^{12}(r) = \alpha^2 \left(\int_{-M(r)}^{\bar{u}} \bar{u} du + \int_{0}^{-M(r)} (\bar{u} + u + M(r))\, du \right) = \alpha^2 \left(\bar{u}^2 - \frac{M(r)^2}{2} \right) \end{aligned}$$

Thus, irrespective of the sign of $M(r)$, the total retention rate is

$$Q_2(r) = \alpha^2 \left(-\frac{1}{2} M(r)^2 - \underline{u} M(r) - \underline{u}\bar{u} + \bar{u}^2 \right) = \alpha^2 \left(-\frac{1}{2} M(r)^2 - \underline{u} M(r) \right) + \alpha \bar{u}.$$

Let us assume now $M(r_1) \geq 0 > M(0)$. For $K = 2$, the gain from precaution writes

$$\begin{aligned} P_2(r_N) \quad \equiv \quad & \delta^F a \left\{ \lambda \left[Q_2(r_N) - Q_2(r_G) \right] + \theta_1 \left[Q_2(r_N) - Q_2(r_B) \right] \right\} \\ & + \delta^F v_2 \left\{ \lambda \left[\hat{Q}_2(r_N) - \hat{Q}_2(r_G) \right] + \theta_1 \left[\hat{Q}_2(r_N) - \hat{Q}_2(r_B) \right] \right\}. \end{aligned}$$

Using the above expressions for the unique and total retention rates, we get

$$\begin{aligned} P_2(r_N) \quad \equiv \quad & \delta^F a \alpha^2 \left\{ \begin{array}{c} \lambda \left[\left(-\frac{1}{2} M(r_N)^2 - \underline{u} M(r_N) \right) - \left(-\frac{1}{2} M(r_1)^2 - \underline{u} M(r_1) \right) \right] \\ + \theta_1 \left[\left(-\frac{1}{2} M(r_N)^2 - \underline{u} M(r_N) \right) - \left(-\frac{1}{2} M(0)^2 - \underline{u} M(0) \right) \right] \end{array} \right\} \\ & + \delta^F v_2 \alpha^2 \left\{ \begin{array}{c} \lambda \left[\left(-\frac{1}{2} M(r_N)^2 - \underline{u} M(r_N) - \underline{u}\bar{u} \right) - \left(-\frac{1}{2} M(r_1)^2 - \underline{u} M(r_1) - \underline{u}\bar{u} \right) \right] \\ + \theta_1 \left[\left(-\frac{1}{2} M(r_N)^2 - \underline{u} M(r_N) - \underline{u}\bar{u} \right) - \left(-\underline{u} M(0) - \underline{u}\bar{u} \right) \right] \end{array} \right\}, \end{aligned}$$

which can be rewritten as

$$\begin{aligned} P_K(r_N) \equiv\ & \delta^F(a+v_2)\alpha^2\lambda\left[-\frac{1}{2}M(r_N)-\frac{1}{2}M(r_1)-\underline{u}\right](M(r_N)-M(r_1)) \\ & +\delta^F a\theta_1\alpha^2\left[-\frac{1}{2}M(r_N)-\frac{1}{2}M(0)-\underline{u}\right](M(r_N)-M(0)) \\ & +\delta^F v_2\alpha^2\theta_1\left[-\frac{1}{2}M(r_N)^2-\underline{u}M(r_N)+\underline{u}M(0)\right], \end{aligned}$$

Moreover, we have by interiority

$$\begin{aligned} \alpha\left[-\frac{1}{2}M(r_N)-\frac{1}{2}M(r_1)-\underline{u}\right] &< 1, \\ \alpha\left[-\frac{1}{2}M(r_N)-\frac{1}{2}M(0)-\underline{u}\right] &< 1, \end{aligned}$$

$$\alpha\left[-\frac{1}{2}M(r_N)^2-\underline{u}M(r_N)+\underline{u}M(0)\right] < -\alpha\underline{u}(M(r_N)-M(0)) < M(r_N)-M(0).$$

In the single-homing case, we have

$$P(r_N)=\delta^F(a+v_2)\alpha\lambda(M(r_N)-M(r_1))+\delta^F(a+v_2)\alpha\theta_1(M(r_N)-M(0)).$$

Therefore,

$$P_K(r_N) < P(r_N).$$

Proof of Lemma 2. The website's expected payoff is $L(\phi(X))+XP(\phi(X))+(1-X)v_1$. The proof follows immediately from the fact that, under *ex post* transparency, the website's net marginal gain from precaution is given by $P(\phi(X))+\phi'(X)[L'(\phi(X))+P'(\phi(X)]-v_1$. Notice an equilibrium always exists because finding an equilibrium reduces to a simple maximization problem in this case.

Proof of Lemma 3. Consider an equilibrium with $0<x<1,\ X=x^K,\ r_N=\phi(X)$. For any $p\leq v_1$, replicating the reasoning used to derive the partial precaution equilibrium in Proposition 3, we must have a cumulative distribution function for prices given by

$$H(p)=\frac{1-X^{\frac{1}{K}}\left(\frac{v_1}{p}\right)^{\frac{1}{K-1}}}{1-X^{\frac{1}{K}}} \text{ on an interval } [X^{\frac{K-1}{K}}v_1, v_1].$$

Now, consider an equilibrium with $0 < x \leq 1$ and suppose that a website deviates to x_i and $p \leq v_1$. Then website i's expected payoff is $L_K\left(\phi(x_i x^{K-1})\right) + x_i x^{K-1} P_K\left(\phi(x_i x^{K-1})\right) + (1 - x_i) x^{K-1} v_1$. Therefore, the website's net marginal gain from precaution at a symmetric equilibrium with aggregate precaution X is given by

$$P_K(\phi(X)) + \left[X P_K'(\phi(X)) + L_K'(\phi(X))\right] \phi'(X) - v_1,$$

which yields the result.

Notice that website i's expected payoff is quasi-concave in x_i if $L_K(\phi(X)) + X P_K(\phi(X)) - X v_1$ is quasi-concave in X.[43] In this case, there exists a unique non-trivial symmetric equilibrium (i.e. an equilibrium different from the trivial one with zero precaution which always exists) if and only if $v_1 > P(\phi(0)) + L_K'(0)\phi'(0)$.

Furthermore, if $L_K(\phi(X)) + X P_K(\phi(X)) - X v_1$ is convex in X then is an equilibrium with $X^* = 1$ if and only if $L_K(\phi(1)) + P_K(\phi(1)) - v_1 \geq L_K(\phi(0))$, and the only equilibrium is $X^* = 0$ otherwise.

Proof of Proposition 5. Consider first the case $K = 1$. As $r_N = \phi(X)$ is decreasing in X, the marginal gain of the website from increasing X is lower when it can commit to its strategy: $\frac{\partial \Pi(r_N, X)}{\partial X} + \frac{\partial \Pi(r_N, X)}{\partial r_N}\phi'(X) < \frac{\partial \Pi(r_N, X)}{\partial X}$. This implies that a full precaution equilibrium exists for a smaller range of values v_1/V_2 while a no precaution equilibrium exists for a wider range. Consider now an equilibrium with no commitment featuring an interior level of precaution $X^* \in (0, 1)$. Then for any $X > X^*$, $\Pi(\phi(X^*), X^*) > \Pi(\phi(X^*), X) > \Pi(\phi(X), X)$. Therefore, the website chooses $X \leq X^*$. Moreover, it holds that $\frac{\partial \Pi}{\partial X}(\phi(X^*), X^*) + \frac{\partial \Pi}{\partial r_N}(\phi(X^*), X^*)\phi'(X^*) = \frac{\partial \Pi}{\partial r_N}(\phi(X^*), X^*)\phi'(X^*) < 0$, which implies that the website chooses $X < X^*$.

Consider now $K \geq 2$ and a symmetric equilibrium with $\tilde{X}_K^* \in (0, 1)$. We have

$$P_K\left(\phi\left(\tilde{X}_K^*\right)\right) - v_1 = -\left(\tilde{X}_K^* P_K'\left(\phi\left(\tilde{X}_K^*\right)\right) + L_K'\left(\phi\left(\tilde{X}_K^*\right)\right)\right)\phi'\left(\tilde{X}_K^*\right) > 0 = P_K(\phi(X_K^*)) - v_1,$$

which implies that $\tilde{X}_K^*$ is weakly lower than X_K^*.

Proof of Lemma 4. Denote X_2 the website's precaution level in period 2 and $V_2(X_2) = a + (1 - X_2) v_2$ the expected revenue generated in period 2 by a returning consumer. Denoting

[43] To see why, notice that, for given precaution x by aother websites, maximizing website i's individual payoff amounts to maximizing $L_K\left(\phi(x_i x^{K-1})\right) + x_i x^{K-1} P_K\left(\phi(x_i x^{K-1})\right) - x_i x^{K-1} v_1$ with respect to $x_i \in [0, x^{K-1}]$, which is the same as maximizing $L_K(\phi(X)) + X P_K(\phi(X)) - X v_1$ with respect to $X \in [0, x^{K-1}]$.

first-period precaution X_1 (instead of X as in the rest of the paper), we have

$$\frac{\partial \Pi}{\partial X_2} = -\mathbb{E}\left\{F'\left(-\left(1-X_2\right)M(r)\right)M(r) \mid X_1\right\}V_2\left(X_2\right) + \mathbb{E}\left\{1-F\left(-\left(1-X_2\right)M(r)\right) \mid X_1\right\}V_2'\left(X_2\right).$$

The second term is negative. The first term is non-positive if $M(r_B) \geq 0$. Therefore, $\partial\Pi/\partial X_2$ is negative –and consequently $X_2 = 0$ is optimal– if $M(r_B)$ is not too negative. Alternatively, if the probability of a bad event $p_B\left(X_1\right)$ is small enough, which is the case if θ_1 is small enough, then the first term is non-positive and, again, $\partial\Pi/\partial X_2$ is negative. Finally, since $\mathbb{E}\left\{r \mid X_1\right\} = r_1$, another sufficient condition for the first term to be non-positive is that

$$\mathbb{E}\left\{-F'\left(-\left(1-X_2\right)M(r)\right)M(r) \mid X_1\right\} \leq -F'\left(-\left(1-X_2\right)M(r_1)\right)M(r_1) \leq 0$$

which holds if $-F'\left(-\left(1-X_2\right)M(r)\right)M(r)$ is a concave function of r, or equivalently (recalling that $M(r)$ is linear), that $uF'\left(u\right)$ is concave over the relevant range.

Proof of Proposition 6. Follows from the analysis above the proposition.

Proof of Lemma 5. The first step is to derive the two firms' second-period equilibrium profits when one firm, say firm 1, chooses $x_1 \geq 0$ in the second period while the other one exerts no precaution in that period (for the sake of notational convenience, we use x_i to denote firm i's second period precaution level in this proof).

Denote by $\hat{Q}^i$ the expected unique retention rate of firm $i \in \{1,2\}$ and by Q^{12} the expected joint retention rate of both firms (the expectation is over beliefs of the mass of consumers joining both websites). Also denote $Q^i = \hat{Q}^i + Q^{12}$. Notice that these retention rates depend on x_1. Notice also that if firm i serves a consumer, it does not know whether it is unique or not. The conditional probability of a unique retention is $\hat{Q}^i/Q^i$.

Assume firm 2 sets $x_2 = 0$ while firm 1 sets $x_1 \geq 0$ in the second period. In any equilibrium, with the same argument as before, the support of prices is the same for both firms, $[\underline{p}, v_2]$, but one firm has a mass point at v_2.We proceed by considering two types of equilibria depending on which firm has a mass point.

Assume that firm 2 sets a mass point μ_2 at v_2. Denote Π_i the *ex ante* profit firm i makes from selling information in the second period. This profit can be written as $\Pi_i = \pi_i Q_i$ where π_i is the profit from selling information on a retained consumer. Then, for any p in $[\underline{p}, v_2)$, profits from selling information are

$$\begin{aligned}
\Pi_1 &= \left(1-x_1\right)p\left(\hat{Q}^1 + Q^{12}\left(1-F_2\left(p\right)\right)\right) = \left(1-x_1\right)v_2\left(\hat{Q}^1 + Q^{12}\mu_2\right), \\
\Pi_2 &= p\left(\hat{Q}^2 + Q^{12}x_1 + Q^{12}\left(1-x_1\right)\left(1-F_1\left(p\right)\right)\right) = v_2\left(\hat{Q}^2 + Q^{12}x_1\right),
\end{aligned}$$

where the last term in each equation is profit at $v_2(-\varepsilon)$. Thus, indifference on the support yields

$$\begin{aligned}\frac{1}{Q^{12}}\left(\frac{v_2\left(\hat{Q}^1+Q^{12}\mu_2\right)}{p}-\hat{Q}^1\right) &= 1-F_2(p),\\ \frac{\hat{Q}^2+Q^{12}x_1}{Q^{12}(1-x_1)}\left(\frac{v_2}{p}-1\right) &= 1-F_1(p).\end{aligned}$$

The lower bound and the mass point are given respectively by :

$$\underline{p}=v_2\frac{\hat{Q}^2+Q^{12}x_1}{Q^2};\ \mu_2=\frac{\left(Q^1x_1-\hat{Q}^1+\hat{Q}^2\right)}{Q^2}.$$

This is an equilibrium if $\mu_2>0$ which is equivalent to $x_1>\frac{\hat{Q}^1-\hat{Q}^2}{Q^1}$.

Finally, the profit of firm 1 from selling information is in this case

$$\Pi_1=(1-x_1)\,v_2\left(\frac{\hat{Q}^2+Q^{12}x_1}{Q^2}\right)Q^1.$$

The same reasoning shows that there is a mass point μ_1 at v_2 for firm 1 if $x_1<\frac{\hat{Q}^1-\hat{Q}^2}{Q^1}$, and in this case the equilibrium is given by

$$\begin{aligned}\frac{1}{Q^{12}}\left(\frac{v_2\hat{Q}^1}{p}-\hat{Q}^1\right) &= 1-F_2(p)\\ \frac{1}{Q^{12}(1-x_1)}\left(\frac{v_2\left(\hat{Q}^2+Q^{12}x_1+Q^{12}\mu_1\right)}{p}-\left(\hat{Q}^2+Q^{12}x_1\right)\right) &= 1-F_1(p_2).\end{aligned}$$

The lower bound and the mass point are now:

$$\underline{p}=v_2\frac{\hat{Q}^1}{Q^1}=v_2\frac{\left(\hat{Q}^2+Q^{12}x_1+Q^{12}\mu_1\right)}{Q^2},\ \mu_1=\frac{\hat{Q}^1-\hat{Q}^2}{Q^1}-x_1.$$

The profit from selling information is then $\Pi_1=(1-x_1)\,v_2\hat{Q}^1$. Notice that the profit is continuous in x_1 and in retention rates $\hat{Q}^1$, $\hat{Q}^2$ and Q^{12}.

We now compare retention rates. In the second period a consumer with beliefs r compares $u^1+(1-x_1)M(r)$, $u^2+M(r)$, $u^1+u^2+M(r)$ and the zero utility from no consumption. Define

$$\begin{aligned}\hat{Q}^1(r) &= \Pr\left\{u^1 > (x_1-1)M(r),\ u^2 < -x_1M(r),\ u^1-u^2 > x_1M(r)\right\},\\ Q^{12}(r) &= \Pr\left\{u^1+u^2 > -M(r),\ u^2 > -x_1M(r),\ u^1 > 0\right\},\\ \hat{Q}^2(r) &= \Pr\left\{u^2 > -M(r),\ u^1 < 0,\ u^2 > u^1 - x_1M(r)\right\}.\end{aligned}$$

Then, for $i=1,2:\hat{Q}^i=\mathbb{E}\left(\hat{Q}^i(r)\right)$ and $Q_i=\mathbb{E}\left(\hat{Q}^i(r)\right)+\mathbb{E}\left(Q^{12}(r)\right)$. Notice that $\hat{Q}^2(r)$ can be rewritten by symmetry (reverting u^1 and u^2):

$$\hat{Q}^2(r)=\Pr\left\{u^1+M(r)>0,\ u^2<0,\ u^2<u^1+x_1M(r)\right\},$$

where all events are more likely than the corresponding events in $\hat{Q}^1(r)$ if $M(r)>0$. Thus, $\hat{Q}^2(r)>\hat{Q}^1(r)$ for all r such that $M(r)>0$, which implies that when $M(r_1)\geq 0$, $\hat{Q}^2>\hat{Q}^1$ if $M(r_B)$ is not too negative or the probability of a bad event is small enough. A sufficient condition for the latter is that θ_1 is small enough.

We conclude that if $M(r_B)$ is not too negative or θ_1 is small enough then

$$\Pi_1^T=\left(a+(1-x_1)v_2\left(\frac{\hat{Q}^2+Q^{12}x_1}{Q^2}\right)\right)Q^1,$$

where all functions are continuous.

By the same reasoning,

$$Q^1=\mathbb{E}\left(\Pr\left\{u^1+\max\left(u^2,-x_1M(r)\right)>-M(r),\ u^1>\min\left(u^2+x_1M(r),0\right)\right\}\right)$$

is strictly decreasing if $M(r_B)$ is not too negative or θ_1 is small enough. As a consequence, Π_1^T is decreasing in x_1 if v_2/a is small enough, which implies that $x_1=0$ is optimal for firm 1; in other words, both firms exerting no precaution in the second period is an equilibrium.

Proof of Proposition 7. See Section 1 of the Online Appendix.

Proof of Proposition 8. Assume that $Q(\bar{r})a\geq Q(r_B)(a+v_2)$. Under the assumption $M(r_1)\geq 0>M(r_B)$, the return to precaution for a given belief r_N under opt out (i.e. the counterpart to $P(r_N)$ in the baseline model) is given by

$$P^{opt}(r_N)=\delta^F\left\{\theta_1\left[(a+v_2)Q(r_N)-aQ(\bar{r})\right]+\lambda(a+v_2)\left[Q(r_N)-Q(r_G)\right]\right\}.$$

Assume that $Q(\bar{r})a\geq Q(r_B)(a+v_2)$. In this scenario, $P^{opt}(r_N)\leq P(r_N)$. Recalling that $P(r_N)$ is increasing in r_N, we distinguish between three cases:

- When $X^* = 0$ in the baseline model we have $P(\phi(0)) \leq v_1$. This implies that $P^{opt}(\phi(0)) \leq v_1$, which implies that equilibrium precaution remains equal to 0.

- When $0 < X^* < 1$ then $P(r_N^*) = v_1$. Since $P^{opt}(r_N) \leq P(r_N)$, the equilibrium posterior belief after a neutral experience under opt-out will be weakly greater than its counterpart in the absence of opt-out.. Since ϕ is decreasing in X this implies that equilibrium precaution is weakly lower under opt-out.

- When $X^* = 1$ then $P(r_N^*) \geq v_1$. In this case, $P^{opt}(r_N) \leq P(r_N)$ implies again that the equilibrium posterior belief after a neutral experience under opt-out will be weakly greater than its counterpart in the absence of opt-out, which entails that equilibrium precaution is weakly lower under opt-out.

Proof of Proposition 9. Straightforward extension of the proof of Proposition 8 to the multi-homing scenario.

Proof of Proposition 10. Follows from the analysis above the proposition.

Proof of Proposition 11. Follows from the analysis above the proposition.

Proof of Lemma 6. Straightforward computations show that $\partial\Phi/\partial X$ has the same sign as $(\theta_l - \theta_1)(1 - Y\lambda)(1 - Y)$, which proves (i), and that $\partial\Phi/\partial Y$ has the same sign as $(1 - X)(\theta_1 - \theta_l)(-1 + \lambda)$, which proves (ii). The proof of (iii) is immediate.

Proof of Proposition 12. *Existence and uniqueness of an equilibrium.* An equilibrium verifies $r_N^{**} = \Phi\left(X^{br}(r_N^{**}), Y^{br}(r_N^{**})\right)$. Therefore, a sufficient condition for a unique equilibrium to exist is that the correspondence $\Phi\left(X^{br}(r_N), Y^{br}(r_N)\right)$ has a unique fixed point. We have:

- For $r_N < r^M$, $X^{br}(r_N) = 0$ and $C_v'(Y^{br}(r_N)) = \max\{\Delta_B(r_N), 0\}$ is non-decreasing, implying that $\Phi\left(X^{br}(r_N), Y^{br}(r_N)\right)$ is non-increasing in r_N.

- For $r_N = r^M$, $X^{br}(r_N) \in [0,1]$ and $C_v'(Y^{br}(r_N)) = \max\left\{\Delta^M, 0\right\}$.

- For $r_N > r^M$, $X^{br}(r_N) = 1$ and $C_v'(Y^{br}(r_N)) = \max\{\Delta_G(r_N), 0\}$ is non-increasing, implying that $\Phi\left(X^{br}(r_N), Y^{br}(r_N)\right)$ is constant in r_N (recall that $\Phi(1, Y) = r_1$ for any Y).

Hence, $\Phi\left(X^{br}(r_N), Y^{br}(r_N)\right)$ is a non-increasing continuous correspondence from $[0,1]$ into itself. This implies that it has a unique fixed point $r_N^{**} = \Phi\left(X^{br}(r_N^{**}), Y^{br}(r_N^{**})\right)$.

Moreover, the graph of the correspondence is continuous in v_1, which implies that r_N^{**} is continuous in v_1.

Proof of (i). Let us first show that r_N^{**} is non-decreasing in v_1 and X^{**} is non-increasing in v_1.

Suppose that $X^{**} = 0$ and let v_1 increase. Then X^{**} remains constant and r_N^{**} cannot decrease. To see why the latter part holds, assume that r_N^{**} decreases locally. Combined with $C_v'(Y^{**}) = \max\{\Delta_B(r_N^{**}), 0\}$, this would imply that Y^{**} is non-decreasing locally,

while combined with $r_N^{**} = \Phi(0, Y^{**})$ and $\Phi(0, Y)$ decreasing with Y, it would imply that r_N^{**} is increasing locally - a contradiction. It follows that Y^{**} is non-increasing in the no-precaution region.

Suppose now that $X^{**} = 1$ and let v_1 increase. Then X^{**} remains constant and so does $r_N^{**} = \Phi(1, Y^{**}) = r_1$.

Suppose finally that $0 < X^{**} < 1$. We known from Proposition 1 that the result holds if $Y^{**} = 0$. Assume now that $C_v'(Y^{**}) = \Delta^M > 0$, and let v_1 increase. We have $r_N^{**} = r^M$ and

$$\frac{dr^M}{dv_1} = \frac{1}{(\lambda + \theta_1) Q'(r^M) V_2} > 0.$$

Moreover, as $r^M = \Phi(X^{**}, Y^{**})$, we have

$$\frac{dr^M}{dv_1} - \frac{\partial \Phi}{\partial Y}\frac{dY^{**}}{dv_1} = \frac{\partial \Phi}{\partial X}\frac{dX^{**}}{dv_1}$$

implying that X^{**} decreases with v_1 if

$$\frac{dr^M}{dv_1} > \frac{\partial \Phi}{\partial Y}\frac{dY^{**}}{dv_1}. \tag{10}$$

Differentiating $C_v'(Y^{**}) = \Delta_B\left(r^M\right) = \theta_1 V_2\left[Q\left(r^M\right) - Q(r_B)\right] - \theta_1 v_1$ with respect to v_1, we get

$$C_v''(Y^{**})\frac{dY^{**}}{dv_1} = \theta_1\left(Q'(r_M)\frac{dr_M}{dv_1}V_2 - 1\right) = \theta_1\left(\frac{Q'(r_M) V_2}{(\lambda + \theta_1) Q'(r^M) V_2} - 1\right) > 0,$$

This, combined with $\partial\Phi/\partial Y < 0$ for $X < 1$, implies that condition 10 holds.

We can therefore conclude that X^{**} is non-increasing in v_1. This, combined with the fact that $X^{**} = 1$ if v_1 is sufficiently small and $X^{**} = 0$ if v_1 is sufficiently large implies that there exist thresholds v^f and v^n such that $X^{**} = 1$ if and only if $v_1 \leq v^f$ and $X^{**} = 0$ if and only if $v_1 \geq v^n$. From $\Phi(1, Y) = r_1$ for all $Y \geq 0$, it follows that $X^{**} = 1$ if and only if $\Delta_G(r_1) \leq \Delta_B(r_1)$, which implies that the threshold v^f is the same as the corresponding threshold in the baseline model, i.e. $v^f = \psi^f V_2$. Moreover, from $\Phi(0, Y) \leq \phi(0)$ and the fact that $\Delta_G(r_N) - \Delta_B(r_N)$ is decreasing in r_N it follows that $\Delta_G(\phi(0)) > \Delta_B(\phi(0))$ whenever $\Delta_G(\Phi(0, Y)) > \Delta_B(\Phi(0, Y))$, which implies that v^n is less than or equal to the corresponding threshold in the baseline model, i.e. $v^n \leq \psi^n V_2$. The subsequent proof of part (ii) shows that $Y^{**} > 0$ when $v_1 = v^n$, which implies that the strict inequality $v^n < \psi^n V_2$ holds.

Proof of (ii). Consider first the case where $X^{**}=1$. Then $r_N^{**}=r_1$, which implies that $\Delta_G(r_1)=(1-\theta_1)v_1>0$. It then follows from $C_v'(Y^{**})=\Delta_G(r_1)$ that $Y^{**}>0$ and Y^{**} is increasing in v_1.

Suppose now that $X^{**}\in(0,1)$. From $\Delta_G\left(r^M\right)=\Delta_B\left(r^M\right)$ it follows that

$$v_1=\left[(\theta_1+\lambda)Q\left(r^M\right)-\theta_1 Q(r_B)-\lambda Q(r_G)\right]V_2.$$

Using this and $\Delta^M=\Delta_B\left(r^M\right)$ we get

$$\Delta^M=\theta_1\left[(1-\theta_1-\lambda)Q\left(r^M\right)+\lambda Q(r_G)-(1-\theta_1)Q(r_B)\right]V_2,$$

which implies that $\Delta^M>0$ because $Q\left(r^M\right)>Q(r_G)>Q(r_B)$. Therefore, $Y^{**}=(C_v')^{-1}\left(\Delta^M\right)>0$.

Let us now turn to the case where $X^{**}=0$. Then $Y^{**}=0$ if and only if $\Delta_B(\phi(0))\leq 0$ which writes as

$$v_1\geq\bar{v}\equiv\left(Q(\phi(0))-Q(r_B)\right)V_2.$$

From $Q(\phi(0))-Q(r_B)>(\theta_1+\lambda)Q(\phi(0))-\theta_1Q(r_B)-\lambda Q(r_G)=\psi^n$, it follows that $\bar{v}>\psi^nV_2$ and consequently $\bar{v}>v^n$.

Finally, consider the way Y^{**} is affected by v_1. In the full precaution regime, $C_v'(Y^{**})=\Delta_G(r_1)$ is increasing in v_1, which implies that Y^{**} is increasing in v_1. In the no precaution regime, Y^{**} is non-increasing in v_1 because $r_N^{**}=\Phi(0,Y^{**})$ is non-decreasing in v_1 and Φ is decreasing in Y. Finally, in the partial precaution region, we have

$$C_v'(Y^{**})=\lambda Q(r_G)-\lambda Q(r^M)+(1-\theta_1)v_1=\theta_1Q(r^M)-\theta_1Q(r_B)-\theta_1v_1,$$

which yields

$$Q(r^M)=\frac{v_1+\lambda Q(r_G)+\theta_1Q(r_B)}{\lambda+\theta_1}$$

and, therefore,

$$C_v'(Y^{**})=\theta_1\left(\frac{v_1+\lambda Q(r_G)+\theta_1Q(r_B)}{\lambda+\theta_1}-Q(r_B)-v_1\right).$$

Hence, Y^{**} is increasing in v_1 in the partial precaution region.

Proof of Proposition 13. The proof for the fact that verification reduces the equilibrium level of precaution is immediate in all cases except when $0<X^{**}<1$ and $r_N^{**}=r^M$. However, in that case, straightforward computations show that $1-X^{**}=\left(\frac{1-\lambda Y^{**}}{1-Y^{**}}\right)(1-X^*)>$

$1 - X^*$.

If $C_v'(Y)$ decreases uniformly, then $Y^{br}(r_N)$ (weakly) increases uniformly and $\Phi\left(X, Y^{br}(r_N)\right)$ (weakly) decreases. Suppose first that $Y^{**} > 0$ and $X^{**} = 0$. Then a marginal reduction in C_v' leads to a decrease in $r_N^{**} = \Phi\left(0, Y^{br}(r_N^{**})\right)$ and an increase in Y^{**}. Suppose now that $Y^{**} > 0$ and $X^{**} > 0$. In this case, a marginal reduction in C_v' leaves $r_N^{**} = r^M$ unchanged and raises $Y^{**} = Y^{br}\left(r^M\right)$. As $\Phi(X^{**}, Y^{**}) = r^M$, the level of precaution X^{**} must decrease.

Proof of Proposition 14. From equation (7) we have

$$1 - X^{**} = \frac{1 - \lambda Y^{**}}{1 - Y^{**}} \frac{r_N - r_1}{(\lambda + \theta_1) r_N - (\lambda + \theta_l) r_1}.$$

The resulting probabilities of events G, B and N are then given, respectively, by

$$\begin{aligned}
p_G^{**} &= \frac{\lambda (r_N - r_1)}{(\lambda + \theta_1) r_N - (\lambda + \theta_l) r_1} + \frac{\lambda (\theta_1 r_N - \theta_l r_1)}{(\lambda + \theta_1) r_N - (\lambda + \theta_l) r_1} Y^{**}, \\
p_B^{**} &= \frac{\theta_1 (r_N - r_1)}{(\lambda + \theta_1) r_N - (\lambda + \theta_l) r_1} (1 - \lambda Y^{**}), \\
p_N^{**} &= \left[1 - (\lambda + \theta_1) \frac{r_N - r_1}{(\lambda + \theta_1) r_N - (\lambda + \theta_l) r_1}\right] (1 - \lambda Y^{**}).
\end{aligned}$$

Suppose first that the posterior belief r_N is not affected by the possibility to verify and screen out bad experiences. Then the availability of a verification technology leads to an increase in the likelihood of a good experience and a decrease in the likelihood of a neutral or bad experience. It follows that the short-term consumer utility increases.

The distribution of the posterior beliefs r undergoes a mean-preserving contraction. Hence, the long-term consumer utility declines reflecting a decrease in the informativeness of the signal after a neutral experience.

Consider now the scenario in which the posterior belief r_N declines as a result of the availability of a verification technology. This occurs only when the resulting level of precaution is zero, thus leading to the following probabilities of experiences: $p_G(0, Y^{**}) = \lambda$, $p_B(0, Y^{**}) = (1 - Y^{**})\theta_1$, and $p_N(0, Y^{**}) = 1 - \lambda - (1 - Y^{**})\theta_1$. As the probability of a bad experience is reduced and the probability of a good experience is either increased or unchanged, the effect of verification on short-term consumer utility is positive.

The change in the distribution of posterior beliefs satisfies a single-crossing property so that verification induces a reduction in risk in the sense of second-order stochastic dominance. This implies that the effect on long-term consumer utility is again negative.

References

[1] Acemoglu, D., A. Malekian, and A. Ozdaglar (2016), "Network Security and Contagion," *Journal of Economic Theory*, 166, 536-585.

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

[3] Admati, A. R., and P. Pfleiderer (1986), "A Monopolistic Market for Information," *Journal of Economic Theory*, 39(2), 400-438.

[4] Admati, A. R., and P. Pfleiderer (1990), "Direct and Indirect Sale of Information," *Econometrica*, 58, 901-928.

[5] Bagwell, K. and A. Wolinsky (2002), "Game Theory and Industrial Organization," *Handbook of Game Theory with Economic Applications*, 3, 1851-1895.

[6] Belleflamme, P., Lam, W. M. W., and W. Vergote (2019), "Competitive Imperfect Price Discrimination and Market Power," CESifo Working Paper No. 7964.

[7] Bergemann, D. and A. Bonatti (2015), "Selling Cookies," *American Economic Journal: Microeconomics*, 7, 259-294.

[8] Bergemann, D., Bonatti, A., and A. Smolin (2018), "The Design and Price of Information," *American Economic Review*, 108, 1-48.

[9] Bergemann, D. and A. Bonatti, (2019), "Markets for information: An introduction," *Annual Review of Economics*, 11.

[10] Bloch, F., and G. Demange (2018), "Taxation and Privacy Protection on Internet Platforms," *Journal of Public Economic Theory*, 20 (1), 52-66 .

[11] Board, S., and M. Meyer-Ter-Vehn (2013), "Reputation for Quality," *Econometrica*, 81(6), 2381-2462.

[12] Bounie, D., Dubus, A., and P. Waelbroeck (2018), "Selling Strategic Information in Digital Competitive Markets," CESifo Working Paper No.7078.

[13] Casadesus-Masanell, R., and A. Hervas-Drane (2015). "Competing with Privacy," *Management Science*, 61, 229-246.

[14] Dwyer, C., and A. Kanguri (2017), "Malvertising-A Rising Threat To The Online Ecosystem." *Journal of Information Systems Applied Research*, 10, 29.

[15] Federal Trade Commission (2009), "Self-Regulatory Principles for Online Behavioral Advertising," FTC Staff Report.

[16] Fudenberg, D. and J. Tirole (1986), "A "Signal-jamming" Theory of Predation," *The RAND Journal of Economics*, 17(3), 366-376.

[17] Goldfarb, A. and C. Tucker (2011), "Online display advertising: Targeting and obtrusiveness," *Marketing Science*, 30(3), 389-404.

[18] Holmström, B. (1999), "Managerial Incentives: A Dynamic Perspective," *The Review of Economic Studies*, 66(1), 169-182.

[19] Judd, K.L. and M.H. Riordan (1994), "Price and Quality in a New Product Monopoly," *The Review of Economic Studies*, 61(4), 773-789.

[20] Lefouili, Y. and Y.L. Toh (2018), "Privacy Regulation and Quality Investment," TSE Working Paper 17-795.

[21] Montes, R., Sand-Zantman, W., and T. Valletti (2019), "The Value of Personal Information in Online Markets with Endogenous Privacy," *Management Science*, 65, 1342-1362.

[22] O'Brien, D.P., and D. Smith (2014), "Privacy in Online Markets: A Welfare Analysis of Demand Rotations," *FTC Bureau of Economics Working Paper*.

[23] Raith, M. (1996), "A General Model of Information Sharing in Oligopoly," *Journal of Economic Theory*, 71, 260-288.

[24] Riordan, M.H. (1985), "Imperfect Information and Dynamic Conjectural Variations," *The RAND Journal of Economics*,16(1), 41-50.

[25] RiskIQ (2016), "RiskIQ's 2016 Malvertising Report,", available at https://www.riskiq.com/infographic/riskiqs-2016-malvertising-report/

[26] Rothschild, M., and J. E. Stiglitz (1971), "Increasing Risk II: Its Economic Consequences," *Journal of Economic Theory*, 3, 66-84.

[27] Spiegel, Y. (2013), "Commercial Software, Adware, and Consumer Privacy," *International Journal of Industrial Organization*, 31, 702-713.

[28] Toh, Y.L. (2018), "Incentivizing Firms to Protect Consumer Data: Can Reputation Play a (Bigger) Role?", working paper.

[29] U.S. Senate Committee on Homeland Security and Governmental Affairs (2014), "Online Advertising and Hidden Hazards to Consumer Privacy and Data Privacy", hearing before the permanent subcommittee on investigations, available at https://www.gpo.gov/fdsys/pkg/CHRG-113shrg89686/pdf/CHRG-113shrg89686.pdf.

[30] Vives X. (1988), "Aggregation of Information in Large Cournot Markets," *Econometrica*, 56:851–76

[31] White, T. B., Zahay, D. L., Thorbjørnsen, H., and S. Shavitt (2008), "Getting too personal: Reactance to highly personalized email solicitations," *Marketing Letters*, 19, 39-50.

Appendix for Online Publication

1 *Ex ante* transparency in the multi-homing scenario

Consider the case of *ex ante* transparency where websites may refuse to sell. In this scenario, all websites observe all levels of precaution before setting prices. Refusal to sell is possible so that the strategy is a public choice of x_i, followed by a private choice of a probability to refuse to sell y_i and a price p_i. The total probability of not selling is then $z_i = x_i + (1 - x_i) y_i$. We focus on symmetric equilibria $x_i = x$.

Suppose first that $P_K\left(\phi\left(x^K\right)\right) > v_1$. In this case the payoff of a website choosing $x_i \neq x$ is

$$L_K(r_N) + z_i z^{K-1} P_K(r_N) + (1 - z_i) p\left(z + (1 - z)(1 - H(p))\right)^{K-1} \text{ with } z_i \in (x_i, 1).$$

Assume that the market anticipates $z^K = X_K^* > x^K$. Then it follows from above that the equilibrium obtains at $z = x + (1 - x) y = (X_K^*)^{\frac{1}{K}}$. Hence, when $X < X_K^*$, the equilibrium probability of sale and the distribution of prices is the same as without transparency.

Suppose now that $P_K\left(\phi\left(x^K\right)\right) < v_1$. Then the symmetric equilibrium distribution of prices is as above. Suppose a website deviates to $x_i \neq x$ while preserving $P_K(r_N) < v_1$. Then all websites set $p \leq v_1$. It can be seen that the lower bound of the support of the price must be the same for all firms (this is because $K - 1$ firms set higher prices than K firms so that there would a contradiction if a smaller number of firms were to charge lower prices). However, one website may have a mass point at v_1 or may not charge high prices.

We consider first the scenario in which the deviating website i puts a mass at v_1. Then, on the support of prices the following indifference condition must hold for website i: $p\left(x + (1 - x)(1 - H(p))\right)^{K-1} = x^{K-1} v_1$, and the following condition must hold for the other websites (where we use the fact that other websites can set a price below but arbitrarily close to v_1):

$$p\left(x + (1 - x)(1 - H(p))\right)^{K-2} \left(x_i + (1 - x_i)(1 - H_i(p))\right) = \left(x_i + (1 - x_i)(1 - H_i(v_1))\right) x^{K-2} v_1.$$

This yields

$$H(p) = \frac{1 - x\left(\frac{v_1}{p}\right)^{\frac{1}{K-1}}}{1 - x} \text{ on an interval } [x^{K-1} v_1, v_1]$$

and

$$(x_i + (1-x_i)(1-H_i(p))) = (x_i + (1-x_i)(1-H_i(v_1)))\left(\frac{v_1}{p}\right)^{\frac{1}{K-1}}.$$

This equilibrium holds provided that the supports are the same, that is, when $x = x_i + (1-x_i)(1-H_i(v_1))$, which holds for $x_i < x$. The payoff of the deviating website with $x_i < x$ is then

$$L_K\left(\phi\left(x_i x^{K-1}\right)\right) + x_i x^{K-1} P_K\left(\phi\left(x_i x^{K-1}\right)\right) + (1-x_i)\, x^{K-1} v_1.$$

Suppose now that $x_i > x$. If $K = 2$ we can apply the previous analysis reverting the role of i and the other website. As the latter sets a mass point at v_1 such that $x + (1-x)(1-H(v_1)) = x_i$, the payoff of website i is then given by

$$L_K(\phi(x_i x)) + x_i x P_K(\phi(x_i x)) + (1-x_i)\, x_i v_1.$$

Assume now that $K > 2$. In this case we investigate an equilibrium where the deviating website does not charge high prices but only prices between $p_{\min}$ and $\hat{p} < v_1$. The following conditions must hold. First, on the interval $(p_{\min}, \hat{p})$ we must have

$$p\,(x + (1-x)(1-H(p)))^{K-1} = \hat{p}\,(x + (1-x)(1-H(\hat{p})))^{K-1},$$

and

$$p\,(x + (1-x)(1-H(p)))^{K-2}(x_i + (1-x_i)(1-H_i(p))) = x_i x^{K-2} v_1.$$

Second, on the interval $(\hat{p}, v_1)$ we must have

$$p\,(x + (1-x)(1-H(p)))^{K-2} x_i = x_i x^{K-2} v_1.$$

Thus, we get:

$$H(p) = \frac{1 - x\left(\frac{v_1}{p}\right)^{\frac{1}{K-2}}}{1-x} \text{ on } [\hat{p}, v_1]$$

Then, we must have $x + (1-x)(1-H(p)) = (\hat{p}/p)^{\frac{1}{K-1}} (v_1/\hat{p})^{\frac{1}{K-2}} x^{K-1}$ on $[p_{\min}, \hat{p}]$, and $x_i + (1-x_i)(1-H_i(p)) = (\hat{p}/p)^{\frac{1}{K-1}} x_i x^{K-2}$ on $[p_{\min}, \hat{p}]$. For the support to be the same we need that $(v_1/\hat{p})^{\frac{1}{K-2}} = x_i/x > 1$ or $x_i > x$, which is the case. To complete the equilibrium we verify that the deviating website setting $p > \hat{p}$ would obtain

$$L_K\left(\phi\left(x_i x^{K-1}\right)\right) + x_i x^{K-1} P_K\left(\phi\left(x_i x^{K-1}\right)\right) + (1-x_i)\, p x^{K-1}\left(\frac{v_1}{p}\right)^{\frac{K-1}{K-2}},$$

which decreases with p and thus is lower than the payoff at $\hat{p}$:

$$L_K\left(\phi\left(x_i x^{K-1}\right)\right) + x_i x^{K-1} P_K\left(\phi\left(x_i x^{K-1}\right)\right) + (1-x_i)\,\hat{p} x_i^{K-1}.$$

To summarize we find that the deviation payoff is given by:

$$\begin{array}{rcl} L_K\left(\phi\left(X_K^*\right)\right) + X_K^* P_K\left(\phi\left(X_K^*\right)\right) + \left(1-\left(X_K^*\right)^{\frac{1}{K}}\right)\left(X_K^*\right)^{\frac{K-1}{K}} v_1 \text{ if } x_i & < & x \text{ and } x^K \leq X_K^*\ , \\ L_K\left(\phi\left(x_i x^{K-1}\right)\right) + x_i x^{K-1} P_K\left(\phi\left(x_i x^{K-1}\right)\right) + (1-x_i)\, x^{K-1} v_1 \text{ if } x_i & < & x \text{ and } x^K > X_K^*, \\ L_K\left(\phi\left(x_i x^{K-1}\right)\right) + x_i x^{K-1} P_K\left(\phi\left(x_i x^{K-1}\right)\right) + (1-x_i)\, v_1 x^{K-2} x_i \text{ if } x_i & > & x \text{ and } x^K > X_K^*. \end{array}$$

The left derivative is negative at $x^K > X_K^*$ and is strictly smaller than the right derivative at that point. The reason for this kink in the payoff is that committing to $x_i > x$ induces a strategic effect that leads the other websites to raise their prices. Hence, the only candidate for a pure precaution equilibrium is at $x^K = X_K^*$ and it exists only if

$$\left(X_K^* P_K'\left(\phi\left(X_K^*\right)\right) + L_K'\left(\phi\left(X_K^*\right)\right)\right)\left(X_K^*\right)^{\frac{1}{K}} \phi'\left(X_K^*\right) + \left(1-\left(X_K^*\right)^{\frac{1}{K}}\right) v_1 \leq 0.$$

and, therefore, only if X_K^* is large enough. Given that X_K^* decreasing in v_1, this holds if v_1 is small enough.

2 Equilibrium analysis under opt out

We focus here on the single-homing case ($K = 1$). Our equilibrium analysis in the scenario in which customers are granted an opt-out option focused on the case where consumers never opt out in the first. period and opt out in the second period if and only if they have a bad experience in the first period. In this section, we still focus on the case when consumers never opt out in the first period but allow for all possible scenarios in the second period.

As a preliminary remark, note that when $\bar{r} < r_B$, the equilibrium is not affected by the possibility of opting out as consumers never choose this option. Similarly, in the case where $\bar{r} > \phi(0)$, consumers always opt out in the second period. Therefore, the website always sells customer information to third parties in the first period, i.e. $X = 0$. We focus in what follows on the scenario in which $r_B < \bar{r} \leq \phi(0)$.

Under this assumption, it cannot be the case that all consumers opt out after a neutral experience because this would imply that $X = 0$ and $r_N = \phi(0) < \bar{r}$. Thus, in equilibrium, consumers should opt in with a positive probability, denoted $P_\emptyset$, after a neutral experience. As we shall see, this probability may be less than 1. We need to distinguish between two

types of equilibria depending on whether consumers opt out or not after a good outcome is observed. In the main text, we considered the scenario in which consumers opt out after a bad experience but not after a good one (i.e. $r_B < \bar{r} < r_1$). We now investigate the existence of an equilibrium in which consumers opt out after a good experience as well. This is the case when $r_1 < \bar{r} < \phi(0)$.

Note first that an equilibrium requires that $r_N \geq \bar{r}$ because some consumers must opt in. This condition is trivially verified if $\bar{r} \leq \phi(1)$, in which case we can replicate the proof of Proposition 1. Defining

$$\begin{aligned} \bar{v}^f &= (\lambda + \theta_1)\left[Q(\phi(1))V_2 - \bar{Q}\bar{V}_2\right] \\ \bar{v}^n &= (\lambda + \theta_1)\left[Q(\phi(1))V_2 - \bar{Q}\bar{V}_2\right], \end{aligned}$$

we get the following result.

Proposition 15 *Consider the single-homing scenario ($K = 1$) and assume $r_1 < \bar{r} \leq \phi(1)$. Then, a unique equilibrium exists. Moreover, there exist thresholds $\bar{v}^f$ and $\bar{v}^n$ such that:*

(i) the website provides full precaution ($X^{opt} = 1$) if $v_1 \leq \bar{v}^f$;

(ii) the website provides partial precaution ($0 < X^{opt} < 1$) if $\bar{v}^f < v_1 < \bar{v}^n$;

(iii) the website provides no precaution ($X^{opt} = 0$) if $v_1 \geq \bar{v}^n$.

Proof. Since $r_\emptyset \geq \phi(1) \geq \bar{r}$, it is sufficient to replace $Q(r_G)V_2$ and $Q(r_B)V_2$ by $\bar{Q}\bar{V}_2$ in the proof of Proposition 1 to get this result. ■

The website would not offer the opt-out option on a voluntary basis if $[\lambda Q(r_G) + \theta_1 Q(r_B)]V_2 > (\lambda + \theta_1)\bar{Q}\bar{V}_2$. In this case, notice that $\bar{v}^f$ and $\bar{v}^n$ are greater than $\psi^f V_2$ and $\psi^n V_2$, respectively, and that $X^{opt} > X^*$ in the range of partial precaution. Thus, the effect of a mandatory opt-out policy is to raise (weakly) the level of precaution.

Let us now turn to the scenario in which $\phi(1) < \bar{r} < \phi(0)$. The analysis in the case $r_1 < \bar{r} < \phi(0)$ no longer applies because the level of precaution cannot be too large in equilibrium. Let us define $\bar{X}$ as the unique solution to $\phi\left(\bar{X}\right) = \bar{r}$. Then, in any equilibrium we must have $X^{opt} \leq \bar{X}$. Notice that the equilibrium level of precaution X^{opt} decreases in v_1. Since for sufficiently large values of v_1 we have $X^{opt} = 0$, there must exist some critical level $\bar{v}^o$ such that $X^{opt} < \bar{X}$ if and only if $v_1 > \bar{v}^o$. In this range, the equilibrium is similar to the one when $r_1 < \bar{r} \leq \phi(1)$. However, for lower values of v_1, the equilibrium must be such that $X^{opt} = \bar{X}$ and consumers randomize between opting in and opting out.

Proposition 16 *Consider the single-homing scenario ($K = 1$) and assume $\phi(1) < \bar{r} < \phi(0)$. Then, a unique equilibrium exists. Moreover, there exists a threshold $\bar{v}^o$ such that:*

(i) the website provides partial precaution and $X^{opt} = \bar{X}$ *if* $v_1 \leq \bar{v}^o$*;*
(ii) the website provides partial precaution and $X^{opt} \in \left(0, \bar{X}\right)$ *if* $\bar{v}^o < v_1 < \bar{v}^n$*;*
(iii) the website provides no precaution ($X^{opt} = 0$*) if* $v_1 \geq \bar{v}^n$*.*

Proof. The result is the same as before when $v_1 > \bar{v}^o$ where $\bar{v}^o$ is defined by

$$\bar{v}^o = (\lambda + \theta_1) \left[Q\left(\phi\left(\bar{X}\right)\right) V_2 - \bar{Q}\bar{V}_2 \right]$$

For smaller values of v_1, we have $X = \bar{X}$ and $r_N = \bar{r}$ in equilibrium, and

$$P_\emptyset = \frac{v_1}{\bar{v}^o} < 1.$$

It then suffices to replace $Q(r_G) V_2$ and $Q(r_B) V_2$ by $\bar{Q}\bar{V}_2$ in the proof of Proposition 1. ■