Revealing Transactions Data to Third Parties: Implications of Privacy Regimes for Welfare in Online Markets

by

Michael R. Baye* and David E. M. Sappington**

Abstract

We examine the effects of privacy policies regarding transactions (e.g., price/quantity) data on online shopping platforms. Disclosure of transactions data induces consumer signaling behavior that affects merchant pricing decisions and the welfare of platform participants. A profit-maximizing platform prefers the disclosure policy that maximizes total welfare. Although this policy benefits sophisticated consumers, it harms unsophisticated (myopic) consumers. Consequently, the welfare effects of alternative privacy policies, data breaches, deceptive privacy policies, and opt-in/opt-out requirements can differ sharply, depending on the level of consumer sophistication and on other factors such as the prevailing status quo.

JEL Numbers: D04, D18, D4, D6, D8, L00, L5.

Keywords: Platforms, Privacy, Signaling, Consumer Protection, Information Economics.

October 2019

* Kelley School of Business, Indiana University, 1309 East Tenth Street, Bloomington, IN 47401 (mbaye@indiana.edu).

** Department of Economics, P.O. Box 117140, University of Florida, Gainesville, Florida 32611-7140 (sapping@ufl.edu).

We thank Paul Belleflamme, Rick Harbaugh, John Morgan, Jeff Prince, Dan Sacks, Curtis Taylor, Liad Wagman, anonymous referees, the editor, and the co-editor for helpful observations and suggestions. We also benefited from comments by participants at the Federal Trade Commission (FTC)'s *Hearings on Competition and Consumer Protection in the 21st Century* in November, 2018 and participants at the FTC's *Consumer Protection Economics Symposium* in December, 2018. We are indebted to Roger Blair for suggesting a related line of research and assisting with early model formulation.

1 Introduction

Consumer protection agencies in the United States and the European Union continue to scrutinize the use of "big data" by online platforms such as Amazon, Apple, Facebook, and Google. In the U.S., dozens of regulators are investigating the type of data that platforms collect, how the data are employed, and whether consumers are well served when a platform's ability to use transactions data or share them with third parties is restricted.[1] The goal of this paper is to provide a better understanding of the transactions data privacy policies that consumers and platforms prefer and the welfare effects of related regulatory restrictions.

The economic literature has established that restricting the use or limiting the sharing of transactions data can either benefit or harm consumers.[2] Restricting a platform's ability to track the purchases of individual consumers can harm consumers by limiting the platform's ability to efficiently match consumers with products and/or advertisers (Evans, 2009).[3] However, the same privacy policy can benefit consumers by preventing a monopolist from exploiting in one transaction information it learns about a customer in a different transaction (Acquisti and Varian, 2005). The literature also notes that the welfare effects of privacy policies depend on whether consumers are sophisticated, i.e., whether they fully anticipate how their present purchase decisions may affect the prices they face in subsequent transactions (Taylor, 2004).[4]

We extend this literature in four primary ways. First, we identify conditions under which a platform's preferred privacy policy maximizes total welfare. Second, we examine how mandated alternatives to the platform's preferred policy affect the welfare of sophisticated and

[1] Platform use of transactions data is an important issue in the Federal Trade Commission's extensive *Hearings on Competition and Consumer Protection in the 21st Century* and in the European Union's ongoing investigation of Amazon's business practices. In addition, fifty attorneys general from U.S. states and territories recently announced investigations of "Big Tech" (Grimaldi and Kendall, 2019). Spulber (2019) provides a useful analysis of price setting on "platforms" and an informative discussion of the relation between platforms and entities such as market makers and intermediaries that operate in multi-sided markets and networks.

[2] Taylor and Wagman (2014) employ four common models of oligopoly competition to demonstrate that winners and losers from privacy policies can vary with the prevailing form of market competition. Acquisti, Taylor and Wagman (2016) provide an excellent survey of the literature on privacy.

[3] Campbell, Goldfarb, and Tucker (2015) demonstrate that policies that require firms to protect consumer data (e.g., prevent third parties from accessing it) can harm consumers by placing smaller firms at a competitive disadvantage, thereby affecting market structure adversely.

[4] Tucker (2012) provides an excellent survey of the empirical literature on these and other tradeoffs.

unsophisticated consumers. Third, we assess the welfare effects of "opt-in" mandates (that require consumers to give their explicit consent before platforms can share transactions data with third parties) and "opt-out" mandates (that require platforms to allow consumers to request and thereby receive a personal exemption from default sharing of transactions data). Fourth, we analyze the welfare effects of data breaches (caused by hackers, for example), deceptive privacy policies, and requirements to remove personally identifiable information before data is shared with third parties.

In our model, consumers purchase two distinct (non-competing) products from different merchants on an online platform. When transactions data are shared on the platform, a consumer's interaction with one merchant may reveal to other merchants the consumer's reservation value for their products. The other merchants may modify the prices they charge the consumer accordingly. A sophisticated consumer who recognizes this effect of data sharing takes it into account when interacting with all merchants. In contrast, when he decides whether to purchase a merchant's product, an unsophisticated consumer only considers whether the price the merchant sets exceeds his reservation value for the product.

We find that sophisticated consumers and the platform generally benefit when the platform shares all transactions data with third parties (i.e., other merchants on the platform). The data sharing provides a channel through which sophisticated consumers can credibly signal when their reservation values for the merchants' products are low. Such signaling induces price concessions from merchants.[5] When the platform does not share transactions data, it effectively closes the signaling channel, thereby harming sophisticated consumers. Closing the channel also reduces platform profit and total welfare by limiting the consummation of welfare-enhancing transactions.

In contrast, unsophisticated consumers benefit when the platform never shares transactions data with third parties. This privacy policy prevents merchants from exploiting

[5] Belleflamme and Vergote (2016) document in a distinct setting the consumer welfare gains that can arise when a merchant is better able to discern consumers' reservation values. The authors consider a setting where consumers interact once with a single monopolist. If she is not prevented from doing so, the monopolist can choose whether to discover consumers' reservation values for her product (with a specified probability). At personal cost, a consumer can eliminate the monopolist's ability to discern the consumer's personal reservation value. The authors show that consumers may be better off when they are unable to limit the monopolist's ability to discern reservation values.

unsophisticated consumers by charging them higher prices after they are observed to pay high prices to other merchants. Thus, the privacy policy that best serves unsophisticated consumers harms sophisticated consumers. Consequently, the formulation of privacy regulations for online platforms can be challenging even when the sole objective of the regulations is to maximize consumer welfare.

The varying impacts of a platform's privacy policy on the welfare of sophisticated and unsophisticated consumers might lead one to conclude that consumer welfare would unambiguously rise under regulations giving each consumer property rights over his data, such that the platform can only share the data with third parties if the consumer opts in. We show that these (and related opt-out) policies that allow each consumer to select his or her optimal privacy policy are not a panacea. The impact of opt-in and opt-out mandates also varies with the degree of consumer sophistication and with the magnitude of the costs that consumers must incur to opt in to or opt out of the platform's prevailing privacy policy.[6] In our model, requiring explicit consumer consent before transactions data are shared with third parties can harm sophisticated consumers.

We also examine how violations of a platform's stated privacy policy (through data breaches by hackers or deception by the platform) affect the welfare of the platform's customers.[7] We find, for example, that when the platform announces it will implement the privacy policy that maximizes the welfare of unsophisticated consumers, both sophisticated and unsophisticated consumers are harmed by an unanticipated violation of the platform's announced privacy policy. However, some consumers are not harmed, and harm only arises under certain configurations of merchant costs.[8] In contrast, when the platform adopts the

[6] Athey, Catalini, and Tucker (2017) demonstrate that even small costs of opting in or opting out can greatly affect consumers' privacy choices.

[7] Such violations might stem from cyber attacks by outside parties or from willful actions by platforms. The Federal Trade Commission (FTC)'s complaint against Uber alleges a breach of Uber's computer network by an outside entity that revealed customer data to third parties (*Complaint in the Matter of Uber Technologies, Inc.*, Docket No. C-4662, October 26, 2018, https://www.ftc.gov/enforcement/cases-proceedings/152-3054/uber-technologies-inc). The FTC's investigation of Pay Pal entails allegations that the company reneged on its promise to keep customer data private (*Decision and Order* in the Matter of PayPal, Inc., Docket No. C-4651, May 24, 2018, https://www.ftc.gov/enforcement/cases-proceedings/162-3102/paypal-inc-matter).

[8] Specifically, consumers with low reservation values for the merchants' products are not harmed. A consumer with a high reservation value is harmed when the data from his transaction with a high-cost merchant is shared with a low-cost merchant (who would set a low price for her product in the absence of data sharing).

policy that maximizes the welfare of sophisticated consumers, a data breach does not harm consumers (because their transactions data are already known to all merchants on the platform). However, a violation of this privacy policy can harm sophisticated consumers by foreclosing the signaling channel through which they can secure price concessions. Therefore, the effects of data breaches and violations of privacy policies can differ for sophisticated and unsophisticated consumers. In addition, the effects of data breaches can differ from the effects of deceptive practices or violations of the platform's stated privacy policies.

We find that total welfare, platform profit, and the welfare of sophisticated consumers are maximized when the platform provides transactions data to third parties. Consequently, under a *laissez faire* policy that permits the platform to implement its preferred privacy policy, the platform will adopt the privacy policy that maximizes the welfare of sophisticated consumers. This privacy policy is not ideal for unsophisticated consumers, however. It is also not the best policy for all merchants.

We also examine the impact of removing all information about a consumer's identity before transactions data are shared with third parties. The removal of such personal information can benefit unsophisticated consumers, but does not always do so. The removal generally harms sophisticated consumers by effectively closing the channel through which they might signal their low reservations values for the merchants' products.

Our analysis differs from the seminal work of Taylor (2004) and Acquisti and Varian (2005) by analyzing platform incentives, opt-in and opt-out mandates, and requirements to remove all information about a consumer's identity before transactions data are shared with third parties.[9] Taylor (2004) focuses on the impact of limits on the ability of individual merchants (rather than the platform) to sell customer transactions data to other merchants.[10]

[9] Our analysis also differs in this respect with the important related work of Conitzer, Taylor, and Wagman (2012) (CTW). CTW analyze a setting where consumers interact with a monopolist in each of two periods. If the monopolist can track individual consumers, she can charge a higher price in period 2 to consumers who purchased her product in period 1. When consumers can preclude such tracking at low personal cost, they will do so to avoid exploitation in period 2. CTW demonstrate that the monopolist also can gain when she is unable to track consumers. This inability allows the monopolist to commit not to exploit consumers in period 2, which makes them willing to pay more for the monopolist's product in period 1, thereby increasing her two-period expected profit.

[10] Calzolari and Pavan (2006) analyze a related setting in which an agent (A) interacts sequentially with two principals (P1 and P2). A is privately informed about a personal characteristic that affects the value he derives from his interaction with the principals. The authors examine the policy (including the information disclosure policy) that maximizes P1's expected welfare in a setting where payments from P2 to P1 can reflect

We find that the incentive of the platform to disclose transactions data to its merchants can differ significantly from the incentive of an individual merchant to disclose its transactions data to other merchants. In our model, a merchant whose data are released to other merchants often suffers a reduction in profit (which it cannot recoup by charging a fee for the data in our model). In contrast, the platform maximizes its profit by disclosing all transactions data to its merchants.[11]

Our analysis proceeds as follows. Section 2 describes the key elements of our model. Section 3 characterizes equilibrium outcomes under different privacy policies. Section 4 identifies the distinct privacy policies that maximize the welfare of sophisticated and unsophisticated consumers. Section 4 also explains how data breaches, violations of stated privacy policies, and opt-in or opt-out policies affect consumer welfare. Section 5 examines the impact of privacy policies on total (rather than consumer) welfare, identifies the privacy policy that maximizes platform profit, and explores the effects of requirements to remove all information about a consumer's identity before transactions data are shared with third parties. Section 6 discusses extensions and concludes.

2 Elements of the Model

We analyze a parsimonious model that allows us illustrate most simply the impacts of privacy policies in settings where a consumer's valuations of multiple products are correlated. We consider a setting where $M \geq 2$ distinct merchants can sell differentiated products to $N \geq 1$ potentially heterogenous consumers via an online shopping platform. The platform enables consumers to identify sellers of the idiosyncratic products they seek to purchase (consistent with the empirical evidence in Brynjolfsson et al. (2003) and Ellison and Ellison

the policy that P1 adopts. The authors identify conditions under which P1 does not disclose any relevant information to P2. The authors also demonstrate that P1's optimal policy can entail some information disclosure and that such disclosure can secure Pareto gains.

[11] Our model also differs from Taylor (2004)'s model by allowing merchants to have distinct production costs. Consequently, the welfare effects of privacy policies in our model vary with the configuration of firms' costs. Like Taylor (2004), Kim and Wagman (2015) (KW) analyze a setting in which merchants may be permitted to sell information about their customers. In KW's model, consumers interact sequentially with two merchants, M1 and M2. Through costly effort, M1 can acquire more accurate information about the cost of serving individual consumers. KW identify conditions under which consumers are better off (and total welfare increases) when M1 is permitted to sell information about its customers to M2. The welfare gains arise in part because the potential to profit from the sale of information induces M1 to acquire better information about its customers, which helps to screen out consumers who are unduly costly to serve.

(2018), for example). For simplicity, we assume each consumer seeks to purchase at most two products, each of which is sold by a different merchant. Furthermore, a consumer's demand for each relevant product is rectangular, so the consumer purchases $n_i > 0$ units (0 units) of Merchant i's product if the relevant price of the product does not exceed (exceeds) the consumer's reservation value for the product.[12] A given consumer's reservation value is the same for each of the two products he may purchase.[13]

For simplicity, we assume each consumer's reservation value can take on one of two possible values. These values can vary across consumers. Unless otherwise noted, the reservation values of different consumers are assumed to be independent. For expositional ease, it will be convenient to consider the activities of a generic consumer whose reservation value (r) for each unit of the products he may purchase is either low ($\underline{r}$) or high ($\overline{r}$) (with $\overline{r} > \underline{r}$). Each consumer knows his reservation value from the outset of his interaction with the merchants. The merchants cannot directly observe any consumer's reservation value, but they initially believe that $r = \overline{r}$ with probability $\phi \in (0, 1)$ for the generic consumer.

Merchants can engage in consumer-specific price discrimination if they find it profitable to do so. For example, a merchant might charge a higher price for her product to a consumer who was observed to have paid a relatively high price for the product he purchased from a different merchant. Merchant i produces her product at constant average cost c_i, which is strictly less than the low reservation value of each consumer. Consequently, in principle, gains from trade are possible in every consumer-merchant interaction.

This parsimonious model may constitute a reasonable caricature of settings like the following. Suppose an academic economist is interested in purchasing two specific books (e.g.,

[12] Section 5 discusses an extension where consumer demand varies continuously with price. For the case of rectangular demand, we assume that when a consumer is indifferent between purchasing n_i units and purchasing 0 units, he purchases n_i units. This assumption is without essential loss of generality because, as in the search literature (e.g., Baye and Morgan, 2001), a minor model extension ensures that a consumer with reservation value $r_i > 0$ strictly prefers to purchase n_i units than to purchase 0 units at the monopoly price, r_i. Specifically, suppose demand is constant at $n_i > 0$ for any price below r_i, and declines continuously to zero as the price increases above r_i to the choke price, $r_i + \varepsilon_i$. If $\varepsilon_i > 0$ is sufficiently small, the profit-maximizing monopoly price is r_i and the consumer secures strictly higher surplus by purchasing n_i units at this price than he secures by purchasing 0 units.

[13] This formulation abstracts from explicit strategic interactions between the merchants (as in Baye and Morgan (2001)). Instead, each merchant has market power in the spirit of monopolistic competition (e.g., Reinganum, 1979).

Order Statistics and *Auction Theory*). Further suppose some economists (e.g., professors) typically are able to pay more for their reference books than others (e.g., graduate students), but a book merchant cannot observe directly how much any specific consumer is willing to pay for the book she sells. In such a setting, the merchant selling *Order Statistics* may be able to infer something about how much the consumer is willing to pay for this book by observing whether the consumer purchased *Auction Theory* and the price he paid for that book. The consumer's rectangular demand is natural in this setting because n_i can be viewed as the number of pages in the book sold by Merchant i. For a given price per page, the consumer either buys all of the pages in the book or none of the pages.

Although consumers, merchants, and products can all be heterogeneous in our model, it is convenient to suppress indicators for specific consumers, merchants, and products. Instead, we use the indices 1 and 2 to denote the *order* in which a generic consumer makes his purchase decisions for the two products of potential interest. Specifically, we let p_1 denote the price a generic consumer is charged by the first merchant he visits (henceforth, Merchant 1), whereas p_2 denotes the price the consumer is charged by the second merchant he visits (henceforth, Merchant 2).

The price each merchant sets for the product she sells to a particular consumer depends in part on her beliefs about the consumer's reservation value. These beliefs, in turn, may be influenced by the platform's privacy policy. We initially focus on two such policies: one where the platform reveals all transactions data to other merchants on the platform ("third parties"), and one where the platform reveals no such data. Then, in Section 5, we allow for the possibility that merchants might learn some, but not all, details of other consumer-merchant interactions. In the setting of primary interest, if the platform reveals all transactions data to third parties, then when Merchant 2 sets the price at which she will sell her product to a particular consumer, she knows the price that Merchant 1 set for the consumer and whether the consumer purchased n_1 units or 0 units of Merchant 1's product at this price. In contrast, if the platform reveals no transactions data to third parties, then at the time Merchant 2 sets her price for a customer, she does not know the price that Merchant 1 set for the consumer or whether the consumer purchased n_1 units of Merchant 1's product or declined to purchase Merchant 1's product.

Under *privacy*, i.e., when the platform reveals no transactions data to third parties, each consumer and Merchant 1 know that Merchant 2 will learn nothing about their interaction. Consequently, Merchant 1's deliberations are straightforward. The merchant knows that if she charges a generic consumer (with $r \in \{\underline{r}, \overline{r}\}$) price $p_1 = \underline{r}$ for her product, the consumer will purchase n_1 units of the product. Merchant 1 secures payoff $n_1 [\, \underline{r} - c_1 \,]$ from this transaction. Alternatively, if Merchant 1 charges a generic consumer price $p_1 = \overline{r}$, the consumer will buy n_1 units of the merchant's product if and only if his reservation value is $r = \overline{r}$. Merchant 1's corresponding (expected) payoff from this transaction is $\phi\, n_1 [\, \overline{r} - c_1 \,]$.[14] Therefore, Merchant 1 will set the lower price ($p_1 = \underline{r}$) and always sell to the generic consumer if and only if her unit cost of production is sufficiently low and the consumer's reservation value is sufficiently likely to be low, i.e.

$$n_1 [\, \underline{r} - c_1 \,] \;\geq\; \phi\, n_1 [\, \overline{r} - c_1 \,] \;\;\Leftrightarrow\;\; c_1 \;\leq\; \widehat{c} \;\equiv\; \overline{r} - \frac{\overline{r} - \underline{r}}{1 - \phi}\,.$$

Merchant 2's considerations are identical to those of Merchant 1 under privacy. Therefore, $\widehat{c}$ is the unit cost for which a merchant's payoff is the same whether she sets price $\underline{r}$ or $\overline{r}$ for the generic consumer under privacy.

Additional considerations arise in the absence of privacy, i.e., when all transactions data are revealed to third parties. In this event, Merchant 2 may infer something about a consumer's reservation value (r) from the details of the consumer's prior interaction with Merchant 1. For example, if Merchant 2 learns a generic consumer bought Merchant 1's product at price $p_1 = \overline{r}$, Merchant 2 might infer the consumer's reservation value is $r = \overline{r}$. Because he recognizes that Merchant 2 might draw such an inference, the consumer may decline to buy Merchant 1's product at price $p_1 = \overline{r}$ even when $r = \overline{r}$. Merchant 1 is aware that each consumer's action may be affected by his concerns about what Merchant 2 will infer from the action. This awareness, in turn, may influence the price that Merchant 1 charges a generic consumer for her product.

We consider pure-strategy perfect Bayesian equilibria (PPBE) when transactions data are revealed to third parties. In such equilibria: (i) each party's action is rational, given

[14] For expositional ease, we will employ the term "payoff" to denote "expected payoff" in the ensuing discussion. If Merchant 1 sets $p_1 \in (\underline{r}, \overline{r})$, her per-transaction payoff is $\phi\, n_1 [\, p_1 - c_1 \,] < \phi\, n_1 [\, \overline{r} - c_1 \,]$. If Merchant 1 sets $p_1 > \overline{r}$, her per-transaction payoff is $0 \,<\, \min\{\, n_1 [\, \underline{r} - c_1 \,],\, \phi\, n_1 [\, \overline{r} - c_1 \,]\,\}$. The term "price" is employed to denote "unit price" throughout the analysis.

prevailing beliefs about r; and (ii) Merchant 2's beliefs about r reflect Bayes rule for all equilibrium actions. In a separating (pooling) PPBE, the number of units the consumer buys from Merchant 1 varies (does not vary) with r.

The timing in the model is as follows. First, the platform announces whether transactions data will be revealed to third parties. Then each consumer uses the platform to identify and transact with the relevant merchants – first with Merchant 1 and then with Merchant 2. Next, Merchant 1 sets price p_1 for the generic consumer and the consumer purchases either 0 or n_1 units of Merchant 1's product. Merchant 2 learns this information when the platform reveals transactions data to third parties. Otherwise, Merchant 2 learns nothing about the consumer's interaction with Merchant 1. Finally, Merchant 2 sets price p_2 for the generic consumer and the consumer purchases either 0 or n_2 units of Merchant 2's product.

3 Equilibrium Transaction Outcomes

Outcomes of individual consumer-merchant interactions can vary with the consumer's reservation value, his sophistication, and merchants' costs. To focus on the effects of incomplete information about consumers' reservation values, we assume all parties know from the outset of their interaction firms' costs and whether any particular consumer is sophisticated or unsophisticated. (Section 6 considers the extension where merchants have incomplete information about a particular consumer's sophistication.)

As noted in the Introduction, an unsophisticated consumer considers only the interaction in which he is presently engaged when he decides whether to buy a merchant's product. Specifically, even when all transactions data are revealed to third parties, the consumer will purchase a merchant's product if and only if the merchant's price does not exceed his reservation value (r). In contrast, a sophisticated consumer considers all future interactions with other merchants each time he interacts with a merchant. In the absence of privacy, the sophisticated consumer recognizes that his decision to purchase or not purchase Merchant 1's product may affect Merchant 2's belief about r and therefore may affect the price that Merchant 2 charges the consumer for her product.

3.1 Unsophisticated Consumers

We first characterize equilibrium transaction outcomes for a generic unsophisticated consumer. Observe that under privacy, the consumer's behavior is the same whether he is sophisticated or unsophisticated. In particular, the consumer will purchase n_1 units from Merchant 1 if and only if the merchant charges him a price that does not exceed r. Merchant 1 will charge an unsophisticated consumer price $p_1 = \underline{r}$ (and thereby ensure he always buys her product) when c_1 and ϕ are sufficiently small. In contrast, Merchant 1 will charge the unsophisticated consumer price $p_1 = \overline{r}$ (and consequently sell him the product at this higher price only if $r = \overline{r}$) when c_1 and ϕ are sufficiently large.

These conclusions are summarized in Lemma 1. Baye and Sappington (2019) provides a formal proof of Lemma 1 and all subsequent lemmas.

Lemma 1 *Merchant* 1*'s payoff from an interaction with an unsophisticated consumer is* $\phi\, n_1 [\, \overline{r} - c_1 \,]$ $(n_1 [\, \underline{r} - c_1 \,])$ *when* $c_1 > \widehat{c}$ $(c_1 \leq \widehat{c})$, *both under privacy and in the absence of privacy.*

Under privacy, Merchant 2 receives no information about a consumer's interaction with Merchant 1. Consequently, Merchant 2 acts exactly as Merchant 1 does. The consumer never has an opportunity to purchase at a price below $\underline{r}$. Consequently, his welfare (i.e., the difference between the value he derives from the products he purchases and the amount he pays for the products) is 0 when $r = \underline{r}$. When $r = \overline{r}$, the consumer secures strictly positive welfare when and only when ϕ is relatively small and a merchant has a relatively low cost and so charges the consumer $\underline{r}$ for her product. In summary:

Lemma 2 *Under privacy, Merchant* 2*'s payoff from an interaction with an unsophisticated consumer is* $\phi\, n_2 [\, \overline{r} - c_2 \,]$ $(n_2 [\, \underline{r} - c_2 \,])$ *when* $c_2 > \widehat{c}$ $(c_2 \leq \widehat{c})$. *A consumer's overall welfare is* 0 *when* $r = \underline{r}$. *His overall welfare when* $r = \overline{r}$ *is: (i)* $[\, n_1 + n_2 \,] [\, \overline{r} - \underline{r} \,]$ *if* $c_1 \leq \widehat{c}$ *and* $c_2 \leq \widehat{c}$; *(ii)* $n_1 [\, \overline{r} - \underline{r} \,]$ *if* $c_1 \leq \widehat{c}$ *and* $c_2 > \widehat{c}$; *(iii)* $n_2 [\, \overline{r} - \underline{r} \,]$ *if* $c_1 > \widehat{c}$ *and* $c_2 \leq \widehat{c}$; *and (iv)* 0 *if* $c_1 > \widehat{c}$ *and* $c_2 > \widehat{c}$.

In the absence of privacy, if Merchant 1 charges an unsophisticated consumer price $p_1 = \overline{r}$ for her product (because $c_1 > \widehat{c}$), Merchant 2 learns that $r = \overline{r}$ $(r = \underline{r})$ if the consumer buys (does not buy) the product from Merchant 1. Merchant 2 charges the consumer a price equal to the revealed value of r in this case. In contrast, if Merchant 1 charges the unsophisticated consumer price $p_1 = \underline{r}$ (because $c_1 \leq \widehat{c}$), the consumer always buys her product. Consequently, Merchant 2 learns nothing about r from observing the details of the consumer's interactions with Merchant 1, and so acts as she does under privacy. These observations provide:

Lemma 3 *In the absence of privacy, Merchant 2's payoff from an interaction with an unsophisticated consumer is: (i) $n_2 [\, \underline{r} - c_2 \,]$ if $c_1 > \widehat{c}$ and $r = \underline{r}$; (ii) $n_2 [\, \overline{r} - c_2 \,]$ if $c_1 > \widehat{c}$ and $r = \overline{r}$; (iii) $\phi \, n_2 [\, \overline{r} - c_2 \,]$ if $c_1 \leq \widehat{c}$ and $c_2 > \widehat{c}$; (iv) $n_2 [\, \underline{r} - c_2 \,]$) if $c_1 \leq \widehat{c}$ and $c_2 \leq \widehat{c}$. A consumer's overall welfare is 0 when $r = \underline{r}$ or $c_1 > \widehat{c}$. His overall welfare when $r = \overline{r}$ and $c_1 \leq \widehat{c}$ is $n_1 [\, \overline{r} - \underline{r} \,]$ ($[\, n_1 + n_2 \,] [\, \overline{r} - \underline{r} \,]$) when $c_2 > \widehat{c}$ ($c_2 \leq \widehat{c}$).*

3.2 Sophisticated Consumers

Now consider equilibrium transaction outcomes for a generic sophisticated consumer. Under privacy, a sophisticated consumer knows that the details of his initial interaction with Merchant 1 will not be revealed to Merchant 2. Therefore, the consumer's concern in each interaction with a merchant is solely with the details of that interaction. Consequently, a sophisticated consumer acts exactly as an unsophisticated consumer acts, and the corresponding outcomes are precisely those that arise under privacy when the consumer is unsophisticated.

Lemma 4 *When a consumer is sophisticated, the PPBE transaction outcomes under privacy are as specified in Lemmas 1 and 2 (where the consumer is unsophisticated).*

Important differences arise in the absence of privacy when a consumer is sophisticated. In this setting, when $r = \overline{r}$, the consumer may decline to purchase Merchant 1's product at price $p_1 = \overline{r}$ to avoid having Merchant 2 raise her price to $\overline{r}$ after inferring that $r = \overline{r}$. In such a setting, Merchant 1 has to reduce the price she charges the consumer below $\overline{r}$ (her preferred price when $c_1 > \widehat{c}$) to convince the consumer with reservation value $\overline{r}$ to purchase her product even though doing so causes Merchant 2 to infer that $r = \overline{r}$. Let $\widehat{p}_1$ denote the highest price the consumer with reservation value $\overline{r}$ will pay for Merchant 1's product when doing so causes Merchant 2 to infer that $r = \overline{r}$ and so increase the price she charges the consumer from $\underline{r}$ to $\overline{r}$. This price is determined by:

$$n_1 [\, \overline{r} - \widehat{p}_1 \,] \; = \; n_2 [\, \overline{r} - \underline{r} \,] \quad \Leftrightarrow \quad \widehat{p}_1 \; = \; \frac{n_2}{n_1} \, \underline{r} + [\, 1 - \frac{n_2}{n_1} \,] \, \overline{r} \, . \tag{1}$$

Expression (1) implies that $\widehat{p}_1 \; \geq \; \underline{r}$ if and only if $n_1 > n_2$. Therefore, Merchant 1 will charge the consumer price $\widehat{p}_1$ (and sell her product to the consumer only when $r = \overline{r}$) rather than charge price $\underline{r}$ (and always sell her product to the consumer) if $n_1 > n_2$ and

$$\phi \, n_1 [\, \widehat{p}_1 - c_1 \,] \; \geq \; n_1 [\, \underline{r} - c_1 \,] \;\; \Leftrightarrow \; \phi \; \geq \; \frac{\underline{r} - c_1}{\widehat{p}_1 - c_1} \;\; \Leftrightarrow \; c_1 \; > \; c^* \equiv \widehat{c} + \phi \, \frac{n_2}{n_1} \left[\frac{\overline{r} \; - \; \underline{r}}{1 \; - \; \phi} \right] . \tag{2}$$

The price reduction that Merchant 1 is compelled to deliver to a sophisticated consumer in the absence of privacy reduces her per-transaction payoff by $\phi \, n_1 [\, \overline{r} - \widehat{p}_1 \,]$. Merchant 2's corresponding payoff increases by the same amount, $\phi \, n_2 [\, \overline{r} - \underline{r} \,]$, because she sells her product to the consumer at price $p_2 = \overline{r}$ rather than at price $p_2 = \underline{r}$ when $r = \overline{r}$. The consumer's welfare is the same under privacy and in its absence because the welfare gain the consumer secures from the lower price he pays to Merchant 1 in the absence of privacy ($n_1 [\, \overline{r} - \widehat{p}_1 \,]$) is offset by the welfare reduction he incurs from the higher price he pays to Merchant 2 ($n_2 [\, \overline{r} - \underline{r} \,]$).

Lemma 5 *In any separating PPBE that arises in the absence of privacy when a consumer is sophisticated: (i) Merchant* 1*'s per-transaction payoff is* $\phi \, n_1 [\, \widehat{p}_1 - c_1 \,]$*; (ii) Merchant* 2*'s per-transaction payoff is* $n_2 [\, \phi \, (\overline{r} - c_2) + (1 - \phi) \, (\underline{r} - c_2) \,]$*; and (iii) a consumer's welfare is* 0 *when* $r = \underline{r}$ *and* $n_1 [\, \overline{r} - \widehat{p}_1 \,] \; = \; n_2 [\, \overline{r} - \underline{r} \,]$ *when* $r = \overline{r}$*. Such equilibria arise when* $n_1 > n_2$ *,* $c_1 > c^*$*, and* $c_2 \leq \widehat{c}$*.*

When $n_1 > n_2$ and Merchant 1's cost and ϕ are sufficiently low (so $c_1 < c^*$), the merchant prefers to reduce the price she charges a sophisticated consumer to $\underline{r}$ and always sell her product to the consumer than to charge the consumer $p_1 = \widehat{p}_1$ and sell to him only when $r = \overline{r}$. When the consumer always purchases Merchant 1's product, Merchant 2 infers nothing about the consumer's reservation value from his interaction with Merchant 1. Therefore, Merchant 2 sets price $p_2 = \underline{r}$ for the consumer when $c_2 \leq \widehat{c}$. A similar outcome arises when $n_1 \leq n_2$. In this case, a consumer sacrifices relatively little when $r = \overline{r}$ if he declines to purchase Merchant 1's product when she sets a price $p_1 \in (\underline{r}\,, \overline{r}\,]$ for the product. Consequently, Merchant 2 will not conclude that $r = \underline{r}$ when she learns the consumer did not purchase Merchant 1's product at such a price. Therefore, the only PPBE that arises when $n_1 \leq n_2$ and $c_2 \leq \widehat{c}$ is the pooling equilibrium in which both merchants charge the consumer price $\underline{r}$ for their products. To summarize:

Lemma 6 *In any pooling PPBE that arises in the absence of privacy when a consumer is sophisticated, Merchant* 1*'s per-transaction payoff is* $n_1\,[\,\underline{r} - c_1\,]$*, Merchant* 2*'s per-transaction payoff is* $n_2\,[\,\underline{r} - c_2\,]$*, and the consumer's overall welfare is* 0 *(*$[\,n_1 + n_2\,]\,[\,\overline{r} - \underline{r}\,]$*) when* $r = \underline{r}$ *(*$r = \overline{r}$*). Such equilibria arise when: (i)* $n_1 > n_2$*,* $c_1 \in (\,\widehat{c}, c^*)$*, and* $c_2 \leq \widehat{c}$*; or (ii)* $n_1 \leq n_2$ *and* $c_2 \leq \widehat{c}$*.*

4 Privacy Regimes and Consumer Welfare

We now employ the findings in Section 3 to determine: (i) how the platform's privacy policy affects consumer welfare; (ii) whether data breaches or violations of a platform's privacy policy harm consumers; and (iii) the effects of policies that require consumers to "opt in" before their transactions data with one merchant are provided to other merchants. We find that the effects of privacy regimes, privacy breaches, and opt-in policies vary with the level of consumer sophistication. In particular, privacy regimes that best serve unsophisticated consumers can harm sophisticated consumers, and *vice versa*. Consequently, it typically is difficult to formulate Pareto-improving privacy policies, even if the exclusive concern is with consumer welfare.

4.1 Optimal Privacy Regimes for Consumers

We first examine the implications of our formal conclusions for unsophisticated consumers. Recall that an unsophisticated consumer does not account for the possibility that information about his interaction with one merchant might affect his interaction with another merchant. Further recall from Lemmas 2 and 3 that the *ex post* welfare of an unsophisticated consumer is never higher, and is sometimes lower, when the platform reveals transactions data to other merchants than under privacy. Therefore, the expected welfare of all unsophisticated consumers combined is higher under privacy than in its absence, and strictly so if consumers' reservation values and merchants' costs span all of the configurations in these lemmas.

Intuitively, this is the case because under privacy, a low-cost merchant charges each consumer a low price and a high-cost merchant charges each consumer a high price. In contrast, in the absence of privacy, a low-cost merchant who learns that a consumer is willing to pay a high price will charge that consumer a high price. Keeping transactions data private protects unsophisticated consumers by preventing a low-cost merchant from using information from a prior transaction to extract additional surplus from an unsophisticated consumer with a high reservation value.

Now consider the implications of our formal conclusions for sophisticated consumers. These consumers recognize that if data pertaining to their interaction with Merchant 1 are revealed, the revelation may affect their interaction with Merchant 2. Lemmas 4, 5, and 6 imply that in any PPBE, a sophisticated consumer's payoff is never lower, and is sometimes higher, in the absence of privacy than in its presence.

The welfare gain a sophisticated consumer may experience from the absence of privacy stems from the lower price (p_1) that Merchant 1 may charge the consumer for her product. In the absence of privacy, Merchant 1 must reduce p_1 to convince the consumer with a high reservation value ($r = \overline{r}$) to buy her product even though doing so leads Merchant 2 to infer that $r = \overline{r}$, and therefore to charge the consumer a high price ($p_2 = \overline{r}$) for her product.[15] In the absence of privacy, the sophisticated consumer with a high reservation value will intentionally sacrifice some surplus in his interaction with Merchant 1 if doing so convinces Merchant 2 that $r = \underline{r}$ and thereby leads her to charge the consumer a low price ($p_2 = \underline{r}$) rather than a high price ($p_2 = \overline{r}$) for her product. Merchant 1 recognizes that a sophisticated consumer with $r = \overline{r}$ will only purchase her product if she sets p_1 so low that the surplus the consumer secures when he buys Merchant 1's product ($n_1 [\overline{r} - p_1]$) exceeds the reduction in surplus the consumer suffers ($n_2 [\overline{r} - \underline{r}]$) when Merchant 2 infers that $r = \overline{r}$ and so increases

[15] This discussion pertains to the settings characterized in Lemma 5, where $n_1 > n_2$ and Merchant 1's cost is sufficiently high ($c_1 > c^*$) that she prefers to set p_1 above $\underline{r}$ and sell her product to the consumer only when he has the high reservation value ($r = \overline{r}$). As Lemma 6 reports, when Merchant 1's costs are more moderate (so $c_1 \in (\widehat{c}, c^*)$ and $n_1 > n_2$) or when $n_1 \leq n_2$ and $c_2 \leq \widehat{c}$, she will reduce her price from $\overline{r}$ all the way to $\underline{r}$ in the pooling equilibrium that arises in the absence of privacy.

the price she charges the consumer from $\underline{r}$ to $\overline{r}$. The resulting reduction in the price that Merchant 1 sets for her product increases the equilibrium welfare of a sophisticated consumer with the high reservation value $r = \overline{r}$.

Under privacy, Merchant 2 does not observe the details of the consumer's interaction with Merchant 1. Consequently, the consumer's strategic considerations (and the corresponding considerations of Merchant 1) no longer arise. In the absence of these strategic considerations, Merchant 1 will not offer a price concession to induce the consumer to reveal his high reservation value. The absence of this price concession harms the sophisticated consumer. In essence, concealing transactions data harms a sophisticated consumer by limiting his ability to protect himself against rent extraction by Merchant 2. In summary, we have:

Proposition 1 *Under the privacy policy that maximizes the welfare of unsophisticated (sophisticated) consumers, the platform reveals no (all) transactions data.*

Proposition 1 indicates that the optimal privacy regime varies with the extent of a consumer's sophistication. More strikingly, the proposition implies that the privacy regime that best serves unsophisticated consumers in the present setting is the worst privacy regime for sophisticated consumers, and *vice versa*.

4.2 Consumer Harm from Violations of Privacy Policies

We now examine how unanticipated data breaches or violations of the platform's privacy policy affect consumer welfare. Possible violations include deceptive or bait-and-switch policies under which a platform promises to keep transactions data private but reneges on the promise by releasing the data. Conversely, a platform might announce it will share data with third parties, but subsequently fail to do so. This failure might arise from a software or computer error, for instance. Alternatively, the failure might stem from the threat of a lawsuit alleging that data sharing facilitates anticompetitive actions by merchants. The ensuing discussion focuses on settings where the platform announces the privacy policy that maximizes consumer welfare (perhaps to help attract consumers to its platform or in response to a government mandate, for example).[16]

[16] Tsai et al. (2011) find that online shoppers tend to frequent retailers who announce policies that better protect shopper privacy.

Initially suppose the platform promises not to reveal transactions data to third parties—a policy that, if adhered to, maximizes the welfare of unsophisticated consumers. If the platform violates its stated privacy policy (either due to deception by the platform or a data breach caused by hackers), Merchant 2 will observe the details of prior transactions. Lemmas 2 and 3 imply that no unsophisticated consumers benefit from such a breach, and some may be harmed (for the reasons discussed below). Consequently, the breach (weakly) harms unsophisticated consumers. Lemma 4 implies the same is true for sophisticated consumers if they do not anticipate the breach.

The magnitude of the harm the unanticipated breach imposes on a consumer varies with his reservation value and the merchants' costs. Furthermore, the breach may not cause any consumer harm. To explain these conclusions, observe from Lemmas 2, 3, and 4 imply that a consumer with a low reservation value ($r = \underline{r}$) is not harmed by the breach because his welfare is zero both in the presence of a breach and in its absence. Likewise, a consumer with a high reservation value ($r = \overline{r}$) is not harmed by the breach when Merchant 1 has a low cost.[17] In contrast, a consumer with a high reservation value is harmed by the breach when Merchant 1 has a high cost ($c_1 > \widehat{c}$) and Merchant 2 has a low cost ($c_2 \leq \widehat{c}$). In this case, the breach reveals that the consumer's reservation value is $\overline{r}$, which leads Merchant 2 to increase the price she charges the consumer from $p_2 = \underline{r}$ to $p_2 = \overline{r}$.

To illustrate the corresponding *expected* aggregate consumer harm from an unanticipated breach, it is convenient to consider the *symmetric setting*, which has the following three features. First, each consumer's reservation value is either $\underline{r}$ (low) or $\overline{r}$ (high). Second, the probability that each of the N_S (N_U) sophisticated (unsophisticated) consumers has a high reservation value is θ_S (θ_U). Third, each of the Merchant i's ($i \in \{1, 2\}$) has the same cost (c_i) and faces the same consumer demand for her product. (Note that c_1 can differ from c_2.)

As noted above, no consumer with a low reservation value is harmed by the breach. Consequently, $\theta_U N_U + \theta_S N_S$ is an upper bound on the expected number of consumers who are harmed by the breach in the symmetric setting. For consumers who are potentially harmed, actual harm is zero if Merchant 1 has a low cost ($c_1 \leq \widehat{c}$) or if Merchant 2 has a

[17] When $c_1 \leq \widehat{c}$, Merchant 1 sets $p_1 = \underline{r}$ for the generic consumer, thereby ensuring the consumer always buys her product. Consequently, Merchant 2 learns nothing about the consumer's reservation value by observing the details of his transaction with Merchant 1.

high cost ($c_2 > \widehat{c}$). In both of these cases, the price Merchant 2 charges to each consumer is not affected by the breach. Only when Merchant 1 has a high cost and Merchant 2 has a low cost are these consumers harmed by the breach. The magnitude of the harm to each of the $\theta_U N_U + \theta_S N_S$ consumers is the product of the price increase he faces ($\overline{r} - \underline{r}$) and the number of units of Merchant 2's product he buys (n_2).

These observations underlie the conclusions in Proposition 2.

Proposition 2 *Suppose the platform announces it will implement the privacy policy that maximizes the welfare of unsophisticated consumers. Then consumers are harmed if transactions data are revealed, contrary to the platform's announcement. Furthermore, in the symmetric setting: (i) the expected number of consumers harmed by the unanticipated data revelation is at most* $\theta_U N_U + \theta_S N_S$*; and (ii) expected aggregate consumer harm is (a)* $[\theta_U N_U + \theta_S N_S] [\overline{r} - \underline{r}] n_2 > 0$ *if* $c_1 > \widehat{c}$ *and* $c_2 \leq \widehat{c}$*; and (b)* 0 *otherwise.*

Proposition 2 demonstrates that the violation of an announced privacy policy can harm both sophisticated and unsophisticated consumers. Furthermore, the expected number of consumers harmed by the violation and the expected aggregate harm can vary with the distribution of consumers' reservation values and with the magnitudes of merchants' costs. For some parameter configurations, the violation does not harm consumers.

In fact, the violation of a privacy policy can benefit rather than harm consumers. To see why, suppose the platform announces it will implement the privacy policy that maximizes the welfare of sophisticated consumers. Because all transactions data are revealed to third parties under this policy, it is apparent that a data breach will not affect the welfare of any consumer. Furthermore, suppose that, contrary to its announced policy, the platform does not reveal any transactions data to third parties. Proposition 1 implies that the resulting data concealment can benefit unsophisticated consumers with high reservation values ($r = \overline{r}$). This is the case because when Merchant 2 does not observe data from previous transactions, she cannot opportunistically charge a consumer a high price when she would otherwise charge him a low price.

Although this violation of the announced privacy policy benefits unsophisticated consumers, it can harm sophisticated consumers with high reservation values. It does so because under the platform's announced privacy policy, a sophisticated consumer may reject a high price from Merchant 1 in order to influence Merchant 2's beliefs about his reservation value. This behavior can, in turn, induce Merchant 1 to reduce the price she charges the consumer for her product. However, when Merchant 2 cannot observe data involving transactions with Merchant 1, a sophisticated consumer no longer has an incentive to reject a high price from Merchant 1, which can lead Merchant 1 to increase the price she sets for her product.[18]

In summary, we have:

Proposition 3 *Suppose the platform announces it will implement the privacy policy that maximizes the welfare of sophisticated consumers. Then a data breach does not harm consumers. In contrast, if the platform violates this privacy policy: (i) unsophisticated consumers are never harmed, and they secure strict gains when* $r = \overline{r}$*,* $c_1 > \widehat{c}$*, and* $c_2 \leq \widehat{c}$*; whereas (ii) sophisticated consumers never benefit, and are strictly harmed when* $n_1 > n_2$*,* $c_1 > c^*$*,* $c_2 \leq \widehat{c}$*, and* $r = \overline{r}$*.*

Proposition 3 has three primary implications. First, the effects of data breaches (by hackers, say) and violations of privacy policies can differ for sophisticated and unsophisticated consumers. Second, the effects of data breaches can differ from the effects of the platform's failure to abide by its stated privacy policy. Third, a privacy violation can harm sophisticated consumers, regardless of whether they anticipate the violation.

4.3 Impact of Opt-in and Opt-out Policies on Consumer Welfare

We now consider the effects of mandated "opt-in" and "opt-out" policies. Under an opt-in policy, the platform does not reveal a consumer's transactions data to third parties unless the consumer explicitly "opts in," thereby authorizing the sharing of his transactions data with third parties. Under an opt-out policy, the platform shares a consumer's transactions data with third parties unless the consumer explicitly "opts out," thereby ensuring that his transactions data are not shared with third parties.

[18] If a sophisticated consumer with a high reservation value does not learn of the violation of the announced privacy policy, he may continue to reject a price from Merchant 1 that is below $\overline{r}$. The consumer will suffer if this unsuccessful attempt to signal a low reservation value does not lead Merchant 2 to reduce the price she charges the consumer to $p_2 = \underline{r}$.

It is well-known that the welfare effects of opt-in or opt-out policies can vary with the prevailing status quo (e.g., Federal Trade Commission, 2009). We first consider the impact of an opt-in policy when the platform does not reveal transactions data to third parties under the initial status quo. Recall from Proposition 1 that this status quo maximizes the welfare of unsophisticated consumers. Consequently, a regulation that requires the platform to adopt an opt-in policy cannot increase the welfare of unsophisticated consumers above the level they secure under the initial status quo.

Different considerations arise for sophisticated consumers. Recall from Proposition 1 that the welfare of sophisticated consumers is lowest under the postulated status quo, where the platform does not reveal transactions data to third parties. Consequently, sophisticated consumers cannot be harmed by the adoption of an opt-in policy. Furthermore, if opting in is costless, a sophisticated consumer will ensure his preferred privacy regime by opting in, thereby authorizing the platform to reveal all of his transactions data to third parties.

In summary, we have:

Proposition 4 *Suppose the platform does not reveal transactions data to third parties under the initial status quo. Then a regulation that requires the platform to adopt an opt-in policy: (i) does not improve the welfare of unsophisticated consumers; but (ii) improves the welfare of sophisticated consumers in the absence of hassle costs, and strictly so when* $n_1 > n_2$, $c_1 > c^*$, *and* $c_2 \leq \widehat{c}$.

We now consider the alternative status quo in which the platform reveals transactions data to third parties. Suppose a regulation requires the platform to adopt an opt-out policy, so the platform cannot reveal a consumer's transactions data to third parties if the consumer explicitly opts out by requesting privacy for his transactions data. Proposition 1 implies that this regulation has the potential to increase the welfare of unsophisticated consumers by effectively allowing them to replace the status quo with the policy that maximizes their welfare (by not revealing transactions data to third parties). Also recall from Lemmas 1 and 3 that no unsophisticated consumer is harmed and unsophisticated consumers with high reservation values benefit when transactions data are concealed from, rather than revealed to, third parties. Therefore, if it is costless to opt out, all unsophisticated consumers will benefit if they opt out of the platform's status quo privacy policy. In doing so, unsophisticated consumers would ensure that the platform implements the privacy policy for their transactions data that maximizes their welfare.

More subtle considerations arise if consumers must incur a positive (but possibly negligible) hassle cost to opt out. In this case, only unsophisticated consumers with a high reservation value ($\overline{r}$) have an incentive to opt out (when $c_1 > \widehat{c}$, $c_2 \leq \widehat{c}$, and the hassle cost is less than the potential gain from opting out, $n_2 [\overline{r} - \underline{r}]$). However, by opting out, these consumers would effectively reveal their high reservation values by behaving differently than consumers with low reservation values, who do not opt out because they secure no strict gain by doing so. (Recall Lemmas 4, 5, and 6.) Such revelation would eliminate the potential gain from opting out. If an unsophisticated consumer recognizes that opting out would not allow him to secure a welfare gain, he will not bear the cost of opting out. In this event, the mandated opt-out policy will not affect his welfare. If an unsophisticated consumer fails to recognize the inference that would be drawn from his opting out and so incurs the hassle cost required to opt out, the mandated policy would harm him.

Now consider the corresponding considerations that arise in this same setting when a consumer is sophisticated. Proposition 1 implies that the welfare of sophisticated consumers is maximized under the platform's status quo policy of revealing transactions data. Consequently, sophisticated consumers will not exercise their option to opt out of this policy, even if doing so is costless. Proposition 5 summarizes these observations.

Proposition 5 *Suppose the platform reveals transactions data to third parties under the initial status quo. Then a regulation that requires the platform to adopt an opt-out policy does not affect the welfare of sophisticated consumers. In contrast, if it is costless to opt out, such a regulation increases the welfare of unsophisticated consumers who opt out, and strictly so when* $c_1 > \widehat{c}$ *and* $c_2 \leq \widehat{c}$.

Propositions 4 and 5 demonstrate that the effects of mandated opt-in or opt-out policies can vary widely, depending on the prevailing status quo, the costs of opting in or opting out, and the level of consumer sophistication.

5 Welfare, Platform Profit, and Alternative Policies

We now assess the impacts of privacy policies on platform profit and on total welfare (the sum of consumer and merchant welfare). We also consider the effects of two additional policies: (i) removing personally identifiable information (PII) from transactions data that are revealed to third parties; and (ii) precluding price discrimination.

5.1 Total Welfare and Platform Incentives

The analysis to this point has focused on consumer welfare, reflecting the primary concern of most antitrust and consumer protection agencies. However, consideration of total welfare is relevant for at least two reasons. First, even when agencies are instructed to protect consumer welfare, they often consider the impact of proposed policies on total welfare.[19] Second, the incentives of two-sided platforms, such as the online shopping platform in our model, typically are not fully aligned with the welfare of participants on just one side of the platform (e.g., Baye and Morgan, 2001). In particular, if the platform in our model sought to maximize its profit and could charge consumers and merchants to use the platform, the platform would adopt the privacy policy that maximizes the combined welfare of all consumers and merchants on the platform and employ fixed fees to extract their rent.

Transactions involving unsophisticated consumers make the same contributions to total welfare whether the platform reveals or does not reveal transactions data to third parties. This conclusion reflects two observations. First, Lemma 1 implies that Merchant 1's payoff is the same under privacy and in its absence. Second, Lemmas 2 and 3 imply that any increase (or reduction) in welfare an unsophisticated consumer experiences under one of the privacy regimes is exactly offset by a reduction (or increase) in Merchant 2's payoff. Therefore, total welfare does not vary across privacy regimes for transactions involving unsophisticated consumers.

Now consider transactions involving sophisticated consumers. Lemmas 4, 5, and 6 imply that the welfare of sophisticated consumers is highest when the platform reveals their

[19] The mission of the U.S. Federal Trade Commission is to protect consumers. However, Section 5 cases (those alleging an "unfair business practice") require an accounting of countervailing benefits to consumers or to competition. See the Federal Trade Commission Act Incorporating U.S. SAFE WEB Act amendments of 2006 at § 45 (Section 5), available at https://www.ftc.gov/sites/default/files/documents/statutes/federal-trade-commission-act/ftc_act_incorporatingus_safe_web_act.pdf.

transactions data to third parties. Total welfare from these transactions is also higher under this policy because surplus-enhancing sales are consummated more often when transactions data are revealed. The expanded sales arise in the absence of privacy because Merchant 1 reduces her price to account for the sophisticated consumer's incentive to reject an otherwise favorable price in an attempt to conceal information from Merchant 2.[20] To summarize:

Proposition 6 *Suppose the platform adopts the privacy policy that maximizes the welfare of sophisticated consumers by revealing transactions data to third parties. Then total welfare is maximized regardless of the prevailing mix of sophisticated and unsophisticated consumers in the population.*

Proposition 6 implies that if a profit-maximizing platform can fully extract surplus from merchants and consumers, then a laissez-faire policy that allows the platform to implement its preferred privacy policy will ensure the welfare of sophisticated consumers is maximized. However, this laissez-faire policy will not necessarily maximize the welfare of all relevant parties. In particular, the policy will leave unsophisticated consumers worse off than they are under a policy that prohibits the platform from providing transactions data to third parties. Furthermore, Lemmas 1, 4, 5, and 6 imply that Merchant 1 is never better off under the laissez-faire policy than when the platform cannot reveal transactions data, and she may be strictly worse off when she interacts with a sophisticated consumer.

5.2 Removing PII

We now consider the effects of an "intermediate" privacy policy in which PII is removed before transactions data is shared with third parties. We take PII to include any information that would allow Merchant 2 to infer from Merchant 1's transactions data the identity of the consumer with whom Merchant 2 is presently interacting. Observe that PII is not limited to a consumer's name, address, and telephone number. PII can also include the consumer's IP address or identifying information gleaned from cookies, for example.[21]

[20] This welfare improvement does not reflect the increased surplus that typically arises from a price reduction in the presence of a downward-sloping demand curve. Recall that a consumer's demand is completely price-inelastic below his reservation value in our model.

[21] The FTC (2009, footnote 47) observes that "Traditionally, PII has been defined as information that can be linked to a specific individual including, but not limited to, name, postal address, email address,

In our model, unsophisticated consumers are never harmed and may benefit when PII is removed from transactions data. In this case, Merchant 2 learns nothing about the reservation value of any particular consumer from Merchant 1's transactions. Consequently, if Merchant 2 has a low cost ($c_2 \leq \widehat{c}$), she will charge price $p_2 = \underline{r}$ to all consumers. In contrast, if Merchant 2 could observe all of Merchant 1's transactions data (including PII), she would charge a high price ($p_2 = \overline{r}$) for her product to any unsophisticated consumer that paid $p_1 = \overline{r}$ for Merchant 1's product. Therefore, removing PII increases the welfare of unsophisticated consumers with high reservation values ($r = \overline{r}$) when Merchant 1 has a high cost ($c_1 > \widehat{c}$) and Merchant 2 has a low cost ($c_2 \leq \widehat{c}$).

In contrast, sophisticated consumers do not benefit from the removal of PII and may be harmed. When PII is removed from transactions data in this setting, each sophisticated consumer recognizes that his interaction with Merchant 1 will reveal nothing about his personal reservation value (r) to Merchant 2. Consequently, each consumer acts precisely as unsophisticated consumers act. In particular, a sophisticated consumer gains nothing by rejecting a price below r from Merchant 1 because Merchant 2 cannot link this rejection to the identity of any particular consumer. Consequently, when she has a high cost ($c_1 > c^*$), Merchant 1 will charge sophisticated consumers a higher price when PII is removed from transactions data (when $n_1 > n_2$ and $c_2 \leq \widehat{c}$). Therefore, the welfare of sophisticated consumers declines, as does total welfare.[22] These observations provide:

Proposition 7 *Removing PII from transactions data benefits unsophisticated consumers but harms sophisticated consumers and reduces total welfare.*

Social Security number, or driver's license number ... [but in online markets] the traditional notion of what constitutes PII versus non-PII is becoming less and less meaningful ...".

[22] This conclusion refects our assumption that consumers' reservation values are independent. Suppose instead it is common knowledge that the reservation values of unsophisticated consumers are perfectly correlated. Further suppose Merchant 1 has a high cost ($c_1 > \widehat{c}$), so she charges each unsophisticated consumer price $p_1 = \overline{r}$. Then as long as transactions data reveals that some consumer purchased n_1 units of Merchant 1's product, Merchant 2 can infer that all unsophisticated consumers have the high reservation value ($r = \overline{r}$). Consequently, Merchant 2 will set $p_2 = \overline{r}$, which leaves unsophisticated consumers with the same welfare they achieve when PII is not removed from transactions data in the absence of privacy.

5.3 Banning Price Discrimination

We now consider the effects of precluding price discrimination by merchants. Observe first that under privacy, merchants do not acquire consumer-specific information that would allow them to benefit from price discrimination. Consequently, a ban on price discrimination would not affect merchant behavior (or merchant payoffs or consumer welfare).

Now consider the impact of such a ban in the absence of privacy when $r \in \{\underline{r}, \overline{r}\}$ for each of the $N > 1$ consumers. The ban does not affect merchant behavior when Merchant 1 has a low cost ($c_1 \leq \widehat{c}$). In this case, Merchant 1 optimally sets the same price ($p_1 = \underline{r}$) for all consumers when price discrimination is feasible, so she will do the same when price discrimination is prohibited. Therefore, Merchant 2 learns nothing about the reservation value of any particular consumer from observing Merchant 1's transactions data, so Merchant 2 optimally charges the same price to all consumers ($p_2 = \underline{r}$ if $c_2 \leq \widehat{c}$; $p_2 = \overline{r}$ if $c_2 > \widehat{c}$).

The analysis is more complex when $c_1 > \widehat{c}$. The effects of a ban on price discrimination in this case can vary with the number of consumers ($N > 1$) and the timing of their interactions with merchants. When N is small or a relatively large number of consumers interact with Merchant 1 before any consumers interact with Merchant 2, Merchant 2 may find it profitable to alter the non-discriminatory price she sets simply because the reservation values of the consumers who interact Merchant 1 before interacting with Merchant 2 differ from the reservation values of the general population of consumers.[23] To abstract from this consideration, the ensuing discussion considers the *large numbers setting* in which N is sufficiently large and the fraction of consumers who interact with Merchant 1 before Merchant 2 acts with any consumer is sufficiently small that the single, profit-maximizing price Merchant 2 sets for all consumers is not influenced by the transactions data she observes before her first interaction with a consumer.

In the large numbers setting, the first consumer to interact with merchants has no in-

[23] For instance, suppose each of N consumers act sequentially, interacting first with Merchant 1 and then with Merchant 2. Further suppose Merchant 2 learns from Merchant 1's initial transaction that the initial consumer's reservation value is $\overline{r}$. Then if Merchant 2 sets price $p_2 = \overline{r}$ for all consumers, her expected payoff is $E\,\pi(\overline{r}) = [\,1 + \phi\,(N-1)\,]\,[\,\overline{r} - c_2\,]\,n_2$ because Merchant 2 expects the fraction ϕ of her subsequent $N-1$ visitors to purchase at this price. Alternatively, if Merchant 2 sets price $p_2 = \underline{r}$ for all consumers, she knows all consumers will purchase her product, yielding expected payoff $E\,\pi(\underline{r}) = N\,[\,\underline{r} - c_2]\,n_2$. In this setting, Merchant 2 will set $p_2 = \underline{r}$ if $E\,\pi(\underline{r}) > E\,\pi(\overline{r}) \Leftrightarrow N > \widetilde{N} \equiv \frac{[\,1-\phi\,][\,\overline{r}-c_2\,]}{\underline{r}-c_2-\phi\,[\,\overline{r}-c_2\,]}$.

centive to reject a price from Merchant 1 that is below his reservation value. This is the case because such a rejection would not influence the price that Merchant 2 subsequently sets. Consequently, even when he is sophisticated, this first consumer acts myopically, just as unsophisticated consumers act. All other consumers do the same when merchants must charge the same price to all consumers.

When Merchant 1 recognizes that sophisticated consumers will act exactly as unsophisticated consumers act, she sets $p_1 = \overline{r}$ for all consumers when $c_1 > \widehat{c}$. Merchant 1 thereby sets a higher price for all consumers than she sets for sophisticated consumers when price discrimination is feasible. The higher price reduces the welfare of sophisticated consumers with $r = \overline{r}$ relative to the setting where the pooling PPBE (with $p_1 = \underline{r}$) identified in Lemma 6 arises when price discrimination is permitted.[24] A ban on price discrimination in this setting affects sophisticated consumers much like privacy does. The ban eliminates the incentive of a sophisticated consumer to curtail his purchase of Merchant 1's product in order to reduce Merchant 2's assessment of his reservation value. When this incentive is eliminated, the sophisticated consumer with $r = \overline{r}$ is harmed (when $c_1 > c^*$ and a pooling PPBE would arise if price discrimination were feasible) because Merchant 1 no longer reduces p_1 below $\overline{r}$ to convince the consumer not to conceal his high reservation value.

In contrast, a ban on price discrimination benefits an unsophisticated consumer with $r = \overline{r}$ when $c_1 > \widehat{c}$ and $c_2 \leq \widehat{c}$. In this case, the ban induces Merchant 2 to charge $p_2 = \underline{r}$ (because $c_2 \leq \widehat{c}$) even though she would charge $p_2 = \overline{r}$ to an unsophisticated consumer who buys Merchant 1's product at price $p_1 = \overline{r}$ when price discrimination is permitted. In essence, a ban on price discrimination protects unsophisticated consumers with high reservation values from selective exploitation by Merchant 2.

These observations are summarized in Proposition 8. Propositions 8 and 9 are proved in Baye and Sappington (2019).

[24] When $c_1 > \widehat{c}$, a sophisticated consumer with $r = \overline{r}$ pays a higher price for Merchant 1's product when price discrimination is banned than he pays for the product in the separating PPBE identified in Lemma 5. However, the consumer's overall welfare is the same whether price discrimination is permitted or banned. This is the case because the relatively low price ($p_1 = \widehat{p}_1$) the consumer pays for Merchant 1's product when price discrimination is permitted is offset by the high price ($p_2 = \overline{r}$) Merchant 2 charges the consumer who buys Merchant 1's product in the separating PPBE. This conclusion reflects expression (1).

Proposition 8 *In the absence of privacy in the large numbers setting, a ban on price discrimination: (i) harms sophisticated consumers, and strictly so when $r = \overline{r}$, $c_1 > \widehat{c}$, and a pooling PPBE would arise if price discrimination were permitted; but (ii) benefits unsophisticated consumers, and strictly so when $r = \overline{r}$, $c_1 > \widehat{c}$, and $c_2 \leq \widehat{c}$.*

Although a ban on price discrimination can increase the welfare of unsophisticated consumers, it never increases total welfare. When $c_1 > \widehat{c}$, the ban induces Merchant 1 to set $p_1 = \overline{r}$, which exceeds the price she sets for sophisticated consumers in the pooling PPBE identified in Lemma 6 when price discrimination is feasible. Sophisticated consumers with $r = \underline{r}$ do not purchase Merchant 1's product when $p_1 = \overline{r}$, so total welfare declines. A ban on price discrimination also reduces the total welfare from transactions with unsophisticated consumers with $r = \underline{r}$ when $c_1 > \widehat{c}$ and $c_2 > \widehat{c}$. In this case, the ban induces Merchant 2 to set $p_1 = \overline{r}$ for all consumers, whereas if price discrimination were feasible, she would set $p_2 = \underline{r}$ for unsophisticated consumers who do not purchase Merchant 1's product at price $p_1 = \underline{r}$. These observations provide:

Proposition 9 *In the absence of privacy in the large numbers setting, a ban on price discrimination reduces total welfare, and strictly so when: (i) a sophisticated consumer with $r = \underline{r}$ interacts with Merchant 1 when her cost is $c_1 > \widehat{c}$ and a pooling PPBE would arise in the absence of privacy if price discrimination were permitted; or (ii) an unsophisticated consumer with $r = \underline{r}$ interacts with Merchants 1 and 2 when their costs are $c_1 > \widehat{c}$ and $c_2 > \widehat{c}$, respectively.*

Propositions 8 and 9 imply that although a ban on price discrimination benefits unsophisticated consumers, it harms sophisticated consumers and reduces total welfare for much the same reason that privacy harms sophisticated consumers and reduces total welfare.

6 Extensions, Caveats, and Conclusions

The growing prevalence of "big data" has raised widespread concern about the use of these data. We have employed a simple model in the spirit of important predecessors (especially Taylor (2004) and Acquisti and Varian (2005)) to examine the effects of sharing transactions (price and quantity) data on an online platform. We found that such sharing can have important effects on consumer, merchant, and platform welfare. Relatively subtle effects can arise because the sharing of transactions data opens a channel through which sophisticated consumers may attempt to signal or conceal their reservation values for merchants' products.

We found that total welfare, the welfare of sophisticated consumers, and platform profit are all maximized when the platform provides transactions data to all merchants. In contrast, the welfare of unsophisticated consumers is maximized when no transactions data are shared with third parties. Consequently, an important tension arises. Privacy policies that best protect unsophisticated consumers may do so at the expense of sophisticated consumers. These policies may also reduce total welfare (and platform profit).

This tension between policies that best serve different types of consumers raises subtle considerations in the formulation of platform privacy policies. For example, opt-in or opt-out requirements can benefit unsophisticated consumers but harm sophisticated consumers. In addition, data breaches and violations of platform privacy policies can have different effects, and can affect sophisticated and unsophisticated consumers in different ways. Consequently, the most appropriate privacy policy for online shopping platforms typically will vary with the relevant social objective and with prevailing institutional features, including the status quo policy, the costs of opting into and out of a privacy policy, and the degree of consumer sophistication.[25]

Although our model is highly stylized, corresponding tensions arise in more complex settings. For example, in Baye and Sappington (2019), we consider the possibility that merchants cannot determine *ex ante* whether any particular consumer is sophisticated or unsophisticated. The key qualitative conclusions drawn above persist in this setting provided the fraction of sophisticated consumers in the population is sufficiently large. We also allow consumer demand for each merchant's product to vary continuously with the product's price. We identify conditions under which equilibria of the type that drive the key findings above continue to arise.

[25] Taylor and Wagman (2014, p. 81) similarly caution "that studies of consumer privacy must be understood within their individual context and industries, and that their conclusions depend on the specific competitive landscapes at play – and may not necessarily apply more broadly."

Of course, we have only considered policies that pertain to the privacy of basic transactions data—price and quantity data and the customer's identity. We have not considered the additional considerations that arise when transactions data include potentially sensitive financial or personal information (e.g., the consumer's health status). Explicit analysis of the appropriate treatment of such additional information merits further study.

In concluding, we note that our model has potential implications for antitrust policy, as well as consumer protection policy. We found that information sharing though a third party (the platform in our model) can increase total welfare in part by promoting the consummation of welfare-enhancing transactions. This finding lends support to the current antitrust practice in the U.S. which recognizes that information exchanges are not necessarily anti-competitive. However, this finding also suggests that current requirements for information sharing to fall in an "antitrust safety zone" may be unduly restrictive in some settings (e.g., when firms do not compete directly). Under current policy,

> "...the agencies will not [generally] challenge a data exchange if: (1) the exchange is managed by a third-party, like a trade association; (2) the information provided by participants is more than three months old; and (3) at least five participants provide the data underlying each statistic shared, no single provider's data contributes more than 25% of the "weight" of any statistic shared, and the shared statistics are sufficiently aggregated that no participant can discern the data of any other participant" (Bloom, 2014).

In our model, even the sharing of data that are current and that explicitly identify the specific data source can enhance welfare.

References

Acquisti, A., Taylor, C., & Wagman, L. (2016). The economics of privacy. *Journal of Economic Literature*, 52(2), 442-92.

Acquisti, A., & Varian, H. (2005). Conditioning prices on purchase history. *Marketing Science*, 24(3), 367-381.

Athey, S., Catalini, C., & Tucker, C. (2017). The digital privacy paradox: Small money, small costs, small talk. National Bureau of Economic Research Working Paper w23488.

Baye, M., & Morgan, J. (2001). Information gatekeepers on the internet and the competitiveness of homogeneous product markets. *American Economic Review*, 91(3), 454-474.

Baye, M., & Sappington, D. (2019). Technical appendix to accompany 'Revealing transactions data to third parties: Implications of privacy regimes for welfare in online markets. Available at http://nash-equilibrium.com/PDFs/Appendix.pdf.

Belleflamme, P., & Vergote, W. (2016). Monopoly price discrimination and privacy: The hidden cost of hiding. *Economics Letters*, 149, 141-144.

Bloom, M. (2014). Information exchange: Be reasonable. Federal Trade Commission. Available at https://www.ftc.gov/news-events/blogs/competition-matters/2014/12/information-exchange-be-reasonable.

Brynjolfsson, E., Hu, Y., & Smith, M. (2003). Consumer surplus in the digital economy: Estimating the value of increased product variety at online booksellers. *Management Science*, 49(11), 1580-1596.

Calzolari, G., & Pavan, A. (2006). On the optimality of privacy in sequential contracting. *Journal of Economic Theory*, 130(1), 168-204.

Campbell, J., Goldfarb, A., & Tucker, C. (2015). Privacy regulation and market structure. *Journal of Economics & Management Strategy*, 24(1), 47-73.

Conitzer, V., Taylor, C., & Wagman, L. (2012). Hide and seek: Costly consumer privacy in a market with repeat purchases. *Marketing Science*, 31(2), 277-292.

Ellison, G., & Ellison, S. (2018). Match quality, search, and the Internet market for used books. National Bureau of Economic Research Working Paper No. w24197.

Evans, D. (2009). The online advertising industry: Economics, evolution, and privacy. *Journal of Economic Perspectives*, 23(3), 37-60.

Federal Trade Commission. (2009). Self-regulatory principles for online behavioral advertising. FTC Staff Report. Available at https://www.ftc.gov/sites/default/files/documents/reports/federal-trade-commission-staff-report-self-regulatory-principles-online-behavioral-advertising/p085400behavadreport.pdf.

Grimaldi, J., & Kendall, B. (2019). The government vs. tech giants. *The Wall Street Journal*, September 10, B4.

Kim, J., & Wagman, L. (2015). Screening incentives and privacy protection in financial markets: A theoretical and empirical analysis. *RAND Journal of Economics*, 46(1), 1-22.

Reinganum, J. (1979). A simple model of equilibrium price dispersion. *Journal of Political Economy*, 87(4), 851-858.

Spulber, D. (2019). The economics of markets and platforms. *Journal of Economics and Management Strategy*, 28(1), 159-172.

Taylor, C. (2004). Consumer privacy and the market for customer information. *RAND Journal of Economics*, 35(4), 631-650.

Taylor, C., & Wagman, L. (2014). Consumer privacy in oligopolistic markets: Winners, losers, and welfare. *International Journal of Industrial Organization*, 34, 80-84.

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

Tucker, C. (2012). The economics of advertising and privacy. *International Journal of Industrial Organization*, 30(3), 326-329.