



Contents lists available at ScienceDirect

## International Journal of Industrial Organization

journal homepage: [www.elsevier.com/locate/ijio](http://www.elsevier.com/locate/ijio)

## Digital monopolies: Privacy protection or price regulation?☆

Simon Loertscher<sup>a</sup>, Leslie M. Marx<sup>b,\*</sup><sup>a</sup> Department of Economics, Level 4, FBE Building, 111 Barry Street, University of Melbourne, Victoria 3010, Australia<sup>b</sup> Fuqua School of Business, Duke University, 100 Fuqua Drive, Durham, NC 27708, USA

## ARTICLE INFO

## Article history:

Received 28 October 2019

Revised 5 March 2020

Accepted 27 April 2020

Available online 19 May 2020

## JEL classification:

C72

D72

## Keywords:

Natural monopoly

Privacy concerns

Big data

Transparency

Ramsey pricing

## ABSTRACT

Increasing returns to scale in data gathering and processing give rise to a new form of monopoly, referred to here as *digital monopoly*. Digital monopolies create new challenges for regulators and antitrust authorities. We address two in this paper: market power arising from improved match values and from reduced privacy. The digital monopoly's profit and social surplus always increase as privacy decreases. However, consumer surplus is non-monotone in privacy. Without privacy, the match value is perfect but completely extracted by the digital monopoly. In contrast, as privacy goes to infinity, match values and social surplus go to zero. With regulated prices, consumer surplus is maximized without privacy protection. As with natural monopolies, price regulation thus remains an appropriate tool in the digital age to capture the social benefits from increasing returns to scale without harming consumers.

© 2020 Elsevier B.V. All rights reserved.

## 1. Introduction

Larger markets are better, all else equal, because they can execute the same trades as smaller, standalone markets, and sometimes execute more or more valuable trades. Consistent with this, Internet-based matchmakers that realize powerful data-driven increasing returns to scale, such as Amazon, Google, and Spotify, have come to prominence in the digital age. Firms that operate in environments for which efficiency dictates that a single firm is optimal are naturally referred to as *digital monopolies*.<sup>1</sup> Just as was the case with natural monopolies, digital monopolies call for antitrust scrutiny and possibly regulation. Indeed, recently digital monopolies have received intense scrutiny from antitrust authorities around the world.<sup>2</sup>

☆ We thank David Byrne, Curt Taylor, Peter Taylor, participants at the JFTC's 18th CIRC International Symposium on Data Concentration on Digital Markets and Competition Policy, 2019 Workshop on Digital Economy and Industrial Organisation at Monash University, and 2019 Competition Economists Network Meeting for useful comments and fruitful discussions. Bing Liu provided excellent research assistance. We gratefully acknowledge support from the Samuel and June Hordern Endowment, a University of Melbourne Faculty of Business & Economics Eminent Research Scholar Grant, and the Australian Research Council under Discovery Project Grant DP200103574.

\* Corresponding author.

E-mail addresses: [simonl@unimelb.edu.au](mailto:simonl@unimelb.edu.au) (S. Loertscher), [marx@duke.edu](mailto:marx@duke.edu) (L.M. Marx).

<sup>1</sup> To the extent that private firms are marketmakers and matchmakers, the idea that, under efficiency, these firms are monopolies—and that there is thus no obvious competitive benchmark—is natural and nothing new. Indeed, as noted by Malamud and Rostek (2017), the basic economic model of a market is a centralized exchange at which all trade occurs. The novelty and virulence of issues associated with such firms arises because digitalization and globalization have increased the force of the underlying returns to scale in marketmaking.

<sup>2</sup> In November 2018, the U.S. Federal Trade Commission held hearings on "The Intersection of Big Data, Privacy, and Competition." In March 2019, the United Kingdom released a report on "Unlocking Digital Competition: Report of the Digital Competition Expert Panel." In June 2019, the Australian

Traditionally, regulation and policy intervention have worked best when they were guided by well-defined objectives such as consumer or social surplus. In this tradition, we analyze the pros and cons of interventions in an environment in which a digital monopoly can use data to either improve matching only or, instead, to improve matching *and* to adjust pricing. Although this distinction has typically not been formulated explicitly, it is a key issue in ongoing antitrust debates. As a case in point, it makes a difference to advertisers whether Google uses its data only to better match advertisers to consumers or, alternatively, to improve matching *and* to adjust the (reserve) prices that it charges advertisers.

Based on a parsimonious model in which more data improves the distribution from which the consumer draws its value, with the improvement being in the sense of hazard rate dominance, we show that the distinction has striking implications for the consumer surplus effects of privacy protection. If data are used exclusively to improve match values, then consumer surplus increases monotonically in the data to which the digital monopoly has access. Put differently, in this case privacy protection unambiguously harms the consumer. In sharp contrast, when the monopoly also uses the information about the consumer's preferences for pricing purposes, the consumer surplus consequences of privacy protection are less clear cut. To a lesser or greater extent, the monopoly extracts part of the additional surplus generated by improvements in matching. In the limit, as the matching becomes perfect, consumers have no private information left and hence lose their entire information rent, while the monopoly captures the entire social surplus. In both cases, social surplus is maximized when all information is revealed to the digital monopoly. However, when data are also used for pricing, the monopoly is not only able to perfectly match the product to the consumer, but also to match the price to the consumer's value, thereby, in the limit, depriving the consumer of all surplus.

As an example, consider the online firm Ziprecruiter, which matches potential employers to jobseekers. The data collected by Ziprecruiter regarding the characteristics of a potential employer both improves match values, to the benefit of the employer, and allows Ziprecruiter to more precisely estimate the employer's willingness to pay for the service, to the detriment of the employer.<sup>3</sup>

From a consumer surplus perspective, the central issue is *not* the protection of privacy but rather the protection of information rents. In our model, fixing the level of data held by the digital monopoly, the protection of information rents can be achieved by regulating prices.<sup>4</sup> If the price is fixed, then data can only be used to improve match values, and improving match values is in the digital monopoly's best interest because it increases the probability of a trade, and, of course, is in the consumer's best interest.

The obvious flip side to the dire implications for consumer surplus when privacy vanishes completely and the digital monopoly's pricing is not restricted is that producer surplus increases and becomes identical to social surplus. Digital monopolies can thus be expected to resist attempts to regulate their pricing. Apart from this natural, and in many ways inevitable, conflict about the division of social surplus, a potential drawback to price regulation is that it may decrease the digital monopoly's incentives to invest in data analytics and product quality. If price regulation decreases equilibrium investments substantively, then there is a tradeoff between the social surplus and consumer surplus that can be achieved via regulating pricing.

That being said, this paper is exclusively concerned with issues pertaining to what are sensibly called *private values* settings, in which matching individuals' preferences as closely as possible is what a benevolent social planner would do.

The remainder of this paper is organized as follows. In [Section 2](#), we describe the model. In [Section 3](#), we derive results and discuss price regulation. In [Section 4](#), we extend the model to allow for investments into data analytics and product quality. In [Section 5](#), we provide an extension to allow competition. In [Section 6](#), we discuss related literature and provide additional discussion of the implications of our results for policy debates surrounding property rights. [Section 7](#) contains conclusions.

## 2. Model

Consider a setup with one consumer (the buyer) and one digital monopoly (the seller), both assumed to be risk neutral. The consumer has value  $v$  for one unit of a product provided by the digital monopoly, which is the consumer's private information, drawn from the distribution  $F_n$ . The parameter  $n > 0$  represents the extent of data collection by the digital monopoly. The digital monopoly does not observe the consumer's value, but knows  $F_n(\cdot)$ . Data collection is modelled by assuming that

$$F_n(v) \equiv F^n(v),$$

Competition & Consumer Commission released its "Digital Platforms Inquiry: Final Report." Also in 2019, the European Commission released "Competition Policy for the Digital Era." The U.S. Department of Justice announced in July 2019 that it is "Reviewing the Practices of Market-Leading Online Platforms" (Press Release 19–799).

<sup>3</sup> According to [Dubé and Misra \(2017\)](#), potential employers purchase a monthly subscription to Ziprecruiter to access resumés that the online system matches with the employer's needs. Job seekers can upload resumes at no charge. The matching process relies on the employer's responses to questions regarding characteristics of its business and the specific job posting. Only after these questions are answered does Ziprecruiter generate a price quote, at which point it must decide whether to subscribe.

<sup>4</sup> Price regulation is a traditional policy instrument for dealing with natural monopolies. Disregarding the digital monopoly's resistance to price regulation, regulating the pricing of a digital monopoly may be straightforward insofar as it "only" requires inspection of the algorithms that the digital monopoly employs.

where  $F$  is a distribution with support  $[\underline{v}, \bar{v}]$  and bounded density  $f(v) > 0$  for all  $v \in (\underline{v}, \bar{v})$ . Increases in  $n$  correspond a higher match value and a correspondingly lower degree of consumer privacy. While it is not strictly necessary to formally define these terms, it is certainly possible and maybe desirable. A natural definition of *match value* is the expectation of  $v$  conditional on exceeding some threshold  $t \in [\underline{v}, \bar{v}]$ :  $\mathbb{E}[v \mid v \geq t]$ . Likewise, a measure of consumer *privacy* is the probability that  $v$  is below some threshold  $t \in (\underline{v}, \bar{v})$ :  $\Pr(v \leq t)$ . This notion of privacy is meaningful because it is the consumer's ability to pretend to be a lower type that is the source of the consumer's information rent. Because  $F_{n'}(v) \leq F_n(v)$  for  $n' > n$ , increases in  $n$  increase the match value and decrease privacy.

Although we treat  $n$  as a continuous variable for technical convenience, there is a natural interpretation when  $n$  is an integer because then the setup is equivalent to the digital monopoly having  $n$  products, with the consumer's value being drawn independently from the distribution  $F$  for each of these products, and the digital monopoly offering the consumer the product for which the consumer has the highest value.<sup>5</sup> For limiting values of  $n$ , we also have a clear interpretation: as data collection goes to infinity, the density  $f_n$  converges to a point mass on  $v = \bar{v}$ , and as data collection goes to zero,  $f_n$  converges to a point mass on  $v = \underline{v}$ .

The optimal price offer of a digital monopolist's with cost  $c \in [\underline{v}, \bar{v}]$  is the price  $p$  that maximizes  $(p - c)(1 - F_n(p))$ , which has first-order condition  $1 - F_n(p) - (p - c)f_n(p) = 0$ . We can write this as  $f_n(p)(c - \Phi_n(p)) = 0$ , where  $\Phi_n$  is the consumer's virtual value function given by

$$\Phi_n(v) \equiv v - \frac{1 - F_n(v)}{f_n(v)}.$$

As noted by [Bulow and Roberts \(1989\)](#), the virtual value function can be interpreted as a buyer's marginal revenue function, treating the (change in the) probability of trade as the (marginal change in) quantity. If  $\Phi_n$  is increasing, then the second-order condition is satisfied when the first-order condition is, implying that the digital monopolist's problem is quasi-concave. We assume that  $\frac{1 - F(v)}{f(v)}$  is nonincreasing in  $v$ , which is sufficient to ensure that  $\Phi_n(v)$  is increasing for all  $n \geq 1$ . This implication follows from the monotonicity of the hazard rate, which we establish, along with other properties, in the following lemma:

**Lemma 1.** For  $n \geq 1$ ,  $\frac{1 - F_n(v)}{f_n(v)}$  is nonincreasing in  $v$ , and for any  $n > 0$  and  $v < \bar{v}$ ,  $\frac{1 - F_n(v)}{f_n(v)}$  increases in  $n$  and goes to infinity as  $n \rightarrow \infty$ .

**Proof.** See [Appendix Appendix A](#). □

The digital monopoly's expected profit when it sets the price  $p$  and has a cost  $c$  is

$$\Pi_n(p) = (p - c)(1 - F_n(p)),$$

which we interchangeably refer to as *producer surplus*. To focus on the interesting case when there are potentially gains from trade, we assume that  $c < \bar{v}$ . Moreover, we assume that  $c \geq \underline{v}$ . This implies that, for any finite  $n$ , the monopoly does not sell with probability one. *Consumer surplus* is

$$CS_n(p) = \int_p^{\bar{v}} (v - p)f_n(v)dv = \int_p^{\bar{v}} (1 - F_n(v))dv. \tag{1}$$

Consequently, *social surplus* is

$$SS_n(p) = CS_n(p) + \Pi_n(p).$$

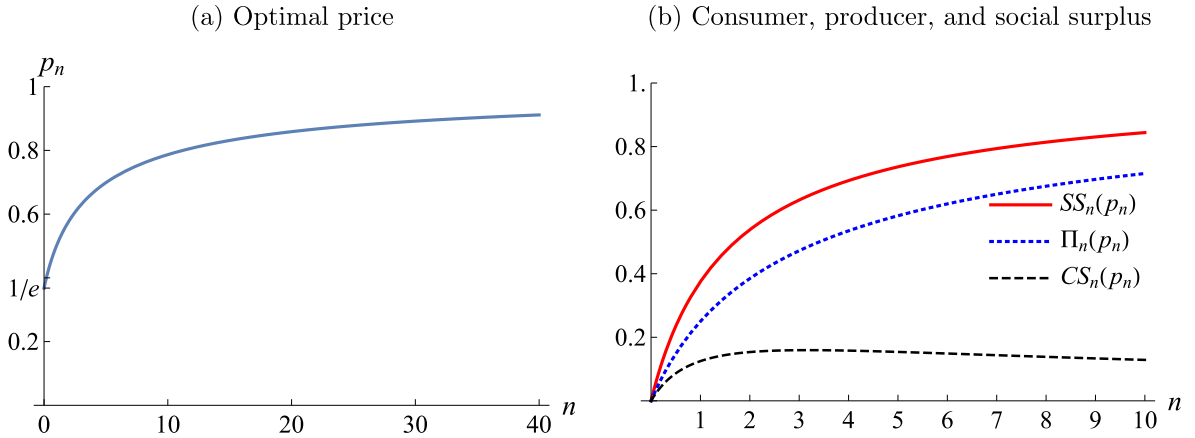
As just described, our model presumes that data collection by the digital monopoly has no direct cost for the consumers or, for that matter, for the monopoly. Data collections that is free for consumers seems an appropriate description for how many of the largest and most controversially debated digital platforms such as Google, Amazon, and Facebook generate data: consumers' user behavior generates the data that is informative and valuable for the platforms. In other applications, such as the Ziprecruiter example mentioned in the introduction, data collection by the platform may be costly for the platform's customer itself. We account for data collection that is costly for the digital monopoly in an extension in [Section 4](#), and we discuss the implications for our main conclusions of costly data collection for consumers in [Section 6.2](#).

### 3. Results

Differentiating  $\Pi_n(p)$  with respect to  $p$  yields a first-order condition that is satisfied with  $p_n$  such that  $\Phi_n(p_n) = c$ . For  $n \geq 1$ , by [Lemma 1](#), the second-order condition is satisfied whenever the first-order condition is.<sup>6</sup> Thus, we have the following characterization of the digital monopoly's optimal price:

<sup>5</sup> Alternatively, one could view the digital monopoly or a regulatory agency as choosing the level of consumer privacy, denoted by  $\rho$ , where the consumer's value distribution is  $F_\rho(v) \equiv 1 - (1 - F(v))^\rho$ . This setup has the interpretation, when  $\rho \geq 2$  is an integer, that a consumer with privacy preference parameter  $\rho$  is willing to forego the top  $\rho - 1$  out of  $\rho$  randomly selected products from the digital monopoly in order to preserve the privacy of its data. The tradeoffs we identify remain the same in this alternative parameterization.

<sup>6</sup> When  $n < 1$ , there may be multiple solutions to the first-order condition. The second-order condition then requires the  $\Phi_n(p_n)$  be increasing at  $p_n$ .



**Fig. 1.** Panel (a) Optimal price  $p_n$  as a function of  $n$ . Panel (b): Expected consumer, producer, and social surplus at the optimal price  $p_n$  as functions of  $n$ . Consumer surplus is maximized at  $n^* \approx 3.17$ . Both panels assume that  $F_n(v) = v^n$  and  $c = 0$ .

**Theorem 1.** *The digital monopoly's optimal price  $p_n$  increases in  $n$  for  $n \geq 1$  and goes to  $\bar{v}$  as  $n$  goes to infinity,  $\lim_{n \rightarrow \infty} p_n = \bar{v}$ .*

*Proof of Theorem 1.* As argued above, for  $n \geq 1$ , the digital monopoly's optimal price satisfies  $\Phi_n(p_n) = c$ . Lemma 1 implies that  $\Phi_n(p)$  increases in  $p$  and that for all  $v < \bar{v}$ ,  $\Phi_n(v)$  decreases in  $n$ , which implies that  $p_n$  increases in  $n$ . Next consider the result that  $\lim_{n \rightarrow \infty} p_n = \bar{v}$ . For any  $n$ ,  $\Phi_n(\bar{v}) = \bar{v}$ , which implies that  $\bar{p}_n$  is bounded above by  $\bar{v}$ . Because  $\{p_n\}_{n=1}^{\infty}$  is an increasing, bounded sequence,  $\lim_{n \rightarrow \infty} p_n$  exists. Suppose that  $\lim_{n \rightarrow \infty} p_n = \bar{p} < \bar{v}$ . Then, by Lemma 1, there exists  $\bar{n}$  sufficiently large that  $\Phi_{\bar{n}}(\bar{p}) < c$ , which implies that  $p_{\bar{n}} > \bar{p}$ , contradicting the supposition that  $\bar{p}$  is the limit. Thus,  $\lim_{n \rightarrow \infty} p_n = \bar{v}$ .  $\square$

Theorem 1 implies that a reduction in the consumer's degree of privacy both increases match quality and also causes the digital monopoly to charge ever increasing prices to the consumer. In addition, using Theorem 1 and the boundedness of the integral in (1), it follows that

$$\lim_{n \rightarrow \infty} CS_n(p_n) = 0.$$

Therefore, given that for all  $n \in [1, \infty)$ ,  $CS_n(p_n) > 0$ , it follows that  $CS_n(p_n)$  achieves its maximum over  $n \in [1, \infty)$  at some finite  $n^*$ .<sup>7</sup> This gives us the following corollary:

**Corollary 1.** *Consumer surplus is maximized at a finite value of  $n$ .*

With reduced data collection, match values deteriorate. Although our paper focuses on the case of  $n \geq 1$ , one can, of course, allow  $n$  to go to zero, in which case total expected surplus goes to zero, harming all parties. In contrast, with high levels of data collection, match values are high and the digital monopoly extracts all of the consumer's information rent, which is good for the digital monopoly and society, but leaves the consumer again with zero expected surplus. The optimum for the consumer is an intermediate level of data collection.

In contrast, the expected surplus of the digital monopoly is increasing in data collection. Assuming that  $n^*$  is unique, then because  $SS_n(p_n)$  is bounded above by  $\bar{v} - c$ , it follows that an increase in  $n$  increases both  $CS_n(p_n)$  and  $SS_n(p_n)$  for  $n < n^*$ . However, for  $n > n^*$ , the digital monopoly has an incentive to increase data collection, whereas the consumer would want to decrease data collection. In other words, for  $n > n^*$ , a social surplus perspective (and the digital monopoly's perspective) gives rise to policy recommendations that are detrimental to the consumer.

As illustrated in Fig. 1, for the case of  $F$  uniform on  $[0,1]$  and  $c = 0$ , the digital monopoly maximizes its profit by setting a personalized price to the consumer of:<sup>8</sup>

$$p_n = \left( \frac{1}{n+1} \right)^{\frac{1}{n}}, \tag{2}$$

with the properties that

$$\lim_{n \rightarrow \infty} p_n = 1, \quad \lim_{n \rightarrow 0} p_n = \frac{1}{e} \quad \text{and} \quad F_n(p_n) = \frac{1}{n+1}.$$

<sup>7</sup> In addition, using  $\lim_{n \rightarrow 0} f_n(x) = \lim_{n \rightarrow 0} nF^{n-1}(x)f(x) = 0$ , it follows that  $\lim_{n \rightarrow 0} CS_n(p_n) = 0$ , so consumer surplus is maximized over  $n \in [0, \infty)$  at some interior value  $n^* \in (0, \infty)$ .

<sup>8</sup> In the parameterization considered here, although the virtual value function  $\Phi_n(v) \equiv v - (1 - F_n(v))/f_n(v)$  is not necessarily strictly increasing for all  $v \in [\underline{v}, \bar{v}]$ ,  $\Phi_n(v)$  is increasing when it is positive, which implies that for any  $p \in (0, 1]$ , there is a unique  $v$  such that  $\Phi_n(v) = p$ . Thus,  $\Phi_n^{-1}(p)$  is well defined for  $p \in (0, 1]$ .

In particular, in the limit as  $n$  goes to infinity, the digital monopoly generates value  $\bar{v}$  for the consumer, but then extracts it all through a price that is also equal to  $\bar{v}$ . Indeed, in the limit the digital monopoly captures all of social surplus:

$$\lim_{n \rightarrow \infty} \Pi_n(p_n) - SS_n(p_n) = 0.$$

Further, if  $\lim_{n \rightarrow \infty} F_n(p_n) = 0$ , as is the case when  $F$  is uniform and  $c = 0$ , then

$$\lim_{n \rightarrow \infty} \Pi_n(p_n) = \bar{v} - c = \lim_{n \rightarrow \infty} SS_n(p_n).$$

In this case, a digital monopoly that is neither restricted in the use of data nor in pricing maximizes social surplus and captures all of it.<sup>9</sup> This reflects the strong incentives digital monopolies have for increasing the amount of data and improving data analytics by, for example, acquiring new data sources. It also suggests that digital monopolies can be expected to have strong incentives to resist policies that restrict pricing.

Because the benefits from big data accrue disproportionately, and in the limit uniquely, to the digital monopoly in the absence of price constraints, one could consider some form of price regulation. A natural approach would seem to be Ramsey pricing, according to which the regulated price  $\tilde{p}_n$  maximizes  $\alpha \Pi_n(p) + (1 - \alpha)SS_n(p)$  for some  $\alpha \in (0, 1)$ . Unfortunately, Ramsey pricing does not solve the problem because, as  $n$  becomes large,  $\tilde{p}_n$  converges to  $\bar{v}$  for any  $\alpha > 0$ .<sup>10</sup> As an alternative, the regulator can choose  $\tilde{p}$  such that the share of social surplus accruing to the consumer stays constant:  $CS_n(\tilde{p}) = \alpha SS_n(\tilde{p})$ , where  $\alpha \in (0, 1)$  measures the consumer's "fair" share. Using the definition of  $SS_n(p)$  and rearranging yields

$$\frac{1 - \alpha}{\alpha} = \frac{(1 - F_n(\tilde{p}))(\tilde{p} - c)}{\int_{\tilde{p}}^{\bar{v}} (1 - F_n(v)) dv}. \tag{3}$$

Because the right side increases in  $\tilde{p}$  for any  $\tilde{p} \leq \Phi_n^{-1}(c)$  and ranges from 0 to infinity as  $\tilde{p}$  varies from  $c$  to  $\bar{v}$ ,<sup>11</sup> it follows that for any fixed  $\alpha$ , there is a unique price that satisfies (3). Moreover, this price decreases in  $\alpha$  and is bounded away from  $\bar{v}$  for any  $\alpha > 0$ .

#### 4. Incentives to invest

The stark (and perhaps dismal) prediction of this model of big data and consumer privacy sheds light on optimal regulatory policies for digital monopolies. However, our results are obtained under the assumption that increasing match value for a given set of data is costless for the monopoly. We now relax this assumption by studying the digital monopoly's incentives to invest. Throughout this section, we assume that the regulated price does not vary with  $n$ .

##### 4.1. Investments in data analytics

We first analyze the digital monopoly's marginal incentives to invest in data analytics. We contrast the marginal incentives for such investment when the price is chosen optimally by the digital monopoly to when the price is fixed at a lower level, e.g., as a result of price regulation.

The effect of an increase in  $n$  on the digital monopoly's profit is

$$\frac{\partial \Pi_n(p)}{\partial n} = -(p - c)F^n(p) \ln(F(p)) > 0.$$

As one would expect, for a given price, the digital monopoly's expected payoff is increasing in the level of data analytics. Further, note that

$$\frac{\partial^2 \Pi_n(p)}{\partial n \partial p} = -F^n(p) \ln(F(p)) - (p - c)nF^{n-1}(p)f(p) \ln(F(p)) - (p - c)F^{n-1}(p)f(p). \tag{4}$$

Because the first two terms of (4) are positive and the final term goes to zero as  $p$  goes to  $c$ , it follows that for  $p$  sufficiently small, we have

$$\frac{\partial^2 \Pi_n(p)}{\partial n \partial p} > 0. \tag{5}$$

The digital monopoly's profit increases in  $n$  because an increase in data increases the probability of trade for any given  $p$ , but the marginal profit associated with an increase in  $n$  depends on the price level. Intuitively, for a price close to  $c$ , margins

<sup>9</sup> If instead  $\lim_{n \rightarrow \infty} 1 - F_n(p_n) < 1$ , then although the digital monopoly captures all of the social surplus created, it does not maximize social surplus because the probability of trade remains below the efficient level. That is, the digital monopoly prices above the level that would maximize social surplus. It is, however, an open question whether distributions  $F$  exist such that  $\lim_{n \rightarrow \infty} F_n(p_n) > 0$ .

<sup>10</sup> This result follows by the same logic as in the proof of Theorem 1. To see this, notice that the first-order condition for the Ramsey price is equivalent to  $\Phi_{\alpha,n}(\tilde{p}_n) = c$ , where  $\Phi_{\alpha,n}(v) = v - \alpha(1 - F_n(v))/f_n(v)$ , which is increasing in  $v$  by assumption. For any  $n$ ,  $\Phi_{\alpha,n}(\bar{v}) = \bar{v}$ , and by Lemma 1, for any  $v < \bar{v}$ ,  $(1 - F_n(v))/f_n(v)$  increases in  $n$  and goes to infinity as  $n$  goes to infinity. Thus,  $\tilde{p}_n$  is increasing in  $n$  and bounded above by  $\bar{v}$ , which implies that  $\lim_{n \rightarrow \infty} \tilde{p}_n$  exists. Suppose that  $\lim_{n \rightarrow \infty} \tilde{p}_n = \tilde{p} < \bar{v}$ . Then there exists  $\bar{n}$  sufficiently large that  $\Phi_{\alpha,\bar{n}}(\tilde{p}) < c$ , which implies that  $p_{\bar{n}} > \tilde{p}$ , contradicting the supposition that  $\tilde{p}$  is the limit. Thus, for any  $\alpha > 0$ ,  $\lim_{n \rightarrow \infty} \tilde{p}_n = \bar{v}$ .

<sup>11</sup> The right side of (3) is zero at  $\tilde{p} = c$ . Using L'Hôpital's rule allows one to show that it goes to infinity as  $\tilde{p}$  goes to  $\bar{v}$ .

are small, so increasing the probability of trade is not particularly valuable. As the price increases above  $c$ , the marginal profit from data increases, consistent with (5). In contrast, for prices that are well above  $c$ , further increases in price do not necessarily increase the marginal profit from data.

This implies that regulation that imposes a sufficiently low price reduces the digital monopoly's marginal incentive to invest in data analytics relative to a digital monopoly that is free to choose its price. Further, if  $F$  is the uniform distribution on  $[0,1]$  and  $c = 0$ , then for any price  $p$  less than the optimal price  $p_n$ , we have  $\frac{\partial^2 \Pi_n(p)}{\partial n \partial p} > 0$ .<sup>12</sup> Thus, in this case, any binding price regulation reduces the digital monopoly's incentive to invest in data.<sup>13</sup> Using these results, we can connect "big data" and privacy concerns with incentives to invest:

**Proposition 1.** *Binding price regulation reduces a digital monopoly's marginal incentive to invest in data analytics if the unregulated price is sufficiently low, and for some distribution and cost assumptions, any binding price regulation reduces the digital monopoly's marginal incentive to invest in data analytics.*

Proposition 1 shows that for the setting we consider here, price regulation imposed on a digital monopoly reduces the digital monopoly's marginal incentive to invest in data analytics.<sup>14</sup>

#### 4.2. Product quality investments

A digital monopoly can also improve match values by directly improving the quality of the product it offers. In our model, such investments can be captured as investments that directly improve the underlying value distribution  $F$ . For example, the streaming giant Netflix has become a vertically integrated firm that produces a considerable amount of content in-house (Koblin, 2017). Arguably, Netflix has an advantage in content production because of its access to viewer data, which allows it to tailor content to fit the preferences of customers. This is one important source of the *golden age* documented by Waldfoegel (2017). For regulatory interventions and policy debates more broadly, it is important to understand how these incentives depend on the amount of available data and on the digital monopoly's price.

To shed light on these questions, we now stipulate that given product quality investment  $I$  by the monopoly, the consumer's value  $v$  is drawn from the distribution  $F(v, I)$  with support  $[\underline{v}, \bar{v}]$ , with increases in  $I$  inducing a first-order stochastic shift in the distribution, i.e.,  $\partial F(v, I)/\partial I \leq 0$ , with a strict inequality for an open set of values of  $v$ . Here, we take the monopoly's price to be fixed at some value  $p$  and ask how the marginal incentives vary with  $n$  and  $p$ .<sup>15</sup>

The monopoly's profit given  $p$ ,  $n$ , and  $I$  is now

$$\Pi(p, n, I) = (p - c)(1 - F^n(p, I)). \quad (6)$$

The first derivative with respect to  $I$  is

$$\frac{\partial \Pi(p, n, I)}{\partial I} = -(p - c)nF^{n-1}(p, I) \frac{\partial F(p, I)}{\partial I} \geq 0,$$

where the inequality is strict if  $\partial F(p, I)/\partial I < 0$ . To see how this marginal incentive varies with  $n$ , note that the cross partial derivative is

$$\frac{\partial^2 \Pi(p, n, I)}{\partial I \partial n} = -(p - c)nF^{n-1}(p, I) \frac{\partial F(p, I)}{\partial I} [1/n + \ln F(p, I)],$$

which is positive if and only if

$$-\ln F(p, I) < \frac{1}{n}.$$

Thus, the marginal incentives for product quality investment increase in  $n$  (i.e., data) if and only if the monopoly's price is high enough.

The intuition is reasonably simple even though still somewhat technical. Inspection of (6) reveals that the cross partial of  $\Pi(p, n, I)$  with respect to  $n$  and  $I$  is positive if and only if the cross partial of  $F^n(p, I)$  with respect to  $n$  and  $I$  is negative because  $p - c$  is positive. Because  $\partial F^n(p, I)/\partial I = nF^{n-1}(p, I)\partial F/\partial I$  with  $\partial F/\partial I < 0$ , this requires the derivative of  $nF^{n-1}(p, I)$

<sup>12</sup> To see this, note that for the uniform case,  $\frac{\partial^2 \Pi_n(p)}{\partial n \partial p} = -p^n(1 + (1+n)\ln p)$ , which is positive if  $p < e^{\frac{1}{1+n}}$ , which, using (2), holds for all  $p < p_n$ .

<sup>13</sup> The result extends straightforwardly beyond the uniform distribution to any distribution  $F(v) = v^k$  on  $[0,1]$  with  $k > 0$ . To see this, let  $\hat{n}$  denote the parameter that measures data for a given  $k$ . Then, by choosing  $n = \hat{n}k$ , the analysis is the same as for the uniform distribution.

<sup>14</sup> Our results can be reinterpreted to speak to the issue of "fake news" on social media (Allcott and Gentzkow, 2017). Suppose that a consumer has value  $v = 1$  for news, which is common knowledge, but incurs a cost to check whether the news is "fake news," and that cost is the consumer's private information. Suppose that the digital monopoly can take steps to fact check content, which reduces the need for the consumer to check for fake news, resulting in a first-order stochastically dominated shift in the consumer's cost distribution. The digital monopoly's "price" may take the form of advertisements or subscription fees. Then if  $c$  is drawn from  $G_n$ , the consumer's value  $x = v - c$  has distribution  $F_n(x) \equiv 1 - G_n(v - x)$ . So, if  $G_n(x) = 1 - (1 - G(x))^n$ , then  $F_n(x) = (1 - G(v - x))^n$ , and our results apply. The consumer prefers some fact checking, but not extreme fact checking (in the extreme, the price is 1 and the consumer has no surplus). An unregulated social media platform has stronger incentives to fact check than does a regulated one.

<sup>15</sup> The first partial derivatives are not affected if  $p$  is chosen optimally given  $n$  and  $I$  because of the envelope theorem. However, the cross-partials become unwieldy when  $p$  is treated as an endogenous variable.

with respect to  $n$  to be positive. For a given  $n$ , this can only be the case if  $F(p, I)$ , and hence  $p$ , is sufficiently large. We summarize this with the following proposition:

**Proposition 2.**  $\frac{\partial^2 \Pi(p, n, I)}{\partial I \partial n} > 0$  if and only if  $p$  is sufficiently large.

Proposition 2 implies that for sufficiently high-priced products, allowing a digital monopoly to have more information on consumers could induce the digital monopoly to invest in increased product quality. Put differently, the proposition means that price regulation has the potential to eliminate the complementarity between product quality investments and investments in data analytics. This highlights a potential drawback of price regulation. It could eliminate the positive feedback between investments in high product quality and data analytics.

### 5. Extension to duopoly

Although the paper is concerned with digital monopolies, it is also of interest to see whether competition between digital content providers, wherever possible, would the remedy negative effects of data collection on consumers that are possible when data is used for both matching and pricing. In light of increasing returns to scale in data, competition may not be the right, and certainly not the first-best, answer. Nevertheless, it is worth exploring.

To this end, we now stipulate that there are two firms, labelled  $A$  and  $B$ , each  $i \in \{A, B\}$ , “producing” a good with value  $v_i$  for the consumer drawn independently from a distribution  $F_{i,n}$  with support  $[0,1]$  and density  $f_{i,n}$ . Initially, we assume that  $n$  is a common level of data available to the firms. This corresponds to a setting in which data is fully transferable between the digital duopolists. Then we consider alternatives in which data is not transferable and in which there is competition from a non-digital fringe, which is defined as a firm that, because it does not have access to data, does not provide better matching as  $n$  increases. We conclude with a comparison of the data transferability regimes.

The consumer has single-unit demand, that is, the consumer buys at most one unit of the good. The consumer’s utility when buying from firm  $i$  at price  $p_i$  is  $v_i - p_i$ . Given that the value of consumer’s outside option is assumed to be 0, the consumer’s maximized utility is  $\max\{v_A - p_A, v_B - p_B, 0\}$ .

#### Digital duopoly with fully transferable data

Given prices  $p_A$  and  $p_B$  and given fully transferable data  $n$ , the expected demand for firm  $A$  is

$$D_{A,n}(p_A, p_B) = \int_{p_A}^1 F_{B,n}(v_A + p_B - p_A) f_{A,n}(v_A) dv_A.$$

Similarly, firm  $B$ ’s expected demand is

$$D_{B,n}(p_B, p_A) = \int_{p_B}^1 F_{A,n}(v_B + p_A - p_B) f_{B,n}(v_B) dv_B.$$

Further, given  $p_A$  and  $p_B$ , consumer surplus is given by

$$CS_n(p_A, p_B) = \int_{p_A}^1 (v_A - p_A) F_{B,n}(v_A + p_B - p_A) f_{A,n}(v_A) dv_A + \int_{p_B}^1 (v_B - p_B) F_{A,n}(v_B + p_A - p_B) f_{B,n}(v_B) dv_B.$$

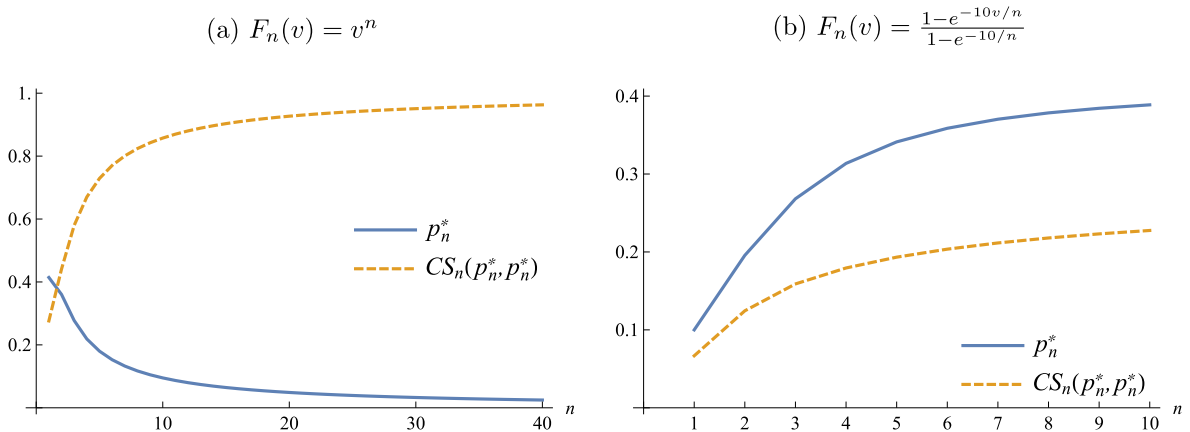
Assuming marginal costs of zero and that first-order conditions are sufficient, for a given  $n$ , the prices  $p_{A,n}^*$  and  $p_{B,n}^*$  constitute a Nash equilibrium if  $p_{i,n}^* = \frac{D_{i,n}(p_{i,n}^*, p_{j,n}^*)}{-\frac{\partial D_{i,n}(p_{i,n}^*, p_{j,n}^*)}{\partial p_{i,n}}}$  for  $i, j \in \{A, B\}$  with  $i \neq j$ . Focusing on symmetric equilibria, where

$F_{A,n} = F_{B,n} = F_n$ , we have

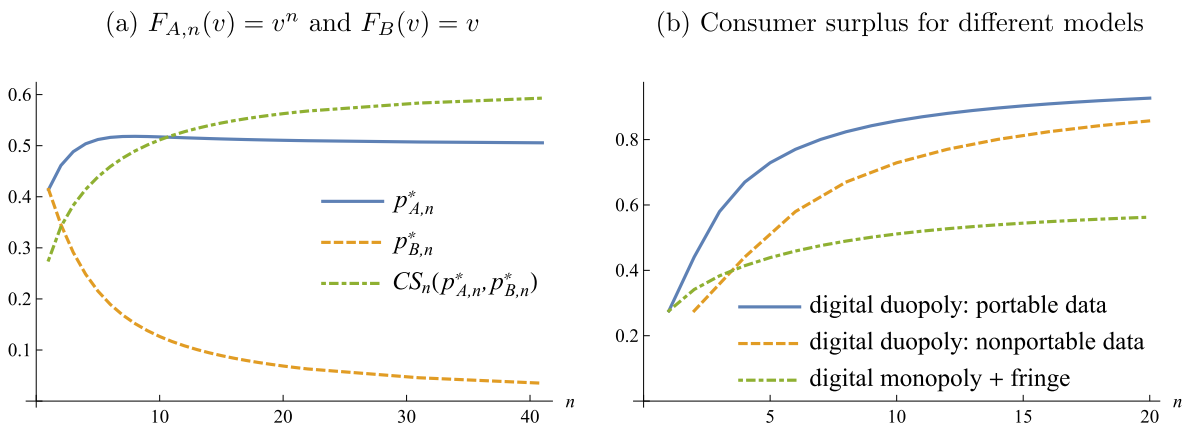
$$p_n^* = \frac{\int_{p_n^*}^1 F_n(x) f_n(x) dx}{F_n(p_n^*) f_n(p_n^*) + \int_{p_n^*}^1 f_n^2(x) dx}. \tag{7}$$

For the parameterization  $F_n(v) = v^n$ , we obtain equilibrium prices  $p_n^*$  that converge to zero as  $n$  goes to infinity,<sup>16</sup> which implies that in the duopoly case, consumer surplus is maximized when the competing firms have unlimited data on consumers, as illustrated in Fig. 2(a). This contrasts with our results for the monopoly case, shown in Fig. 1, where  $p_n^*$  goes to one as  $n$  goes to infinity and consumer surplus is maximized at a finite level of data. For other parameterizations, one obtains different results. For example, Fig. 2(b) considers the parameterization  $F_n(v) = \frac{1 - e^{-10v/n}}{1 - e^{-10}}$  and shows that the price increases in the amount of data, consistent with monopoly results shown in Fig. 1(a). However, in this duopoly example, even though the price increases with data, consumer surplus also increases with data. There are two competing effects: as

<sup>16</sup> For  $F_n(v) = v^n$ , (7) implies that  $p_n^* = \frac{(1 - (p_n^*)^{2n})^{(2 - \frac{1}{n})}}{2((n-1)(p_n^*)^{2n-1} + n)}$ . Taking the limit as  $n$  goes to  $\infty$  and using  $p_n^* \in (0, 1)$  for all finite  $n$ , we have  $\lim_{n \rightarrow \infty} p_n^* = 0$ .



**Fig. 2.** Prices and consumer surplus as a function of the level of data for a duopoly model with symmetric firms and parameterized distributions as indicated.



**Fig. 3.** Panel (a): Prices and consumer surplus as a function of the level of data for a digital firm competing with a non-digital firm with parameterized distributions as indicated. Panel (b): Consumer surplus for three models. Digital duopoly: portable data assumes two digital firms each with  $F_n(v) = v^n$ . Digital duopoly: nonportable data assumes two digital firms each with  $F_n(v) = v^{n/2}$ . Digital monopoly + fringe assumes one digital firm with  $F_{A,n}(v) = v^n$  and one non-digital firm with  $F_B(v) = v$ .

data increase, the distribution of values improves, which all else equal increases consumer surplus, but the price increases, which decreases consumer surplus. As shown in this example, the improvement of the distribution can dominate, so that the overall effect of increased data is to increase consumer surplus.

From this analysis emerges an important difference between the increasing returns to scale in traditional brick-and-mortar industries that derive from the production technology and the increasing returns to scale due to data in the digital age. While marginal cost pricing does not allow firms to break even in the traditional industries (and competition may involve the duplication of fixed investment costs), marginal cost pricing, or more generally, competition may be more viable in digital industries, in particular when data is transferable across providers. This suggests that competition, and the associated downward pressure on prices, may have a role to play in mitigating potential harm to consumers from firms having high levels of data and may obviate the need for price regulation. To explore this further, we now consider the case of a digital monopoly with a non-digital fringe.

#### Digital monopoly competing with a non-digital fringe

Suppose that firm A is a digital firm with data  $n$ , but firm B does not use (or does not have access to) data. Specifically, suppose that  $F_{A,n}(v) = v^n$  and that  $F_B(v) = v$ . Then we find that as  $n$  increases, the digital firm's price initially increases and then decreases, while the non-digital firm's price decreases with  $n$  (see Fig. 3(a)). As  $n$  increases, the digital firm benefits from the improvement in the match value for its product, and captures some of that value by increasing its price, but the effects of competition eventually limit those price increases. The non-digital firm responds to the increasingly effective competition from the digital firm by decreasing its price. As shown in Fig. 3(a), for the parameterization that we consider, consumer surplus increases with data.



### Comparison

To conclude this extension, we briefly consider the impact of data portability by contrasting three scenarios, illustrated in Fig. 3(b). The first one corresponds to the case of fully transferable data, so that digital duopoly firms have access to the same customer data. The second scenario is the one just analyzed, in which a digital monopoly faces a non-digital fringe. It is illustrated in Fig. 3(a). To these we add a third scenario in which data is not portable so that when the level of data is  $n$ , each of the digital duopoly firms operates only with data  $n/2$ . As shown in Fig. 3(b), consumer surplus is greater under a digital duopoly with fully portable data than under either competition from a non-digital fringe or competition from a digital firm but without data portability. This suggests that data portability may be a relevant policy issue when it comes to consumer surplus.

Interestingly, the results displayed in Fig. 3(a) suggest that even competition with a “brick-and-mortar” fringe firm that makes no use of data collection may be enough to effectively induce the digital firm to use data only to improve matching without substantially increasing prices (in fact, for  $n$  large enough, its price decreases with  $n$  in this specification). Thus, even a moderate level of competition may be enough to prevent the digital monopoly from excessive price discrimination via personalized pricing. In turn, this suggests not only a possible explanation for why, empirically, one observes little personalized pricing, but also draws attention to the possibility that maintaining outside competition may be more important than privacy protection to advance consumer surplus.

## 6. Discussion

This paper contributes to current policy debates regarding online privacy and the pricing of digital goods,<sup>17</sup> and to the economics literature related to that. In this section, we first review the related literature and then provide a discussion of policy-relevant issues in the context of our paper.

### 6.1. Related literature

The increasing returns in market making mentioned in the introduction that are at the heart of digital monopolies relate to Williamson's puzzle of selective intervention (Williamson, 1985), according to which there would be no limits to firm size because an integrated firm could always replicate what standalone firms do, and sometimes do better.<sup>18</sup> In this paper, we relate this driving force of digital monopolies to monopoly pricing, including price discrimination and matching.<sup>19</sup> We show that the desirability of privacy protection for consumers, which has recently been studied in a variety of contexts (Shelanski, 2013; Acquisti et al., 2016; Jullien et al., 2018; Goldfarb and Tucker, 2019), critically depends on what data is used for—matching only or matching and pricing.<sup>20</sup>

Of course, the analysis of the effects of price discrimination on consumer and social surplus has a long tradition in economics, pre-dating the digital economy; see, e.g., Hart and Tirole (1988). Subsequent literature on consumer privacy and regulation recognizes that consumers may have an incentive to withhold information (e.g., Villas-Boas, 1999; Fudenberg and Tirole, 2000; Villas-Boas, 2004; Taylor, 2004; Acquisti and Varian, 2005; Fudenberg and Villas-Boas, 2006; Hermalin and Katz, 2006; Conitzer et al., 2012; Acquisti et al., 2016; Goldfarb and Tucker, 2019), and that restrictions on the ability of a firm to price discriminate may harm consumers that do not trade as a result (e.g., Taylor and Wagman, 2014; Kim and Wagman, 2015). For discussion of issues related to policies and regulation for the digital economy, see Shelanski (2013); Waldfogel (2017); Jullien et al. (2018). Goldfarb and Tucker (2012) analyze the empirical link between privacy regulation and investment.

Acquisti et al. (2016) provide extensive discussion of a wide variety of models of privacy, and we refer the reader to that review for details. Two categories of models connect closely with our work. First, there are models in which a consumer or a data provider has private information regarding the consumer's willingness to pay that, if available to sellers, would facilitate their use of personalized prices. Second, there are models of matching between suppliers and consumers with heterogeneous tastes for those suppliers and/or their products in which consumer data has the potential to facilitate (and privacy to frustrate) matching. From a modelling perspective, what sets our paper apart from the more “standard” approaches on privacy as surveyed by Acquisti et al. (2016), is that in our incomplete information setting, consumers have an innate desire to protect their privacy—it is precisely because they are privately informed about their values that they derive any surplus.

The main contribution of our paper is that we identify the distinction between what digital firms use data for—to improve matching only or to improve matching and adjust pricing—as fundamental to understanding critical tradeoffs in the

<sup>17</sup> Since 2016, the U.S. Federal Trade Commission (FTC) has held an annual workshop called PrivacyCon where scholars discuss recent research on privacy and security issues (<https://www.ftc.gov/news-events/events-calendar/privacycon-2020>). See also FTC (2012) and, for FTC efforts related to children's online privacy, see <https://www.ftc.gov/tips-advice/business-center/privacy-and-security/children%27s-privacy>. See also “Big Data and Differential Pricing,” Office of the President of the United States, February 2015, [https://obamawhitehouse.archives.gov/sites/default/files/whitehouse\\_files/docs/Big\\_Data\\_Report\\_Nonembargo\\_v2.pdf](https://obamawhitehouse.archives.gov/sites/default/files/whitehouse_files/docs/Big_Data_Report_Nonembargo_v2.pdf).

<sup>18</sup> See Crémer (2010) for discussion and Loertscher and Riordan (2019) for partial resolution in an incomplete contracting environment.

<sup>19</sup> Our work is distinct from the literature on information design (for an overview, see Bergemann and Morris, 2019) in that in our model (i) consumers do not behave as information designers and (ii) additional data improves the seller's prior rather than changing its posterior.

<sup>20</sup> This distinction is largely absent from concurrent policy debates (see, e.g., Crémer et al., 2019, Chapter 5).

digital age. We do so in a stripped-down model in which this distinction and the associated tradeoff is most transparent. While the prior literature has touched on related issues and similar themes, to the best of our knowledge, this key distinction has nowhere been made explicit. For example, [Belleflamme and Vergote \(2016\)](#), [Shy and Stenbacka \(2016\)](#), and [Montes et al. \(2019\)](#) consider price discrimination based on data without allowing for the possibility that data can be used to improve matching values for consumers. Conversely, [Bergemann and Bonatti \(2015\)](#) and [Lefouili and Toh \(2019\)](#) consider the matching value of data, but do not allow data to be used for pricing to consumers. In [de Cornière and de Nijs \(2016\)](#), data are used either for both matching and pricing or for neither, so the distinction between matching only versus matching and pricing does not arise. The distinction between the use of data for matching only or for matching and pricing is, albeit somewhat implicitly, at the heart of the analysis of [Ichihashi \(2020\)](#), who provides conditions under which a seller prefers not to use information for pricing to incentivize information disclosure by consumers who choose their disclosure rule prior to the realization of their types. In a sense, [Ichihashi](#) considers a version of the problem that we analyze, augmented by an *ex ante* stage in which the consumer chooses how much information to reveal and the seller can commit not to price based on this information.

## 6.2. Policy issues

We conclude this section with a brief discussion of pertinent policy issues that our model and analysis shed light on.

### 6.2.1. Digital monopoly intermediary

Our analysis presumes that the digital monopoly is both a data collector and a seller of products or content. Of course, this is a better description of some digital monopolies—such as Netflix or Amazon—than of others, like Google, that are *digital monopoly intermediaries* insofar as their revenue accrues primarily from selling data to third parties (in the case of Google, to advertisers). We now briefly to what extent our main insights carry over to such digital monopoly intermediaries.

For that purpose, we first consider a pure digital intermediary who has no use for data itself. To fix ideas, we assume that the data is sold to advertisers and that advertisements are informative, and the more so the better they are matched to a consumer's preferences. Under these assumptions, the digital monopoly intermediary is merely a reinterpretation of our model. The key distinction is still whether data are only used to improve matching or to “improve” both matching and pricing. For the social and consumer surplus implications, it is immaterial who does the pricing—the digital monopoly or third parties. An interesting and to the best of our knowledge open question for future research is whether the digital monopoly intermediary is able to capture the additional surplus that its data generate via an appropriately designed sale mechanism. The match value increases with data, which means that the willingness to pay of an advertiser with the “right” product increases, but the number of advertisers who are willing to compete for a given consumer decreases.<sup>21</sup> The only nuances arise when considering the incentives to invest in data analytics, which depend on whether the digital monopoly intermediary is able to capture the additional surplus that its data create in the same way as a digital monopoly. As mentioned in the previous footnote, this is an open question for future research.

Alternatively, one can consider a “hybrid” digital monopoly that can use its data both directly to target its customers and indirectly by selling it to third parties. If the direct effect on consumer surplus from third parties' having access to this data is negligible (which may be a sensible assumption if this consumer surplus effect is sufficiently uncertain and can be either positive), then only the indirect effect matters, which is that with third parties the digital monopoly has stronger incentives to invest in data analytics. Consequently, the key question is again whether the digital monopoly uses the data to improve matching only, in which case the consumers should welcome data sales to third parties whose consumer surplus effect is negligible, or to “improve” matching and pricing, in which case the consumer surplus effects of such data sales to third parties could go either way. (And, of course, if the direct third-party effects are not negligible for consumer surplus, then these have to be spelled out and accounted for as well.)

### 6.2.2. Data collection that is costly for the consumer

In many settings, consumers need incur no additional cost beyond using the products or services of a digital monopoly in order to generate the data that the digital monopoly then uses for matching and/or pricing. However, in some settings, such as the Ziprecruiter example mentioned in the introduction, a consumer may need to incur time or other costs in order to provide data to the digital monopoly. Here we briefly consider the implications of this additional data-related cost on consumers.

To account for data collection costs imposed on consumers, one can amend the model by subtracting from the consumer's surplus a cost  $c(n)$ , which is assumed to be positive, increasing, and convex in  $n$ . Further, instead of viewing the

<sup>21</sup> Although, for example, the analysis in [de Cornière and de Nijs \(2016\)](#) suggests that increasing information revelation unambiguously improves auction outcomes for the intermediary, this conclusion is driven by the assumption in their setting that revealing more information corresponds to having more firms compete for the consumer. However, one can easily and plausibly come up with setups in which this is not the case. Assume, for example, that a consumer likes with equal probability product  $A$  or product  $B$ . Firm  $i \in \{A, B\}$ , whose product the consumer likes, obtains a profit of  $\Pi > 0$  while the other one obtains a profit of 0. In an efficient auction in which the risk-neutral firms do not know what product the consumer likes, each is willing to bid  $\Pi/2$ , which is thus the intermediary's revenue. With complete information revelation, the firms know which product the consumer likes, implying that one of them is willing to bid  $\Pi$  and the other one 0, generating zero revenue for the intermediary.

level of data (or data analytics) as a choice variable of the digital monopoly, as we do in Section 4.1, one could view the level of data as a choice variable of the consumer. In that case, assuming that data are used for both matching and pricing, data would be chosen by the consumer to solve

$$\max_n CS_n(p_n) - c(n),$$

where  $p_n$  is the price chosen by the digital monopoly given data  $n$ . The addition of a data cost to the consumer reinforces the result in Corollary 1 that a finite level of data is optimal for the consumer and changes our result regarding social surplus in that social surplus is no longer everywhere increasing in the level of data.

Alternatively, if data are used only for matching and the price is fixed at  $p$ , then data would be chosen by the consumer to solve  $\max_n CS_n(p) - c(n)$ . The consumer's optimal level of data, would, of course, be diminished in light of the added cost term, and it would no longer be the case that more data is unambiguously good for the consumer.

### 6.2.3. Property rights

Beyond online search engines, which are a prime example of digital monopolies that aim at improving match values, well-documented benefits to consumers and society arise exactly because consumers do *not* protect their privacy. Waldfoegel (2017) calls the current era a *golden age* of music, movies, books, and television, documenting how digitalization has led to this new era.

While everything may look new in the digital age, our analysis suggests that digital monopolies parallel their "natural" counterparts and that policy tools like price regulation that were useful for balancing tradeoffs between producer and consumer surplus may remain valid instruments in the digital age. Other parallels exist and can be used to inform policy. For example, the data that users generate through their online behaviour has a public goods component in that the information gleaned from it can be used to improve other consumers' match values. This problem is similar to the classic public health problem of vaccination, where major benefits from an individual's vaccination accrue to society as a whole rather than the individual who obtains the vaccination.

Concurrent policy debates often evolve along the lines that consumers should be given the property rights to their data, sometimes accompanied by expressions of frustration that consumers do not care (enough) about protecting their data. While this proposition has appeal to economists, and maybe to larger audiences as well, it deserves discussion and context.

First, the vaccination problem provides a useful benchmark. The typical health policy prescription is *not* that every one should be free to choose whether they (or their offspring) obtain vaccination against contagious diseases. Much to the contrary, in many instances policy mandates individuals to take the individually costly action of being vaccinated if the benefits to society are deemed to sufficiently outweigh these costs.

Second, although our analysis in Section 5 shows that consumers benefit from owning their data when there is data portability and competition between digital firms, in many digital settings there are challenges for what it means for consumers to own their data. For example, while the personal data that are shared with, say, a bank when applying for a loan might be well defined, it is less clear what data a consumer might have generated and then have ownership rights over following, for example, an hour spent listening to a music streaming service. If a sequence of songs is streamed in response to an initial seed provided by the consumer, what are the data that then belong to the service and what to the consumer? Further, numerous practical challenges exist for data portability, including the lack of consistent data formats, proprietary databases, and barriers to international data transfers (Engels, 2016).

Third, part of the appeal of the proposition that consumers should have rights over their data stems from the Coase Theorem (Coase, 1960), according to which, if transaction costs are negligible, the initial allocation of property rights only affects the division of social surplus—that is, how the pie is shared, not the size of the pie. Accordingly, absent transaction costs, giving consumers ownership of their data might well shift the balance between consumer and producer surplus towards consumers. The validity of the argument depends on whether transaction costs are negligible. Claiming that they are negligible raises the question why these platforms emerge in the first place. Of course, there is no single model or analysis that captures the rich nature of problems in the digital age. However, the lesson on optimal property rights that emerges from the mechanism design literature, where the source of transaction costs is the private information about values and costs, is *not* that agents on one side of the market (say, consumers) should be given all the property rights.<sup>22</sup> Indeed, the gist of the celebrated impossibility theorem of Myerson and Satterthwaite (1983), with its precursor in Vickrey (1961), is that with extreme ownership structures, efficient incentive compatible and individually rational trade is impossible without running a deficit. In contrast, efficient trade may be possible with shared ownership structures (Cramton et al., 1987; Neeman, 1999; Che, 2006; Figueroa and Skreta, 2012).<sup>23</sup> Obviously, this does not prove that extreme ownership is always suboptimal, but it certainly provides a cautionary tale against the proposition that it is optimal.<sup>24</sup>

<sup>22</sup> For example, in the case of music streaming, where "big data" has successfully played a key role in generating value by matching consumers to music, certain payments to artists and composers are fixed by statute.

<sup>23</sup> Delacrétaz et al. (2019) provide a generalization of the impossibility theorem with extreme ownership.

<sup>24</sup> Loertscher and Wasser (2019) provide conditions for extreme ownership to be optimal when the objective involves profit motives.

## 7. Conclusions

Like natural monopolies, digital monopolies arise because of increasing returns to scale. Exploitation of these increasing returns increases social surplus but, without limits on the use of data for pricing, may reduce consumer surplus. While privacy protection reduces, and in the limit eliminates, the market power of digital monopolies, privacy protection also reduces, and in the limit eliminates, the social surplus created by digital monopolies. In particular, in our setting, consumer harm arises only by the combination of improved match values due to privacy reduction *and* more aggressive pricing by the monopoly. For a fixed price, the consumer always benefits from the improved matches that come with a reduction in privacy. Based on this, we conclude that competition policy should aim at protecting consumers' information rents rather than their privacy. While privacy protection is a possible means to achieve this end, our analysis shows that other, more traditional tools, such as regulating prices, may be preferable from both a consumer and social surplus perspective.

Our paper brings to light new questions regarding the form of optimal price regulation in the digital age. Even within our baseline model without investment, traditional approaches such as Ramsey pricing may not work satisfactorily because the elasticity of demand is endogenous to the amount of data available to the monopoly. In response, we propose price regulation that keeps the ratio of producer to consumer surplus fixed as social surplus grows due to increasing returns to scale, raising the practical question as to how it can be implemented. In richer models that account explicitly for the nature of the data available to the monopoly, the question arises of what prices may depend on. For example, if prices vary with consumer location, they might sensibly be required to only depend on anonymous data such as distance from the closest distribution center and be prohibited from depending on the consumer's delivery address itself. These, and related, questions are excellent problems for future research.

### Credit statement

Simon Loertscher: all aspects

Leslie M. Marx: all aspects

### Appendix A. Appendix: Proofs

*Proof of Lemma 1.* Observe that (dropping the argument  $v$ )

$$\left[ \frac{1 - F_n}{f_n} \right]' = \frac{nF^{n-2}[-f^2F^n - (1 - F^n)f'F - (n - 1)f^2]}{(nfF^{n-1})^2}.$$

This is nonpositive if and only if

$$-f^2F^n - (1 - F^n)f'F - (n - 1)f^2 \leq 0. \quad (8)$$

Given our assumption that  $(1 - F)/f$  is nonincreasing, the smallest value that  $f'$  can take without violating the assumed monotonicity of  $(1 - F)/f$  is  $-f^2/(1 - F)$ . Consequently, we have

$$\begin{aligned} -f^2F^n - (1 - F^n)f'F - (n - 1)f^2 &\leq -f^2F^n + \frac{(1 - F^n)f^2F}{1 - F} - (n - 1)f^2 \\ &= \frac{f^2}{1 - F}[-F^n(1 - F) + (1 - F^n)F - (n - 1)(1 - F)] \\ &= \frac{f^2}{1 - F}[1 - F^n - n(1 - F)]. \end{aligned}$$

It follows that (8) holds if the function  $q(F) \equiv 1 - F^n - n(1 - F)$  defined for  $F \in [0, 1]$  is not more than 0. At  $F = 1$ ,  $q(1) = 0$ . Moreover,  $q'(F) = n(1 - F^{n-1})$  is nonnegative using our assumption that  $n \geq 1$  (and positive for all  $n > 1$  and  $F < 1$ ), proving that  $q(F) \leq 0$  for all  $F \in [0, 1]$ . This completes the proof that (8) holds.

Turning to the next part of the lemma,  $\frac{1 - F_n(v)}{f_n(v)}$  is increasing in  $n$  for all  $v < \bar{v}$  if and only if the function  $Q(n) \equiv \frac{1 - F^n}{nF^{n-1}}$  is increasing in  $n$  for all  $F \in [0, 1)$ . Differentiating, we have

$$Q'(n) = \frac{-\ln(F)F^n nF^{n-1} - (1 - F^n)F^{n-1} - (1 - F^n)\ln(F)nF^{n-1}}{(nF^{n-1})^2},$$

which is greater than 0 if and only if  $-n\ln(F) - (1 - F^n) > 0$ . For  $F = 1$ ,  $-n\ln(F) - (1 - F^n) = 0$ . We now show that  $-n\ln(F) - (1 - F^n)$  is decreasing in  $F$ . Taking the derivative, we obtain  $-n/F + nF^{n-1}$ , which is less than 0 for all  $F < 1$ . This completes the proof that  $\frac{1 - F_n(v)}{f_n(v)}$  is increasing in  $n$ .

To show that  $\frac{1 - F_n(v)}{f_n(v)}$  is unbounded in  $n$  for  $F(v) < 1$ , we first rewrite it as

$$\frac{1 - F_n(v)}{f_n(v)} = \frac{1}{f(v)} \left[ \frac{1}{nF^{n-1}(v)} - \frac{F(v)}{n} \right].$$

Because  $\lim_{n \rightarrow \infty} nF^{n-1}(v) = 0$  for  $v \leq \bar{v}$  (which is equivalent to  $F(v) \leq 1$ ), it follows that for any  $v \leq \bar{v}$

$$\lim_{n \rightarrow \infty} \frac{1}{nF^{n-1}(v)} = \infty.$$

Because  $\lim_{n \rightarrow \infty} \frac{F(v)}{n} = 0$ , for any  $v < \bar{v}$ ,

$$\lim_{n \rightarrow \infty} \frac{1 - F_n(v)}{f_n(v)} = \infty,$$

which completes the proof.  $\square$

## References

- Acquisti, A., Taylor, C., Wagman, L., 2016. The economics of privacy. *J. Econ. Lit.* 54, 442–492.
- Acquisti, A., Varian, H.R., 2005. Conditioning prices on purchase history. *Marketing Science* 24 (3), 367–381.
- Allcott, H., Gentzkow, M., 2017. Social media and fake news in the 2016 election. *Journal of Economic Perspectives* 31 (2), 1–28.
- Belleflamme, P., Vergote, W., 2016. Monopoly price discrimination and privacy: the hidden cost of hiding. *Econ. Lett.* 149, 141–144.
- Bergemann, D., Bonatti, A., 2015. Selling cookies. *American Economic Journal: Microeconomics* 7 (3), 259–294.
- Bergemann, D., Morris, S., 2019. Information design: a unified perspective. *J. Econ. Lit.* 57 (1), 44–95.
- Bulow, J., Roberts, J., 1989. The simple economics of optimal auctions. *Journal of Political Economy* 97 (5), 1060–90.
- Che, Y.-K., 2006. Beyond the coasian irrelevance: Asymmetric information. Unpublished lecture notes.
- Coase, R., 1960. The problem of social cost. *Journal of Law and Economics* 3, 1–44.
- Conitzer, V., Taylor, C.R., Wagman, L., 2012. Hide and seek: costly consumer privacy in a market with repeat purchases. *Marketing Science* 31, 277–292.
- de Cornière, A., de Nijis, R., 2016. Online advertising and privacy. *RAND Journal of Economics* 47 (1), 48–72.
- Cramton, P., Gibbons, R., Klemperer, P., 1987. Dissolving a partnership efficiently. *Econometrica* 55 (3), 615–632.
- Crémer, J., 2010. Solving the “selective intervention” puzzle. *Revue d’Économie Industrielle* 129–130 (1–2), 43–56.
- Crémer, J., de Montjoye, Y.-A., Schweitzer, H., 2019. Competition policy for the digital era. EC Report.
- Delacrétaz, D., Loertscher, S., Marx, L., Wilkening, T., 2019. Two-sided allocation problems, decomposability, and the impossibility of efficient trade. *J. Econ. Theory* 179, 416–454.
- Dubé, J.-P., Misra, S., 2017. Scalable price targeting. University of Chicago.
- Engels, B., 2016. Data portability among online platforms. *Internet Policy Review* 5 (2), 1–17.
- Federal Trade Commission, 2012. Protecting consumer privacy in an era of rapid change: recommendations for businesses and policymakers. Federal Trade Commission Report, March, available at [www.ftc.gov/os/2012/03/120326privacyreport.pdf](http://www.ftc.gov/os/2012/03/120326privacyreport.pdf).
- Figueroa, N., Skreta, V., 2012. Asymmetric partnerships. *Econ. Lett.* 115 (2), 268–271.
- Fudenberg, D., Tirole, J., 2000. Customer poaching and brand switching. *RAND Journal of Economics* 31 (4), 634–657.
- Fudenberg, D., Villas-Boas, J.M., 2006. Behavior-based Price Discrimination and Customer Recognition. In: Hendershott, T. (Ed.), *Handbook on Economics and Information Systems*, Vol. 1. Emerald Group Publishing Limited, Bingley, U.K., pp. 377–436.
- Goldfarb, A., Tucker, C., 2012. Privacy and Innovation. In: Lerner, J., Stern, S. (Eds.), *Innovation Policy and the Economy*, Vol. 12. University of Chicago Press, pp. 65–90.
- Goldfarb, A., Tucker, C., 2019. Digital economics. *J. Econ. Lit.* 57 (1), 3–43.
- Hart, O.D., Tirole, J., 1988. Contract renegotiation and coasian dynamics. *Review of Economic Studies* 55 (4), 509–540.
- Hermalin, B.E., Katz, M.L., 2006. Privacy, property rights and efficiency: the economics of privacy as secrecy. *Quantitative Marketing and Economics* 4 (3), 209–239.
- Ichihashi, S., 2020. Online privacy and information disclosure by consumers. *American Economic Review* 110 (2), 569–595.
- Jullien, B., Lefouili, Y., Riordan, M. H., 2018. Privacy protection and consumer retention. Toulouse School of Economics Working Paper No. TSE - 947.
- Kim, J.-H., Wagman, L., 2015. Screening incentives and privacy protection in financial markets: a theoretical and empirical analysis. *RAND Journal of Economics* 46 (1), 1–22.
- Koblin, J., 2017. Netflix says it will spend up to \$8 billion on content next year. *New York Times*, October 16, 2017.
- Lefouili, Y., Toh, Y. L., 2019. Privacy regulation and quality investment. TSE Working Paper No. 17–795.
- Loertscher, S., Riordan, M.H., 2019. Make and buy: outsourcing, vertical integration, and cost reduction. *American Economic Journal: Microeconomics* 11 (1), 105–123.
- Loertscher, S., Wasser, C., 2019. Optimal structure and dissolution of partnerships. *Theoretical Economics* 14, 1063–1114.
- Malamud, S., Rostek, M., 2017. Decentralized exchange. *American Economic Review* 107 (11), 3320–3362.
- Montes, R., Sand-Zantman, W., Valletti, T., 2019. The value of personal information in online markets with endogenous privacy. *Manage. Sci.* 65 (3), 1342–1362.
- Myerson, R.B., Satterthwaite, M.A., 1983. Efficient mechanisms for bilateral trading. *J. Econ. Theory* 29 (2), 265–281.
- Neeman, Z., 1999. Property rights and efficiency of voluntary bargaining under asymmetric information. *Review of Economic Studies* 66 (3), 679–691.
- Shelanski, H.A., 2013. Information, innovation, and competition policy for the internet. *Univ. PA. Law Rev.* 161, 1663–1705.
- Shy, O., Stenbacka, R., 2016. Customer privacy and competition. *Journal of Economics & Management Strategy* 25 (3), 539–562.
- Taylor, C.R., 2004. Consumer privacy and the market for customer information. *RAND Journal of Economics* 35 (4), 631–650.
- Taylor, C.R., Wagman, L., 2014. Consumer privacy in oligopolistic markets: winners, losers. *Int. J. Ind. Organiz.* 34, 80–84.
- Vickrey, W., 1961. Counterspeculation, auction, and competitive sealed tenders. *Journal of Finance* 16, 8–37.
- Villas-Boas, J.M., 1999. Dynamic competition with customer recognition. *RAND Journal of Economics* 30 (4), 604–631.
- Villas-Boas, J.M., 2004. Price cycles in markets with customer recognition. *RAND Journal of Economics* 35 (3), 486–501.
- Waldfogel, J., 2017. How digitization has created a golden age of music, movies, books, and television. *Journal of Economic Perspectives* 31 (3), 195–214.
- Williamson, O.E., 1985. *The economic institutions of capitalism*. New York: Free Press.