

Digital Monopolies: Privacy Protection or Price Regulation?*

Simon Loertscher[†] Leslie M. Marx[‡]

October 28, 2019

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.

Keywords: natural monopoly, privacy concerns, big data, transparency, Ramsey pricing

JEL-Classification: C72, D72

*We thank David Byrne, Peter Taylor, participants at the 2019 Workshop on Digital Economy and Industrial Organisation at Monash University and the 2019 Competition Economists Network Meeting for useful comments and fruitful discussions. We gratefully acknowledge support from the Samuel and June Hordern Endowment and a University of Melbourne Faculty of Business & Economics Eminent Research Scholar Grant. Bing Liu provided excellent research assistance.

[†]Department of Economics, Level 4, FBE Building, 111 Barry Street, University of Melbourne, Victoria 3010, Australia. Email: simonl@unimelb.edu.au.

[‡]Fuqua School of Business, Duke University, 100 Fuqua Drive, Durham, NC 27708, USA. Email: marx@duke.edu.

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*.¹ 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.²

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 his 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

¹To 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.

²In 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 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).

consumer: Someone who likes *Mumford and Sons* or *Joy Division* will typically be pleased to be referred to a band like *The Killers* or, respectively, *The National*. 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.³

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.⁴ 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

³According 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.

⁴Price 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.

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. With private values, consumers who like *ABBA* are best off if they also get to listen to, say, *Bee Gees*, *Cher*, and *Boney M.*

The remainder of this paper is organized as follows. Section 2 describes the model. In Section 3, we derive results and discuss price regulation. Section 4 extends the model to allow for investments into data analytics and product quality. Related literature is discussed in Section 5, and Section 6 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 his 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

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

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.⁵ 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

⁵While 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.

value.⁶ 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 $v = \underline{v}$.

We assume that $\frac{1-F(v)}{f(v)}$ is nonincreasing in v , which is sufficient to ensure that the consumer's virtual value function, denoted $\Phi_n(v)$ and given by

$$\Phi_n(v) = v - \frac{1 - F_n(v)}{f_n(v)},$$

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 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

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

Consequently, *social surplus* is

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

To be clear and upfront about this, the point of our model is *not* that it is particularly innovative or creative, but rather that it allows us to capture fundamental features and

⁶Alternatively, one could view the digital monopoly or a regulatory agency as choosing the level of consumer privacy, denoted by ρ , 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 ρ is willing to forego the top $\rho - 1$ out of ρ randomly selected products from the digital monopoly in order to preserve the privacy of his data. The tradeoffs we identify remain the same in this alternative parameterization.

trade-offs pertaining to ongoing policy debates. Therefore, we take it to be great news if many other, off-the-shelf models in industrial organization generate the same predictions and give rise to the same policy recommendation as our simple model.

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.⁷ Thus, we have the following characterization of the digital monopoly's optimal price:

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, $\lim_{n \rightarrow \infty} p_n = \bar{v}$. ■

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, the boundedness of the integral in (1), and $\lim_{n \rightarrow 0} f_n(x) = \lim_{n \rightarrow 0} nF^{n-1}(x)f(x) = 0$, it follows that

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

Therefore, given that for all $n \in (0, \infty)$, $CS_n(p_n) > 0$, it follows that $CS_n(p_n)$ is maximized at some value $n^* \in (0, \infty)$. This gives us the following corollary:

Corollary 1. *Consumer surplus is maximized at an interior value of n .*

⁷When $n < 1$, 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 .

With no data collection, match values deteriorate and 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) give rise to policy recommendations that are detrimental to the consumer.

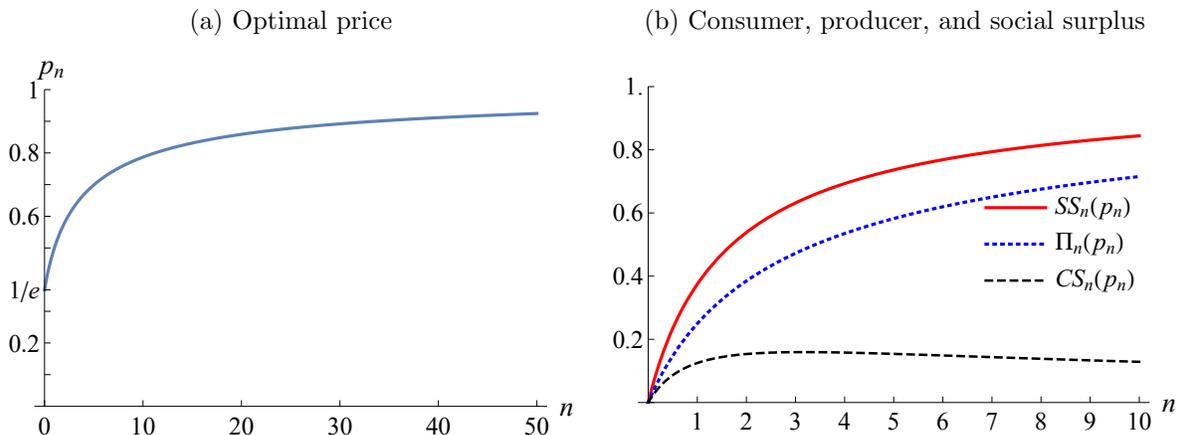


Figure 1: Panel (a) Optimal price p_n as a function of n . Panel (b): Expected consumer, producer, and social surplus at the optimal price p_n as functions of n . Consumer surplus is maximized at $n^* \approx 3.17$. Both panels assume that $F_n(v) = v^n$ and $c = 0$.

As illustrated in Figure 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:⁸

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

⁸In 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 $p \in (0, 1]$.

with the properties that

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

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.⁹ 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 $\alpha > 0$.¹⁰ 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 $SS_n(p)$

⁹If 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 $\lim_{n \rightarrow \infty} F_n(p_n) > 0$.

¹⁰This 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 $\alpha > 0$, $\lim_{n \rightarrow \infty} \tilde{p}_n = \bar{v}$.

and rearranging yields

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

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} ,¹¹ it follows that for any fixed α , there is a unique price that satisfies (3). Moreover, this price decreases in α 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

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

Because the first two terms of (4) are positive and the final term goes to zero as p goes

¹¹The 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} .

to c , it follows that for p sufficiently small, we have

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

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 $\frac{\partial^2 \Pi_n(p)}{\partial n \partial p} > 0$.¹² Thus, in this case, any binding price regulation reduces the digital monopoly’s incentive to invest in data.¹³ Using these results, we can connect “big data” and privacy concerns with incentives to invest:

Proposition 1. *Price regulation reduces a digital monopoly’s marginal incentive to invest in data analytics if the imposed 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.¹⁴

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

¹²To 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 $p < p_n$.

¹³The 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.

¹⁴Our 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.

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 Waldfogel (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 $[v, \bar{v}]$, 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 .¹⁵

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

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

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. We summarize 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.

¹⁵The 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.

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 Related Literature

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. Waldfogel (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, 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.¹⁶ 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).¹⁷ 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.¹⁸

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.¹⁹ In this paper, we relate this driving force of digital monopolies to monopoly pricing, including price discrimination and matching.²⁰ 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),

¹⁶For 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.

¹⁷Delacrétaz et al. (2019) provide a generalization of the impossibility theorem with extreme ownership.

¹⁸Loertscher and Wasser (2019) provide conditions for extreme ownership to be optimal when the objective involves profit motives.

¹⁹See Crémer (2010) for discussion and Loertscher and Riordan (2019) for partial resolution in an incomplete contracting environment.

²⁰Our 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.

critically depends on what data is used for—matching only or matching and pricing.²¹

6 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.

²¹This distinction is largely absent from concurrent policy debates (Chapter 5 Crémer et al., 2019).

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

$$-f^2F^n - (1 - F^n)f'F - (n - 1)f^2 \leq 0. \quad (5)$$

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 (5) 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 (5) 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. ■

References

- ACQUISTI, A., C. TAYLOR, AND L. WAGMAN (2016): “The Economics of Privacy,” *Journal of Economic Literature*, 54, 442–492.
- ALLCOTT, H. AND M. GENTZKOW (2017): “Social Media and Fake News in the 2016 Election,” *Journal of Economic Perspectives*, 31, 1–28.
- BERGEMANN, D. AND S. MORRIS (2019): “Information Design: A Unified Perspective,” *Journal of Economic Literature*, 57, 44–95.
- 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.
- CRAMTON, P., R. GIBBONS, AND P. KLEMPERER (1987): “Dissolving a Partnership Efficiently,” *Econometrica*, 55, 615–632.
- CRÉMER, J. (2010): “Solving the “Selective Intervention” Puzzle,” *Revue d’Économie Industrielle*, 129–130, 43–56.
- CRÉMER, J., Y.-A. DE MONTJOYE, AND H. SCHWEITZER (2019): “Competition Policy for the Digital Era,” EC Report.
- DELACRÉTAZ, D., S. LOERTSCHER, L. MARX, AND T. WILKENING (2019): “Two-Sided Allocation Problems, Decomposability, and the Impossibility of Efficient Trade,” *Journal of Economic Theory*, 179, 416–454.
- DUBÉ, J.-P. AND S. MISRA (2017): “Scalable Price Targeting,” University of Chicago.
- FIGUEROA, N. AND V. SKRETA (2012): “Asymmetric Partnerships,” *Economics Letters*, 115, 268–271.
- GOLDFARB, A. AND C. TUCKER (2019): “Digital Economics,” *Journal of Economic Literature*, 57, 3–43.
- JULLIEN, B., Y. LEFOUILI, AND M. H. RIORDAN (2018): “Privacy Protection and Consumer Retention,” Toulouse School of Economics Working Paper No. TSE - 947.
- KOBLIN, J. (2017): “Netflix Says It Will Spend Up to \$8 Billion on Content Next Year,” *New York Times*, October 16, 2017.
- LOERTSCHER, S. AND M. H. RIORDAN (2019): “Make and Buy: Outsourcing, Vertical Integration, and Cost Reduction,” *American Economic Journal: Microeconomics*, 11,

105–123.

- LOERTSCHER, S. AND C. WASSER (2019): “Optimal Structure and Dissolution of Partnerships,” *Theoretical Economics*, 14, 1063–1114.
- MALAMUD, S. AND M. ROSTEK (2017): “Decentralized Exchange,” *American Economic Review*, 107, 3320–3362.
- MYERSON, R. AND M. SATTERTHWAITTE (1983): “Efficient Mechanisms for Bilateral Trading,” *Journal of Economic Theory*, 29, 265–281.
- NEEMAN, Z. (1999): “Property Rights and Efficiency of Voluntary Bargaining under Asymmetric Information,” *Review of Economic Studies*, 66, 679–691.
- SHELANSKI, H. A. (2013): “Information, Innovation, and Competition Policy for the Internet,” *University of Pennsylvania Law Review*, 161, 1663–1705.
- VICKREY, W. (1961): “Counterspeculation, Auction, and Competitive Sealed Tenders,” *Journal of Finance*, 16, 8–37.
- WALDFOGEL, J. (2017): “How Digitization Has Created a Golden Age of Music, Movies, Books, and Television,” *Journal of Economic Perspectives*, 31, 195–214.
- WILLIAMSON, O. E. (1985): *The Economic Institutions of Capitalism*, New York: Free Press.