Reputation and the Provision of Data Security

Manos Perdikakis*

March 18, 2024

Abstract

In a two-period model, a monopolist chooses unobserved data-security investments. Consumers pay no access fee, but must share their personal data and suffer when data breaches occur. The firm wants to earn a reputation for protecting users' data, to maintain high activity in period two. I analyse two regimes of endogenous data-sharing, differing as to whether the firm or the consumers have ex-post control over it. Ex-ante commitment to data-sharing affects consumer surplus directly, but also via equilibrium investment. Starting at the firm-control equilibrium, the planner can improve total consumer surplus by reducing the amount of data that both high- and low- reputation firms collect. On the other hand, compared to the ex-post consumer optimum, committing to less data-sharing following a breach induces higher security; the ex-ante optimal level trades off the direct benefit of higher security against the cost of reduced learning about the level of cyber-risk. I discuss how these results relate to GDPR-type regulation regarding optional cookies, and also examine penalties and minimum security standards.

*manos.perdikakis@economics.ox.ac.uk. Department of Economics, University of Oxford, and Jesus College. I am grateful to my supervisor, Margaret Meyer, for her guidance on this project and must thank Greg Taylor, Alexei Parakhonyak, and Giulio Gottardo for very helpful discussions. This paper has also benefited from many helpful discussions with speakers of the Nuffield Economic Theory Seminar series, as well with members of the IO Reading Group at Oxford. I gratefully acknowledge funding from the Department of Economics at the University of Oxford, as well as the AG Leventis Foundation.

1 Introduction

Following the surfacing of major data breaches, most notably the Cambridge Analytica scandal, suspicion has arisen regarding the extent to which personal-data handling firms respect their users' privacy. In Mark Zuckerberg's own words, the Cambridge Analytica scandal represented a "breach of trust" between Facebook and its users[1]. In the Senate hearing that followed the scandal[2], he pointed at the crucial importance of "trust" for Facebook's business model, which depends on maintaining long-term relationships with users who share their most personal information with the platform[3].

Trust is necessary because it is difficult for firms to credibly signal that they adopt good data-protection practices. Even when firms try to be fully transparent with their privacy policies, users often do not read them thoroughly, due to the texts being significantly long or convoluted or making extensive use of legal terminology. Or users may treat them as cheap talk, i.e. non-binding statements. Even in the presence of stated privacy policies, firms seem to have significant ex post *discretion* on how to implement those policies and as demonstrated in Mark Zuckerberg's Senate hearing following the Cambridge Analytica scandal, it may be hard to verify the extent to which breaches occur due to firms' poor security practices. This means we can plausibly think of firms' actions to provide high data-protection as both *unobservable* and *non-contractible.*

This discussion implies that *reputations* for good data-security could play a big role in the interaction between long-lived firms and privacy-concerned consumers. I find this consistent with the observation that firms seem concerned with convincing users that they value their privacy[4]. For another example, see Apple's recent campaign from 2023: "Privacy. That's Apple". [5]

[1]https://www.forbes.com/sites/kathleenchaykowski/2018/03/21/mark-zuckerberg-addresses-breach-of-trust-in-facebook-user-data-crisis/#3cc9c33d3e36, 2018

[2]https://www.washingtonpost.com/news/the-switch/wp/2018/04/10/transcript-of-mark-zuckerbergs-senate-hearing/

[3]Senator Gary Peters' quote from the Senate hearing conveys the same message: "Can I believe...who has access to this information about me? So, I think it's safe to say, very simply, that Facebook is losing the trust of an awful lot of Americans as a result of this incident".

[4]Using Facebook as an example, its business website mentions: "..we take data protection and privacy very seriously and are committed to complying to data protection legislation..", while Mark Zuckerberg recently outlined his "Privacy-focused vision for Social Networking".

[5]https://www.apple.com/uk/newsroom/2023/01/apple-builds-on-privacy-commitment-by-unveiling-new-efforts-on-data-privacy-day/

This motivates me to investigate whether firms' concerns to maintain users' trust provide them with sufficient incentives to adopt data-protection practices. The concern is particularly acute because many digital service providers are mostly monetizing either consumer attention via ads, or more broadly consumers' data, offering their services for free, and may thus not have sufficient incentives to offer high quality services to their consumers. Using such a model of reputation incentives, I aim to understand how reputational dynamics interact with common policy remedies.

In my benchmark model, I examine a two-period interaction between a monopolist service provider and consumers, who do not pay a monetary price to access a firm's service, rather must share their personal data with the firm. The firm monetizes this information and chooses its level of *unobservable* data-security investment in order to avoid breaches of its database. I will be considering consumers who value their privacy, and with each data breach suffer disutility, which increases in the amount of data that they must share with the firm in exchange for the service. In the first section, this amount of data-collection will be treated as exogenous.

In addition to unobservable security investments, I use a model with consumer uncertainty about the extent of cyber-risk. The latter realistically captures the fact consumers are uncertain about the riskiness of sharing their personal data with a given firm. They will thus rely on the occurrence or not of data-breaches to *learn* about the risk of sharing data with the firm. After updating their beliefs, they make their activity choices again in the second period.

In a two-period model, firms will be motivated to invest in the first period in order to avoid public data breaches and the resulting harms to their *reputation* as adopters of good data-protection practices. In terms of modelling, I will be using a model of incomplete information in which firms may be Commitment or Normal types, and the Commitment type always provides high level of data-security. A lack of data-breaches is good news, and makes consumers update their belief upwards about the probability they are facing a Commitment type; I will refer to this belief as the firm's *reputation.* Absent regulation, investment incentives are purely implicit, motivated by consumer retention, and the Normal type makes no security investment in the last period.

I first use the benchmark model to identify how reputational incentives interact with common policies. I first examine the impact of two policies on equilibrium investment: penalties on firms that get breached and the specification of minimum security standards.

Both of those policy levers act as *commitment devices* in my model. High level of a minimum standard means that even if a firm has low reputation following a data breach, consumers understand that it will use security at least equal to the mandatory minimum in the second period. But this decreases the harm to the firm from having low reputation, thus erodes the implicit incentive to achieve high reputation in the first period. Unless the planner can specify a sufficiently high level for minimum security standards, adopting such a policy will increase second-period investment but *decrease* first period investment in equilibrium.

The main welfare and policy analysis in this paper follows next. I extend the model to study a setting in which the level of data-sharing is endogenous. I consider two different modes of endogenous data-sharing: the *consumer-control* regime in which consumers make **ex-post** optimal data-sharing decisions, given a firm's current reputation, and a *firm-control* regime, in which the firm chooses the amount of data consumers are required to share with the firm, in order to access the service. The firm chooses the requirement in order to ex-post maximize its profit, taking its current reputation into account.

I'm motivated to study these regimes of data-collection control because I believe they map accurately to how cookies-related data-collection took place before and after the introduction of the EU GDPR "opt-in" regulation. Under the latter, which is also the subject of the empirical study[6] by Aridor et al. [2023], firms must ask for consumers to explicitly opt into data collection and consent is required for *each purpose of data-processing individually.* Prior to this regulation, firms did not need to offer opt-out options and it is plausible to assume that they chose their cookie-collection in a profit-maximizing manner.

Starting at either of these equilibria, I ask: does ex-ante commitment to different *history-dependent* data-sharing levels allow a planner to increase expected consumer surplus relative to equilibrium? The planner observes no informational advantage over consumers but can commit to different data-sharing levels depending on the firm's reputation, i.e. whether it suffered a data-breach or not. I will be asking this question for the equilibrium of each of the two regimes of ex-post data control.

Changes in the levels of data-sharing in the second period have multiple effects on consumer surplus (CS) in this model: the direct, on the utility of active users in the second period, and the indirect, via changing equilibrium investment incentives. In turn,

[6]The authors find a reduction in total cookies by 12.5% caused by this regulation.

equilibrium investment affects consumer surplus via first-period disutility from breaches, and via increasing the relative frequency with which a Normal (low-security) type will have a high reputation in period two. Conditional on facing a Normal type, consumers face ex-post regret in equilibrium following no-breach; they share too much data (in either regime) and too many users are active relative to a perfect information setting, because they entertain the possibility of facing a Commitment type. I call this the "signal-jamming" effect of higher investment and it is always *negative*: high first-period secuirty impedes learning and reduces second-period CS.

Starting from the equilibrium under firm-control, changes in the levels of second-period data sharing have no first-order impact on investment incentives; data-sharing affects investment incentives via changing profits in each of the second period states, and at the ex-post profit maximizing levels of data sharing, profit is insensitive to changes in them. This means that the only first-order impact is the *direct* one on second-period CS. A profit maximizing firm whose revenue per consumer increases with data-sharing will always ask for so much data that CS is decreasing at the margin, so that the direct effect of limiting data-sharing is positive. A CS-maximizing planner who faces a regime of firm-control can therefore set small caps on the levels of data-collection in period 2, on *both* high- and low-reputation firms and achieve an increase in *total* CS.

On the other hand, a planner that starts from a consumer-control equilibrium faces a different situation; in that case, the direct effects on expected CS_2 are zero because data-sharing in period two is chosen optimally by consumers. Therefore, consumers can benefit in the second-period by changes in data-sharing that induce *less* equilibrium security, so that there is more accurate learning about the environment, i.e. less signal-jamming. However such a reduction will come at the expense of first-period security. This is the fundamental policy trade-off that emerges in this model, because of the dual role investment has. It both affects real outcomes, but also impedes learning about the firm's type. At this equilibrium, security investment may be too high or too low relative to the consumer-optimal level. Data-caps for high- and low- reputation firms have effects of opposite direction on equilibrium investment, so that the nature of intervention is different according to the firm's reputation.

I then extend the benchmark model to a duopoly; I find that with linear revenue, equilibrium investment of each firm is always lower relative to monopoly; this is a simple consequence of the fact that under linear revenue in market share, the presence of a competitor will reduce the marginal benefit to achieving high reputation in the second period. As the previous analysis suggests, this does not necessarily imply lower consumer surplus, since it will imply faster learning about firms' types and less ex-post regret in the second period. In Appendix C, I introduce endogenous data-sharing and examine how data caps can affect consumer welfare in a duopoly, where firms simultaneously choose their required levels of data-sharing to attract consumers. I use mostly numerical simulations to find that data caps can consistently increase consumer surplus relative to the firm-control optimum, despite the fact that competition drives firms' equilibrium data extraction down.

To further motivate the model, it is worth it looking at some literature which suggests (1) firms might indeed suffer financial damage following a data breach and (2) consumers do value their privacy. Focusing on public corporations in the US, Kamiya et al. [2021] find significant negative abnormal returns only when cyber attacks induce the loss of personal data; the abnormal returns of firms that do experience negative returns are almost 500 USD million per attack (1 percent of value). Closely related to my model of reputation incentives, the authors argue that in a full-information world where there is no learning about the firm or the environment after a successful cyber attack is disclosed, the firm's loss of value should only reflect *out of pocket fees* (e.g. penalties, legal fees, etc.). Using data on disclosed breaches from 2005 to 2017, they estimate that cyber attacks have substantial additional **reputation** costs on top of those due to expected legal action and penalties. Reputation in their setting, and in mine, is synonymous with the firm-specific distribution of losses due to cyber attacks that the customers perceive. Thus, their paper provides valuable empirical justification both for the learning component of my model and the existence of firm incentives to avoid data breaches.

Regarding (2), Lin [2022] attempts to disentangle between "taste" for privacy and instrumental preferences, i.e. preferences stemming from anticipated surplus loss in the absence of privacy. In my model, I will not take a stance on whether consumers' privacy preferences are intrinsic or instrumental. Using a lab experiment, the paper finds that consumers do have both *intrinsic* and instrumental preferences for privacy.

Even though I have drawn motivation from the Cambridge Analytica case, the concerns described above are not restricted to social media. The example that motivates the analysis of closely related work in Jullien et al. [2020] can be effectively used to provide motivation for my work too; the authors recall an incident in which the Times website, due to *insufficient diligence* in screening third-parties that were allowed to post ads on the newspaper's website, exposed its users to digitally harmful material.

In the next section, I discuss related literature. In Section 3, I present the benchmark monopoly provider model, with exogenously determined levels of data sharing. The main body of policy and welfare analysis is in section 4, in which I introduce the two regimes of endogenous data collection, and discuss the ability of a planner to improve consumer surplus by pre-committing to levels of second-period data sharing that depend on a firm's posterior reputation. Section 5 introduces the extension of the baseline model to a duopoly, Section 6 shortly analyses two simple policies in the duopoly context with exogenously determined data-sharing. The paper then concludes.

2 Related Literature

In this literature review, I find it more worthwhile to discuss few papers in greater length that are closest to mine, rather than attempt to list all papers in the large literature about the economics of privacy and cyber-security provision. There are excellent surveys both very recent by Goldfarb and Tucker [2023], as well as slightly older by Acquisti et al. [2016]. The former also covers the recent empirical work, both on the economic impact of GDPR and on measuring privacy concerns. The latter deals in depth with the theory literature on the economics of privacy.

The paper closest to mine is probably Jullien et al. [2020]. Their model uses a signal-jamming, two-period model of belief formation and in their paper too, firms take unobserved actions, in the form of screening the third-parties they share consumer data with. In their model, as in mine, equilibrium incentives are based on the prospect of consumer retention. However, consumers do not update their beliefs about firm attitudes towards privacy, rather about their own *vulnerability* in the event of a data-breach. Their single-website model is similar to my model of monopoly with fixed data terms, but the focus of their paper is multi-homing competition between websites.

They study this mode of competition for consumers in order to focus on (a) website competition in the *advertising market* and (b) on a novel "public good" problem between websites: as long as consumer vulnerability is positively correlated across websites, a lack of precaution by any one of them means that the consumer is more likely to value using any of the other websites in the next period less. The public-goods aspect means that in the perfect correlation case, a "zero-protection" equilibrium always exists. In contrast to their analysis, I distinguish between data-collection and data-protection, and I focus on *history-dependent* caps on data-sharing and analyse different regimes of endogenous data-collection.

The following papers, de Cornière and Taylor [2021], Ahnert et al. [2022], and Fainmesser et al. [2023], all have as their main focus the impact of the firms' business model on equilibrium incentives for privacy provision or cyber-security investment. They differ from mine in that they model the incentives of cyber-attackers and the frequency of data-breach *attempts* is endogenous in their papers. They all use static models and do not consider reputation-based incentives for security.

In the first one, de Cornière and Taylor [2021], the authors study the interaction between the firms' business models in duopoly and equilibrium levels of cybersecurity. They do this in a static setting with observable investments. The introduction of strategic hackers introduces a *negative* network externality between users of each firm, since a large user base attracts more data-stealing attempts. Their discussion focuses on comparing equilibrium investments between ad-funded duopolists and product funded ones that charge (endogenous) prices. The fact that they use observable security decisions, leads to different findings than mine regarding the efficiency of investment provision in monopoly[7] and duopoly, in the comparable advertising regime, compared with the extension of my model to a duopoly setting.

Ahnert et al. [2022] is a model of security provision and fee choice by financial intermediaries. Unlike de Cornière and Taylor [2021], they study a single business model of the firm that interacts with consumers, but they study different modes of operation by the hackers, who can either choose to ask the firm for ransom or engage in conventional attacks and attempt to steal users' data. Attackers first choose their mode of operation which the firm observes, then the firm chooses fee and security level (which the users may or may not observe, they deal with both cases) and then attack commences. Both papers study the optimal design of liability as well as minimum security standards.

Fainmesser et al. [2023] also models attackers' side in detail. Their innovation is dealing with both data-storage and data-sharing choices of the firm and they analyze those, both for ad-funded and transaction-funded firms. They take a firm's business model as given and find the optimal data collection and data security levels. Firms that are more data-driven,

[7] In their "monopoly" example, they assume full market coverage, hence there exist no incentives for the provision of security.

set both higher levels of data collection and protection. This complementarity arises because higher collection attracts more attackers and thus raises the marginal benefit of protection.

The paper by Markovich and Yehezkel [2021] focuses on the comparison between consumer-and firm-control of data-collection decisions, also motivated by GDPR-style regulation which gives consumers greater control. They do not study cyber-security, but in their model which regime is optimal depends on whether disutility from data-commercialization varies more across users or across data-items.

To the best of my knowledge, and according to the survey by Goldfarb and Tucker [2023], and there is no empirical work documenting the impact of GDPR on cyber-security investments and equilibrium frequency of data-breaches. As Garrett Johnson notes, "we have seen more research on the *unintended* consequences of the GDPR, rather than the *intended*".

Koutroumpis et al. [2022] examine the link between hiring of cyber-security specialists by firms and stronger data-protection laws and enforcement in the UK, and find significant positive impact of the new policy on cyber-security hiring expenditure, using Burning Glass job ads data. They focus specifically on data-breaches as a subset of cyber-attacks, since those both (a) involve a loss of personal information and (b) are often harmless to the victim firm in terms of operations disruption. These two elements lead to potential *under-investment*. This is also the motivating application I will make use of for this paper, i.e. I will not be thinking about *ransomware attacks*, because they are both directly harmful to the firms that have to pay ransom to restore some part of their digital operations, and could plausibly be costless to consumers, if the firms pay the ransom and attackers are 'noble' in the sense that they don't sell data even after receiving a ransom payment.

3 Monopoly model

There is a continuum of consumers with mass one, uniformly distributed over a line on $[0, 1]$. They interact with a single firm over $T = 2$ periods. The firm provides a digital service, is positioned at $A = 0$ and wishes to attract users to register; Registration of users lasts 1 period, while firm and users live for 2 periods. Users make their participation decisions at the beginning of each period. The firm charge users *no registration or usage* *fees*, but users must share their personal data with the firm in order to use the serive[8]. Firms have some ex post discretion on how to treat the data of users they have attracted in a given period. As earlier discussed, this can be thought of as effort that firms exert to better screen third parties that get access to consumer data, or as investment in cyber-security to deter data breaches. Crucially, I assume that this effort is *unobserved* by the consumers and *non-contractible.* I find this assumption reasonable; even if a data breach is made publicly known, it could be quite costly, if at all feasible, to prove that it was due to lack of due diligence by the firm. This variable will be denoted by e and I refer to it as the effort/investment/security level.

Towards attracting privacy-concerned potential users, the firm faces potential gains from maintaining a reputation of caring about users' privacy. I capture this intuition by introducing incomplete information about firms' *types*; a firm can have type N or C, which stand for Normal and Commitment type, respectively. The type of the firm is privately known to it. The Commitment type is non-strategic and always chooses the same action. In particular, a Commitment type is the "good" type and always chooses the highest level of effort, $e = 1$. In this paper, we will be concerned with equilibrium incentives of Normal types.

As we will see in the next subsection, consumers of a digital service care about the level of effort the provider exerts, because this effort level determines the probability with which they experience a breach of their personal data and suffer privacy disutility. As mentioned in an earlier section, I primarily think of this privacy cost as psychological and provide a reduced-form representation, but it could also be motivated by price or non-price discrimination concerns. I define the "outcome" binary random variable s_t, which can take values $\{b, n\}$, standing for *breach* and *no breach.* The value of this random variable becomes publicly known at the end of each period. I also define $P(s_t = b|type, e_t)$ as the probability that firm A will suffer a *breach*, given its type and effort level exerted by the Normal type in period $t \in \{1, 2\}$.

Throughout this paper, I use the following conditional probability mass function for the outcome variable, s_t.

$$P_t^b := P(s_t = b|type, e_t) = \begin{cases} \zeta & , type = C \\ \zeta + (1-\zeta)(1-e_t) & , type = N \end{cases}$$

The interpretation of the above pmf is simple: in each period, there is a probability

[8] Or we could think of this as data being generated by their activity on a website/app.

$\zeta \in [0, 1]$, that a negative breach "shock" will arrive regardless of the firm's effort choice. Conditional on the negative shock not arriving, as the above specification suggests, breach probabilities in period 2 are independent of s_1 and e_1. I will refer to the firm's "reputation" in a given period, as the probability with which users believe that the firm's type is C in that period. The firm has prior reputation μ_1 which is common knowledge. After observing $s_1 \in \{n, b\}$ at the end of $t = 1$, fully rational users update their beliefs the firm's type using Bayes' Rule. The posterior reputation, $\mu_2 \in \{\mu_n, \mu_b\}$, depends on the prior and also on the effort level that users believe the Normal-type exerts in the first period, $\tilde{e_1}$. I assume all users share the same conjectures, and everybody observes the realization of s_1 accurately, so that there is a single posterior reputation for the firm at the end of $t = 1$.

For any $\zeta < 1$, the posterior reputation a firm achieves following a $s_1^A = n$ realization in period one is:

$$\mu_n = \mu(C|s_1^A = n, \tilde{e_1}) = \frac{\mu_1}{\mu_1 + (1 - \mu_1)\tilde{e_1}} \tag{1}$$

The above posterior is not well-defined for $\zeta = 1$, since the probability of a non-breach outcome becomes zero in that case. Posterior reputation following a good outcome takes values in $[\mu_1, 1]$. What is immediately apparent in expression (1), is that posterior reputation is decreasing in the effort conjecture, $\tilde{e_1}$. This is very intuitive; if users anticipate that a Normal type firm exerts a lot of effort to prevent breaches from occurring, it must be that a "no breach" outcome is less surprising, less indicative of a Commitment type and thus has a smaller effect of improving the firm's prior reputation[9]. When $\tilde{e_1} = 0$, then all good news indicate a Commitment type with certainty, so posterior reputation following $s_1 =$ "*no breach*" is equal to one. In contrast, when $\tilde{e_1} = 1$, the Normal type firm perfectly replicates the Commitment type's behaviour in period 1, hence posterior reputations are not updated and $\mu_n = \mu_1$. Note that μ_n does not depend on ζ since the ratio of probabilities with which each type achieves a "no-breach" realization is constant with respect to ζ.

Similarly, the posterior reputation for a firm that has a "breach" outcome in period 1 is:

$$\mu_b = \mu(C|s_1 = b, \tilde{e_1}) = \frac{\zeta\mu_1}{\zeta\mu_1 + [(1 - \zeta)(1 - \tilde{e_1}) + \zeta](1 - \mu_1)} \tag{2}$$

which now is always *smaller* than μ_1 and is *increasing* in ζ. This is intuitive; for $\zeta = 0$, we

[9] This is in contrast with other models of effort provision: For instance, in models using the setup of Holmström [1999], posterior reputation is not a function of effort conjectures, hence the resulting first-order condition is linear in the simplest model.

are in a perfectly revealing bad news setting and a breach realization lets consumers know that they are facing an N type with certainty. Higher ζ allows consumers to entertain the possibility that the breach was a result of a negative shock. This posterior is *increasing* in the consumers' effort conjecture, since a bad result is more likely to be the outcome of a negative shock rather than firm negligence (low effort).

We can think of ζ as inversely related to the quality of public infrastructure and support given to firms to protect against cyber warfare. For instance, as the level of support that firms receive in terms of information provision regarding state-of-the-art cyber attacks. Similarly, we can think of ζ as the probability in each period that firms are attacked using highly sophisticated hacking methods that they could not have protected themselves against, or simply as the minimum probability that firms are exposed because of human error in their processes (e.g. an employer losing their work laptop).

3.1 Monopoly Equilibrium

I now turn to users' payoffs and participation decisions. As mentioned already, users make their participation decisions at the beginning of each period, meaning that users choose between using service A or staying idle. Each user is characterized by location θ on the line, which is the value of their outside option. In this section, I adopt the simplest utility function and I assume that users suffer disutility equal to $(-\delta_t)$ in any period in which they suffer a breach – there is no heterogeneity in privacy preferences and data is purely harmful. I think of δ_t as the data input (in utility terms) required by consumers to use the service. It can differ across periods, and in this section, I will treat it as *exogenously* given. Higher value of δ_t means greater disutility for a consumer who suffers a breach in period t. Expected utility given information I_t is then:

$$EU_t(\delta_t, P_t^b) = v - \delta_t P_t^b \tag{3}$$

Implicit in this utility specification is the following assumption: a user that has suffered a breach in the first period will incur additional disutility of $(-\delta_2)$ if their data is breached again in the second period. The disutility of an active user that experiences a breach in the second period is independent[10] of both first-period activity and s_1. Last, but not least, I abstract from network effects and informational externalities by assuming that the utility

[10]This means, for example, that an agent that uses the service in both periods does not incur higher privacy cost in the event of a second-period breach than a user who just uses A in $t = 2$. Users suffer only because their current period data is exposed.

users derive from using the service is independent of other users' participation decisions (both past and present). If informational externalities in the spirit of Acemoglu et al (2022) were present, then the outside option would be weakly negative and decreasing in the mass of active users. Given the above specification of expected utility, a user is active in period 2 if $\theta < \theta_2(\mu_2)$:

$$\theta_2(\mu_2) = v - \delta_2((1 - \mu_2) + \mu_2\zeta)$$

The firm earns an exogenously set price p per user it serves, which is net of the constant marginal cost of servicing an additional consumer. This is to ease exposition and we will consider data-dependent revenue later. From consumers' perspective, the probability that $s_2 = b$ is $((1 - \mu_2) + \mu_2\zeta)$; as we will see, an N type will optimally exert $e_2 = 0$, i.e. will always suffer a breach in period two in $t = 2$, and the probability that a C type is breached is ζ.

Given the expressions for the cutoffs, the Normal type chooses e_1 to maximize expected profits across both periods net of investment cost:

$$\Pi(e_1; \tilde{e_1}) = p\theta_1(\mu_1, \tilde{e_1}) + P(n|N)p\theta_2(\mu_n) + (1 - P(n|N))p\theta_2(\mu_b) - C(e_1) \tag{4}$$

where $P(n|N) = (1 - \zeta)e_1$ and $P(b|N) = 1 - P(n|N)$. The cost function is increasing and convex in e, $C(e) = \frac{1}{2}ce^2$, and the Normal-type firm chooses effort to maximize the above profit function, taking effort conjectures as given. Note that the first period cutoff and posterior reputation μ_2, $\theta_1(\mu, \tilde{e_1})$, only depends on consumers' *conjecture*, $\tilde{e_1}$, and are not directly influenced by the firm, even though that conjecture will have to be correct in equilibrium. The first-order condition with that an interior solution must satisfy is:

$$\begin{aligned} (1 - \zeta)p(\theta_n - \theta_b) &= ce_1 \iff \\ (1 - \zeta)^2\delta_2(\mu_n - \mu_b) &= \frac{c}{p}e_1 \end{aligned} \tag{5}$$

where $\theta_n := \theta_2(\mu_n)$ and $\theta_b := \theta_2(\mu_b)$. Equation (5) defines the monopolist's optimal effort provision, taking beliefs as exogenously fixed. Investment in security is only motivated by concerns to attract users in period 2. Greater difference between revenue in the two potential outcomes induces higher investment provision. To turn (5) into an equilibrium defining equation, I must impose the equilibrium condition that conjectures are correct, i.e. $\tilde{e_1} = e_1$. Since the difference of posterior reputations $\mu_n - \mu_b$ is *decreasing* in $\tilde{e}$, we obtain equilibrium existence and uniqueness.

Proposition 1 *A unique Perfect Bayesian Equilibrium of the monopoly game exists for all parameter values. Type C plays* $e_1 = 1$ *in every period and type N plays* $e_2 = 0$*. In equilibrium, users' conjectures are correct, i.e.* $\tilde{e_1} = e_1$ *and first-period choices by the firm maximize expected profit given those conjectures. I call the equilibrium level* e^M*:*

- *If* $\zeta = 0$, e^M *is given by the unique positive solution to the equilibrium first-order-condition, if the latter is weakly lower than 1, which happens for* $c > \mu p \delta_2$*. Otherwise, it is given by the corner solution* $e^M = 1$*, and we have a pooling equilibrium.*

- *If* $1 > \zeta > 0$, e^M *is given by the unique solution to the equilibrium first-order-condition* (5) *in* $[0, 1]$ *and is always strictly between* $(0, 1)$*.*

- *If* $\zeta = 1$*, no positive investment can be supported in equilibrium,* $e^M = 0$*.*

We obtain the following intuitive comparative statics results:

Corollary 1 *The following hold regarding the equilibrium effort level* e_M*:*

- *Equilibrium effort is weakly*[11] ***increasing*** *in* p, δ_2*.*

- *Equilibrium effort is weakly* ***decreasing*** *in* β, c*, and* ζ*.*

To interpret these comparative statics results, an argument under fixed conjectures suffices, i.e. they all match the signs of the partial derivatives taken if equilibrium effort were given by equation (5). This is a simple consequence of the fact that equilibrium is found at the intersection of a downward sloping curve of the firm's best-response to consumers' conjecture $\tilde{e_1}$ with the $e_1 = \tilde{e_1}$ line. This implies that the total effect of a change in parameters on equilibrium investment will be of the same direction as if beliefs were fixed, but also of lower magnitude.

It is simple to understand the result with respect to the exogenous shock probability, ζ, by looking at the expected reputation gain from avoiding a breach, in the left-hand side of (5). For any fixed users' conjecture about effort, higher ζ reduces this difference, thus attenuates incentives for achieving high reputation by the N type. Even though this is not a surprising result, it is interesting to rethink about the interpretation of ζ as a parameter that a regulator can affect. Apart from the direct gain of reducing ζ, a regulator would also benefit indirectly by increasing the effort induced by Normal type firms.

[11] Because for some parameters we might have corner solutions $e_1 = 1$.

It is non-trivial to establish whether a firm with a higher prior μ will exert higher effort in equilibrium, since both posterior reputations present in the net gain term are increasing in the prior, holding conjectures fixed.

In the case where $\zeta = 0$, and the posterior following a breach becomes zero for the firm that suffered it (perfect bad news), it is easy to see that equilibrium effort would be increasing in the prior because the expected gain from not getting breached would be higher for every conjecture level held by consumers. By continuity of e_1^M in ζ, we can expect that this result will carry-through for low values of this parameter.

3.2 Discussion of assumptions

Before moving on to the policy analysis and model extension, I discuss some assumptions on which the equilibrium derivation and subsequent analysis does not depend on qualitatively.

1. User homogeneity in data preferences does not matter; what does matter is that demand in each state is decreasing in the probability of a breach that users perceive.

2. Utility does not have to be always decreasing in data. We can repeat the analysis with concave utility such that at any reputation level, the first units of data-sharing are beneficial to consumers. In fact, such a utility function facilitates the analysis of the section with endogenous data-sharing and will be used there.

3. The same applies to the assumption of firm revenue being independent of data-sharing. It does not affect the previous results and will be relaxed in the next section.

4. The firm need not be privately informed about its type. the model would work very similarly as a pure "signal-jamming" model, in which some firms are type C and some are type N, but the firm itself also does **not** know its own type. The marginal benefit on investment to a firm that does not know its own type would have to be multiplied by $(1 - \mu_1)$, since investment is only valuable if the firm is Normal type. This modelling assumption may be more appropriate if we are thinking as data-breach risks coming from, for example, zero-day vulnerabilities, the existence of which the firms are reasonably assumed to be unaware of. Finally, note that the firm's type does not need to be time invariant, just positively correlated across the two periods, for investment incentives to be supported in equilibrium.

5. The model can be extended to allow for a simple treatment of positive consumption externalities. Sufficiently high magnitude means the monopolist achieves full market coverage in period two regardless or μ_2, thus has no investment incentives. For modest magnitude, the analysis remains qualitatively the same.

4 Endogenous data-sharing

In this section, there are two objectives: First, I extend the monopoly model just presented to account for *endogenous* choices of data-policies $\delta_1, \delta_n, \delta_b$, where δ_1 refers to the first period, and the two other terms refer to the two possible states of period 2. I will be focusing on two different regimes of *ex-post* control over data-sharing. In the regime of *consumer-control*, consumers can choose in every period the amount of data they want to share with the firm, if they participate at all. They can thus react to new information about the firm in the second period, by changing how much data they share with it to maximize their second-period expected utility.

In the second regime, of *firm-control*, the firm chooses it's profit maximizing data policy in each period and state; I use *ex-post* control, in the sense that the firm cannot pre-commit in period 1 to how much data it will ask for consumers to share in period 2.

The second objective in this section, is to understand whether a planner that can change the amount of data shared in each state of the second period, i.e. either following a breach or following no-breach, can do so in a way that improves consumer surplus relative to either of the "regulation-free" equilibria of each regime. Under the consumer-regime, this is equivalent to asking whether consumers would collectively benefit from pre-commitment to different levels of 2nd period data-sharing than those that maximize CS ex-post. Maintaining the initial assumptions of no data-sharing externalities between consumers, data-sharing decisions under the consumer regime are ex-post CS-maximizing. I further maintain the homogeneous data-preferences assumption, so that there a unique data-sharing level that maximizes each consumer's surplus from using the firm's service.

Throughout this section, I will be focusing on the perfect bad-news case of $\zeta = 0$, i.e. $\mu_b = 0$, but this only simplifies the exposition of results and is not necessary.

4.1 Equilibrium in the two regimes

To have a more meaningful investigation of the consumer-regime, we adapt the utility function so that ex-post optimal data-sharing is not always trivially zero for consumers, which would be the case with purely harmful data. I will use a specific functional form, but the results discussed will go through for any utility function that is (quasi-)concave in the amount of data-shared, and is such that consumers benefit less from data-sharing when facing a higher probability of a breach. With the specific functional form, a consumer who uses the service has expected utility of:

$$u(\delta, P_t^b) = \alpha\delta - P_t^b\delta^2 \tag{6}$$

which is a function of P_t^b, the perceived probability of suffering a breach in period t, using all information available in that period, and data-sharing, δ. Notice that this is always inverse-U shaped in δ, i.e. the first units of data-sharing always offer positive marginal utility to consumers. All consumers have equal values of α but heterogeneous outside options $\theta \sim U[0,1]$. The mass of users for given reputation and data level is $D(\delta, P_t^b) = \theta^* = u(\delta, P_t^b)$, and we note that $D_{\delta\delta} < 0$, demand is *concave* and the mixed partial derivative is negative $D_{\delta p} < 0$, since $u_{\delta p} < 0$. Higher probability of a breach always decreases the marginal utility from more data, and thus suppresses the location of the indifferent consumer.

The level of data-sharing chosen under consumer-control, δ^C is:

$$\delta^C(P_t^b) = \frac{\alpha}{2P_t^b} \tag{7}$$

The above is decreasing in the probability of a breach, so that second period δ^C is *increasing* in the firm's reputation. Higher reputation increases utility at any level of data-sharing, and thus the location of the indifferent consumer (envelope theorem), but also increases firm revenue via both greater demand and greater revenue per consumer. In other words, if we define $\Pi^C(P_t^b) := \Pi_t(\delta^C(P_t^b), P_t^b) = r(\delta^C(P_t^b))D(\delta^C(P_t^b), P_t^b)$, we obtain the intuitive $d\Pi^C/dP^b < 0$.

To determine data sharing under *firm-control*, remember that the revenue function for both Commitment and Normal types is $\Pi_t(\delta, P_t^b) := r(\delta)D(\delta, P_t^b)$. For exposition, I focus on profit-maximizing level for revenue per consumer linear in data-collection $r(\delta) = p_1\delta$.

$$\delta^F(P_t^b) = \frac{2\alpha}{3P_t^b} \tag{8}$$

which is the maximizer of Π_t, since for any $\alpha > 0$ the profit function is quasi-concave. It is important to notice that we assume there are no data-adjustment costs between periods: thus, Normal and Commitment type firms have the same optimal δ choices at every period and state, and there is no type-signalling[12] from the choice of δ_1. Observe that for any P_t^b, δ^F is larger than δ^C, as predicted by evaluating the firm first-order condition at the consumer optimal data level: both utility and profit are concave in δ, and profit is still increasing at δ^C, because $\Pi_\delta = r'D + D'r = r'D > 0$ at δ^C. I will refer to this feature often, so it is useful to state it as a Lemma.

Lemma 1 *Assume that at any P_t^b, Π and u are quasi-concave in δ. Then, a firm that sets data-sharing requirements to maximize current-period profits will optimally do so at a level that satisfies $u_\delta(\delta, P_t^b) < 0$.*

This means that in equilibrium of the game with firm-control, consumers would always rather that the firm asks for less data in each state.

I emphasize that this model preserves a feature of the previous analysis, namely that the consumer type is not interpreted as data-sensitivity, i.e. $u_{\theta\delta} = 0$. This means that consumers always agree on the optimal level of data sharing δ^C.

Under both regimes, equilibrium existence and uniqueness follows from identical arguments, and very similar to those under exogenous data, because firm profit in period 2 remains increasing in its posterior reputation. An (interior) *equilibrium* under regime C is defined as a combination $(e_1^C, \mu_n^C, d_1^C, \delta_n^C, \delta_b^C)$, such that:

1. Beliefs following no-breach are consistent with Bayes' rule, given investment level e_1^C.

2. Given e_1^C, μ_n^C, consumers choose data-sharing in each period and state according to $\delta = \delta^C(P_t^b)$, and

3. e_1^* satisfies the investment-foc, given that profits in each state and period are $\Pi_t^C(P_t^b)$. The first-order condition for $\zeta = 0$, following the derivation of the previous section, is:

$$\Pi_2^C(1 - \mu_n(e)) - \Pi_2^C(1) = C'(e) \tag{9}$$

[12]Committing to their future data-requirements in period 1 would potentially allow Commitment firms to signal their type. A firm that knows it will have higher reputation in the second period finds it more profitable to commit to a higher level of future data sharing, because higher data requirement makes profit more sensitive to current reputation.

where I have used the fact that the second-period probability of a breach is $(1-\mu_n)$, i.e. the probability of facing a Normal type[13], and I am making explicit that posterior reputation following "no-breach" depends on equilibrium investment beliefs.

Similarly, we can define the unique equilibrium under firm-control, $(e_1^F, \mu_n^F, d_1^F, \delta_n^F, \delta_b^F)$. It is still the case that the firm cannot influence consumer conjectures about investment in period one. Hence, optimal e_1 (and δ_n, δ_b in the firm-regime) is independent of the choice of δ_1, and we will omit δ_1 from most of the discussion below.

4.2 Regulation in the two regimes

In this section, I will be considering a regulator that can impose changes in the amount of data sharing in each state of period 2. This means that the regulator also observes the state realization, and can adjust data-sharing according to beliefs about the firm's type. The regulator has no informational advantage over consumers and can only condition data-collection levels in period two on publicly available information about the firm.

I will be considering again the game with **exogenous** data terms $\delta_n, \delta_b, \delta_1$, whose equilibrium was derived in the first section. I will ask how consumer surplus changes in the equilibrium of this game as the regulator changes δ_n and δ_b, and will see how the answer changes depending on whether we evaluate the changes in CS at the equilibrium or the game under ex-post consumer or firm control.

I will first try to get some insight about the regulator's ability to raise CS_2, i.e. expected consumer surplus in period 2. Starting from any pair (δ_n, δ_b) and the well-defined equilibrium (μ_n, e_1) that these induce:

$$\frac{dCS_2}{d\delta_b} = \frac{\partial CS_2}{\partial e_1}\frac{\partial e_1}{\partial \delta_b} + \frac{\partial CS_2}{\partial \delta_b} \tag{10}$$

$$\frac{dCS_2}{d\delta_n} = \frac{\partial CS_2}{\partial e_1}\frac{\partial e_1}{\partial \delta_n} + \frac{\partial CS_2}{\partial \delta_n} \tag{11}$$

Each data term has two effects on CS_2, a **direct**, via changing utility consumers derive from data in period 2, and an **indirect** via changing first-period investment of the Normal type, which depends on the slope of investment with respect to corresponding data term. We will analyse each of the three terms separately.

Let's start with the direct effects. Starting from a pair (δ_n, δ_b) and the well-defined interior equilibrium (μ_n, e_1) that these induce, the direct effect induced by a change in δ_b

[13] Which following a breach is equal to 1, when $\zeta = 0$.

is *negative* in the relevant region $(\delta_b^C, \delta_b^F]$ between the values of the two 'regulation-free' equilibria[14]. To see that, note that:

$$
\begin{aligned}
CS_2 = \mu \int_0^{\theta_n} \Big(u(\delta_n, 0) - \theta \Big) d\theta \\
+ (1-\mu) \left[e_1 \int_0^{\theta_n} \Big(u(\delta_n, 1) - \theta \Big) d\theta + (1-e_1) \int_0^{\theta_b} \Big(u(\delta_b, 1) - \theta \Big) d\theta \right]
\end{aligned} \tag{12}
$$

so that:

$$
\frac{\partial CS_2}{\partial \delta_b} = (1-\mu)(1-e_1)\theta_b \frac{\partial u(\delta_b, 1)}{\partial \delta_b} \tag{13}
$$

and we know by Lemma 1 the partial derivative in the right-hand side is negative for $\delta_b > \delta_b^C$ and equal to zero at δ_b^C.

I turn to the direct effect induced by changes in δ_n. This is a bit more subtle to analyse, because the sign will depend on the comparison between the exogenous initial δ_n and an endogenous object, $\delta^C(\mu_n)$, which I introduce next. For a δ_n, δ_b pair of exogenous data terms that induces an equilibrium posterior belief μ_n, we can define the ex-post optimal data-sharing levels for either firm or consumers in state n, given those equilibrium beliefs μ_n. Call these $\delta^C(\mu_n), \delta^F(\mu_n)$ and by Lemma 1, it will always hold that $\delta^C(\mu_n) < \delta^F(\mu_n)$; I emphasise that these depend on the exogenously defined parameters δ_n, δ_b, via the μ_n the latter induce.

At any equilibrium beliefs μ_n, an increase in δ_n benefits consumers via the direct effect iff $\delta_n < \delta_n^C(\mu_n)$. We summarize the previous discussion into a Lemma:

Lemma 2 *Starting at an interior equilibrium* e_1, μ_n *induced by a pair of parameters* (δ_n, δ_b)*, the direct effect on* CS_2 *of an increase in* δ_b *is negative if and only if* $\delta_b > \delta_b^C$*, while the direct effect of an increase in* δ_n *is negative if and only if* $\delta_n > \delta^C(\mu_n)$*.*

Next, I discuss the investment slope terms, beginning with $\partial e_1/\partial \delta_b$. For any fixed δ_n, δ_b, an equilibrium is a pair[15] (e_1, μ_n) defined by:

$$
ce_1 = \Pi_2(\delta_n, 1-\mu_n) - \Pi_2(\delta_b, 0) \tag{14}
$$

$$
\mu_n = \frac{\mu}{\mu + (1-\mu)e_1} \tag{15}
$$

[14]The perfect-bad news assumption simplifies this argument because both bounds of the interval are just scalars, independent of δ_n, μ_n. This will not be the case for the analysis of δ_n, so the argument will me modified there.

[15]Ignoring equilibrium δ_1 because it's not directly useful to the discussion.

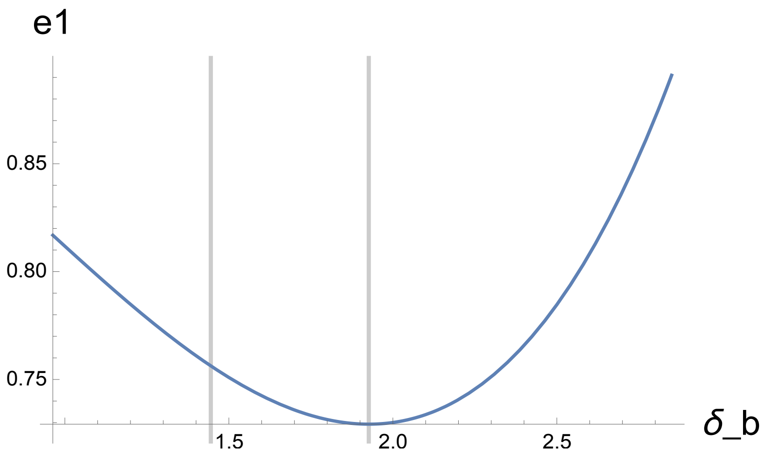


Figure 1: Illustration of Lemma 5: The vertical line that crosses the flat part of the curve corresponds to δ_b^F, while the one on the left corresponds to $\delta_b^C < \delta_b^F$, at which point investment is always decreasing in δ_b. Drawn for: $c = 11, \alpha = 2.5, \delta_n = 2.5, \mu = 0.45$.

Following a change in the parameter δ_b, the investment first-order condition must continue to hold in a new interior equilibrium, so the change in e_1 must satisfy:

$$c\frac{\partial e_1}{\partial \delta_b} = \frac{\partial \Pi_2(\delta_n, 1-\mu_n)}{\partial \mu_n}\frac{\partial \mu_n}{\partial e_1}\frac{\partial e_1}{\partial \delta_b} - \frac{\partial \Pi_2(\delta_b, 1)}{\partial \delta_b} \tag{16}$$

so that:

$$\frac{de_1}{d\delta_b} = -\left[\frac{\partial \Pi_2(\delta_b, 1)}{\partial \delta_b}\right]\left[c - \underbrace{\frac{\partial \Pi_2(\delta_n, 1-\mu_n)}{\partial \mu_n}}_{(+)}\underbrace{\frac{\partial \mu_n}{\partial e_1}}_{(-)}\right]^{-1} \tag{17}$$

and the derivative has the **opposite** sign of $\frac{\partial \Pi_2(\delta_b,1)}{\partial \delta_b}$, which we know is independent of e_1, because of perfect bad news, and positive if and only if $\delta_b > \delta_b^F$; this is due to the concavity of the profit function in the data term, for any fixed consumer beliefs[16]. This is what the Figure shows: as δ_b approaches the firm-optimum from below, the profit from achieving low reputation increases, which implies that incentives to avoid a low reputation decrease. We can state the following Lemma:

Lemma 3 *At any equilibrium of the game with exogenous δ_b, δ_n, the partial derivative $\partial e_1/\partial \delta_b$ is positive if and only if $\delta_b > \delta_b^F$, with equality at $\delta_b = \delta_b^F$. The partial derivative is given by 17.*

Now, let's turn to the slope $\partial e_1/\partial \delta_n$, which requires an argument similar to that of the δ_n-direct-effect to analyse.

[16]As will become clear later, this analysis does not rely on the assumption of perfectly revealing bad news, but exposition is simplified because of it.

For given μ_n, $\Pi(\delta, 1-\mu_n)$ is concave in δ by assumption, I have defined $\delta^F(\mu_n)$ as precisely the global maximizer of profit. Whether a change in the exogenous δ_n increases equilibrium investment, depends only on whether δ_n is larger or smaller than $\delta^F(\mu_n)$ at the original equilibrium.

Lemma 4 *Take any pair of fixed (δ_n, δ_b) that induces an interior equilibrium (e_1^*, μ_n^*). The total derivative of equilibrium investment has the same sign as the partial derivative $(\partial \Pi_2(\delta_n, 1-\mu_n)/\partial \delta_n)$, i.e. the sign of $(\partial e_1/\partial \delta_n)$ only depends on the comparison between δ_n and $\delta_n^F(\mu_n)$ at the initial equilibrium.*

Finally, what about the partial derivative, $\frac{\partial CS_2}{\partial e_1}$, that is present in both (10) and (11) derivatives? Differentiating CS_2, we obtain:

$$\frac{\partial CS_2}{\partial e_1} = (1-\mu)\left[\int_0^{\theta_n} \Big(u(\delta_n, 1) - \theta\Big) d\theta - \int_0^{\theta_b} \Big(u(\delta_b, 1) - \theta\Big) d\theta\right]$$

First, notice that even though changes in e_1 also affect equilibrium posterior reputation μ_n and thus participation decisions, there is no first-order impact of those on CS_2 because there are no externalities in this model; consumers choose participation optimally, given their information, which is on average correct, in equilibrium.

I will refer to the above as the *signal-jamming* effect of investment and I claim that evaluated at either of the "regulation-free" equilibria, it is negative. Each integral in the above expression is over consumer net utilities $(u-\theta)$ conditional on facing a Normal type, i.e. with second period probability of breach equal to 1. That conditional consumer surplus is maximized by the combination $\theta_b(\delta_b^C), \delta_b^C$, so that the integral on the right is always (weakly) larger. Intuitively, as long as lack of a data breach remains informative, i.e. as long as consumers participation and data-sharing decisions are different in the two states of period two, higher investment by the Normal type means that consumers are more frequently misguided into giving away more data than they would, if they knew for a fact that they are facing a Normal type.

A similar argument explains why the signal-jamming effect is negative when evaluated at the equilibrium with firm-control. Conditional on facing a Normal type in period 2, consumers' favourite data-sharing level is δ_b^C. In the equilibrium with firm control, it will be the case that $\delta_n^F > \delta_b^F > \delta_b^C$, so that each active consumer receives lower utility following no-breach relative to the utility following a breach (conditional on facing a Normal type). Their problem is exacerbated by the fact that more consumers are active following no breach, than the optimal level θ_b. Similar arguments reveal that the signal-jamming effect is negative whenever $\delta_n \geq \delta_b \geq \delta_b^C$.

Lemma 5 *The signal-jamming effect of investment on second-period consumers surplus is always negative at the equilibrium with consumer ex-post control, $(\delta_n^C, \delta_b^C, e_1^C)$, and the equilibrium with firm ex-post control over data-sharing, $(\delta_n^F, \delta_b^F, e_1^F)$.*

Finally, given the last few lemmas, we are ready to sign the total derivatives at each of the two equilibria.

Consider changes in either parameter, starting from the equilibrium values of the equilibrium under ex-post control by the firm. By Lemma 1, investment is unchanged by local changes to either data term, so only the direct loss to consumers remains.

Corollary 2 *Starting from the equilibrium values of data-sharing under firm-control, δ_n^F and δ_b^F, a planner can raise total consumer surplus by imposing a marginal reduction in the amount of data that the firm with either high or low reputation can ask for in period two.*

Notice the mention of *total* consumer-surplus; these data terms only affect CS_1 via investment, so there is no first-order impact on CS_1 from either change in δ_b or δ_n from their equilibrium values. Consumers simply benefit from less data-sharing in period 2, because by Lemma 1, firms ask too much data in the firm-control equilibrium.

On the other hand, looking at the total derivatives at the consumer-control equilibrium:

$$\left.\frac{dCS_2}{d\delta_b}\right|_{(\delta_n^C,\delta_b^C)} = \underbrace{\frac{\partial CS_2}{\partial e_1}}_{(-)} \underbrace{\frac{\partial e_1}{\partial \delta_b}}_{(-)} + \underbrace{\frac{\partial CS_2}{\partial \delta_b}}_{=0} > 0 \tag{18}$$

$$\left.\frac{dCS_2}{d\delta_n}\right|_{(\delta_n^C,\delta_b^C)} = \underbrace{\frac{\partial CS_2}{\partial e_1}}_{(-)} \underbrace{\frac{\partial e_1}{\partial \delta_n}}_{(+)} + \underbrace{\frac{\partial CS_2}{\partial \delta_n}}_{=0} < 0 \tag{19}$$

The difference driving the results is that an increase in δ_n from δ_n^C will increase investment (prize of high reputation increases), whereas an increase in δ_b will decrease investment (consolation prize of low reputation increases). Thus, one change causes more and the other causes less signal-jamming, and both have no first-order direct effects.

We learn that locally, CS_2 would benefit by committing to a larger δ_b, i.e. they punish too hard and share too little data with low-reputation firms, but smaller δ_n, they give out too much data to high-reputation firms.

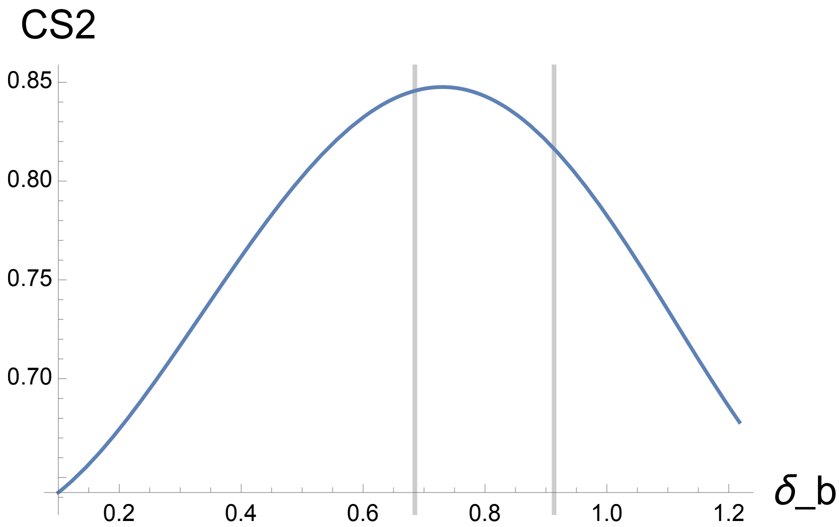


Figure 2: Fixing some δ_n, we plot CS_2 as a function of δ_b. Between the two vertical lines, there exists a local maximum of CS_2: for the given functional form assumptions, it is a global maximum. Drawn for: $c = 11, \alpha = 2.5, \delta_n = 2.5, \mu = 0.45$.

Corollary 3 *Starting at the the equilibrium under ex-post control of the consumers, the planner can increase CS_2 by imposing small caps on data-sharing for high-reputation firms, but not for low-reputation ones.*

These corollaries are meant to relate the model in this paper with the regulation around opt-out rights of consumers, which is interpreted here as consumers choosing how much of their data to share with each firm, given what they believe about the security of the firm. Corollary 2 tells us that some degree of opt-out rights is always beneficial to consumers, in terms of total consumer surplus. Corollary 3 would more reasonably be implemented by controlling the extent to which consumers to opt-out, i.e. the type of information that they can opt-out of providing, for firms that are classified as "high-risk", following a disclosed data-breach.

The two Corollaries jointly imply that for any fixed δ_n, there exists a local maximum of second-period consumer surplus at some $\delta_b^* \in (\delta_b^C, \delta_b^F)$. For the case of the specific functional forms we use, numerical output suggests that it is a global max, as seen in Figure 2.

Of course, there is a caveat, that these changes will now induce changes to CS of period 1, too. The latter, is defined as:

$$CS_1 = \int_0^{\theta_1} \left(u(\delta_1, P_1^b) - \theta \right) d\theta \tag{20}$$

so that the total derivative with respect to e_1 is:

$$\left.\frac{dCS_1}{de_1}\right|_{(\delta_n^C,\delta_b^C,\delta_1^C)} = \int_0^{\theta_1} \Bigg(\underbrace{\frac{\partial u(\delta_1, P_1^b)}{\partial \delta_1}}_{(=0)} \frac{\partial \delta_1}{\partial e_1} + \frac{\partial u(\delta_1, P_1^b)}{\partial e_1} \Bigg) d\theta$$

$$= \theta_1 \frac{\partial u(\delta_1, P_1^b)}{\partial e_1} > 0 \quad (21)$$

Once again, effects via changing equilibrium beliefs about investment and thus changing θ_1 are not of first-order; consumers make activity decisions optimally in equilibrium. In addition, the first-order effects from changing δ_1 are zero at δ_1^C. The purely positive impact of investment on CS_1 then comes from a lower breach probability[17].

Going back to CS_2 and Corollary 3, we observe that both suggested changes that improve CS2 (raising δ_b and decreasing δ_n) have the downside of decreasing e_1.

So how can we compare the induced change to CS_1 and CS_2 by e_1? Observe that at high levels of investment, θ_1 should be larger, **and** δ_1^C should be larger too, because consumers know their data is safer. With our baseline functional form for utility:

$$\frac{\partial u(\delta_1, P_1^b)}{\partial e_1} = (1-\mu)(\delta_1^C)^2 \quad (22)$$

which is higher at higher δ, and thus higher at higher levels of investment. This is intuitive; when consumers feel safe and share a lot of data with the firm, the marginal utility from additional investment rises.

On the other hand, the signal jamming effect, via which CS_2 changes, shrinks at high levels of investment. We take the second order partial derivative to see that:

$$\frac{\partial^2 CS_2}{\partial e_1^2} = \underbrace{\frac{\partial \theta_n}{\partial e_1}}_{(-)} \underbrace{\Big(u(\delta_n, 1) - \theta_n \Big)}_{(-)} > 0 \quad (23)$$

Participation is lower because e_1 lowers the high posterior reputation. Conditional on facing a Normal type, equilibrium participation levels following no-breach the equilibrium marginal consumer always has ex-post regret. Intuitively, as $e_1 \to 1$, the signal-jamming effect should shrink[18], because as $e_1 \to 1$, it implies $\mu_n \to \mu$, bringing $\theta_n(\delta_n^C)$ and δ_n^C closer to their state b corresponding quantities. Essentially at higher e_1, consumers understand

[17] When there is also an effect via δ_1, and given that $\theta_1 = u(\delta_1, P_1^b)$, we observe that CS_1 increases iff the participation cutoff does. This is the case because $u_{\theta\delta} = 0$: whenever *any* consumer benefits by the joint change in δ_1 and e_1, the marginal consumer benefits too, thus the location of the margin shifts up. For our baseline functional form, greater investment increases equilibrium CS_1.

[18] And would indeed always go to zero in a model without perfect bad news.

that a lack of data-breach is often caused by a Normal type and are cautious with participation and data-sharing in the second period. This means that there is less ex-post regret in period two when a Normal type achieves high reputation.

Lemma 6 *At higher levels of e_1^C, the negative signal-jamming effect is of lower magnitude, $\frac{\partial^2 CS_2}{\partial e_1^2} > 0$. At the same time, the positive impact of investment on CS_1 is even higher, $\frac{\partial^2 CS_1}{\partial e_1^2} > 0$. As a result, it is more likely that starting from high e_1^C, increases in e_1 are more likely to raise total consumer surplus.*

For the rest of the paper, I will revert back to settings with exogenously determined data-sharing to look at how my baseline model extends to a setting where there are two competing firms, both seeking to maintain high reputations for high data-security.

5 Duopoly

The model extends naturally to a setting with two firms, A and B, positioned at the two ends of a Hotelling line. Superscript $j \in \{A, B\}$ will denote a variable referring to firm j. I will be assuming that the firms offer the same standalone value and have the same data requirement in each period.

We introduce a transportation cost parameter β and interpret θ as the consumers' relative preference between the two platforms. In period t, user θ joins firm A iff:

$$\beta(1-\theta) - \delta_t P(s_t^A = b|I_t) \geq \beta\theta - \delta_t P(s_t^B = b|I_t) \tag{24}$$

It is assumed that v is large enough so that there is full market coverage for any pair of firm reputations. A sufficient condition is that we have full market coverage when both reputations are zero, which requires $v > (\beta/2) + \delta_2$.

We assume that there are no consumer switching costs in order to streamline the exposition, but the presence of switching costs would not change anything in the baseline model presented here[19]. The realizations of s_1^A and s_1^B are publicly observed, hence the second period information set, I_2 is the same for all users, regardless of which service they used in period 1. In addition, the random variables s_1^A and s_1^B are assumed to be *independent*, i.e. exogenous shocks arrive independently to different firms, hence there is no informational aspect to competition; realizations of s_1^A do not convey information on

[19]Since firms do not have instruments to signal their type with

the effort or type of firm B and vice versa. Thus, each firm's posterior reputation in period two is only a function of that firm's own period 1 realization and users' conjecture about its effort. For given $\tilde{e^A}, \tilde{e^B}$, the relevant second period state is now defined as the pair of firms' period 1 realizations, which can take four different values $\{b, b\}, \{n, b\}, \{b, n\}, \{n, n\}$. Replacing the inequality symbol with an equality one, (24) defines the location of the indifferent user in each of the four period-2 states. The cutoff value (which is also firm A's market share) as a function of the state is given by:

$$\theta(s_1^A, s_1^B) = \frac{1}{2} - \frac{\delta_2}{2\beta}\Big[P(s_2^A = b|I_2) - P(s_2^B = b|I_2)\Big] \tag{25}$$

The cutoff is, naturally, decreasing (increasing) in the posterior probability that firm A (B, respectively) will suffer a breach in period 2 and the posterior differential matters more when δ_2 is larger and product differentiation smaller.

Full market coverage in duopoly after every possible pair of realizations in period 1 requires $\theta_b \geq 1/2$, i.e. $2v \geq \beta + 2\delta_2$. Thus, the following assumption implies full market coverage in duopoly and incomplete market coverage (i.e. positive investment incentives) for monopoly:

$$\beta > v > \frac{\beta + 2\delta_2}{2} \tag{26}$$

This will ensure that cutoffs are always linear in posterior reputation levels and is the assumption we will be working under.

Firm A's expected[20] profit over the two periods, as a function of its own conjecture of firm B's Normal-type effort, e^B, and as a function of users' conjectures, $\tilde{e^A}, \tilde{e^B}$, is defined by:

$$\pi(e^A, \tilde{e^A}, e^B, \tilde{e^B}) = p\theta_1 + e^A p(P^B(n)\theta_{nn} + P^B(b)\theta_{nb}) \\ + (1 - e^A)p(P^B(n)\theta_{bn} + P^B(b)\theta_{bb}) - C(e) \tag{27}$$

In equation (27), $P^B(b) = P(s_1^B = b|I_1)$ is the probability with which firm B is going to suffer a breach in the first period, hence it is given by $P^B(b) = \zeta\mu_1^A + [(1-\zeta)(1-e^B) + \zeta](1-\mu_1^A)$, and $P^B(n) = 1 - P^B(b)$. I am using the shorthand θ_{nn} as the realized indifferent consumer location in the event that both firms' period 1 outcome realizations are $\{no\ breach\}$ and the other θ subscripts are interpreted accordingly; to simplify nota-

[20]Expectation taken over the probability distribution of period 2 states, which is fixed for given effort of the Normal-type competitor.

tion, I omit the fact that these cutoffs also depend on $\tilde{e^A}, \tilde{e^B}$, but those cutoffs are also equilibrium objects thanks to this dependence.

Firm A is assumed to be risk neutral, hence it chooses e^A to maximize expected profit (27), taking $\tilde{e^B}$, $\tilde{e^A}$ and e^B as given. The corresponding first order condition an optimal, interior effort level must satisfy is given by:

$$\begin{aligned} E(\theta_n) - E(\theta_b) &= \frac{c}{p} e \iff \\ (P^B(n)\theta_{nn} + P^B(b)\theta_{nb}) - (P^B(n)\theta_{bn} + P^B(b)\theta_{bb}) &= \frac{c}{p} e \iff \\ \frac{\delta_2(1-\zeta)}{2\beta}(\mu_A^n - \mu_A^b) &= \frac{c}{p} e \end{aligned} \tag{28}$$

In evaluating the probabilities of a breach that appear in the second row, consumers realize that in period two, $P(b|I_2) = P(N|I_2) + P(C|I_2)\zeta$, since Commitment types are only breached with probability ζ and Normal types never exert any effort in the second period in equilibrium. Once again, the above condition simply states that since firms are risk neutral, in an interior optimum the expected marginal benefit from increasing effort provision must equal the (deterministic) marginal cost of doing so. The left-hand side of (9) is simply the expected net increase in market share in the absence of no-breach, relative to suffering a breach. This incentive is increasing in the consumers' degree of privacy disutility, δ_2, and decreasing in the degree of horizontal differentiation, β.

The linearity of payoffs in the competitor's posterior reputation, μ_2^B means that in equilibrium, when $e_B = \tilde{e_B}$, A chooses its optimal effort level as if it is facing a competitor whose period two reputation will certainly be given by μ_1^B. In the baseline linear specification of this section, the prior μ_1^B is *also* irrelevant to effort incentives in equilibrium . We obtain equilibrium effort levels of each firm by imposing that all conjectures are correct and firms maximize profits given those conjectures.

Proposition 2 *In the duopoly version of the game, there exists a unique Perfect Bayesian Equilibrium; Normal type firms choose investment to maximize profits given users' beliefs. Users' conjectures about the investment levels of both players' Normal types are correct. Players' conjectures about the effort levels of their competitor's Normal types are also correct. In that equilibrium:*

- *If $\zeta = 0$, e^A is given by the unique positive solution to the equilibrium first-order-condition, if the latter is weakly lower than 1, which happens for $c > \frac{p\delta_2\mu}{2\beta}$. Otherwise, it is given by the corner solution $e^A = 1$.*

- *If* $1 > \zeta > 0$*, investment of A's Normal type,* e^A *is given by the unique solution to the equilibrium first-order-condition* (28) *in* $[0,1]$ *and is always strictly between* $(0,1)$.
- *If* $\zeta = 1$*, no positive investment can be supported in equilibrium,* $e^A = 0$.

The simple structure of this model allows us to immediately compare between equilibrium investments in monopoly and duopoly.

Lemma 7 *For* e^M *denoting the first-period equilibrium effort level of firm A when it is a monopolist and* e^A *the equilibrium investment under duopoly; it holds that* $e^A \leq e^M$ *for any admissible parameter configuration and, in particular,* $e^A < e^M$ *for all parameter constellations that yield unique interior solutions in both cases.*

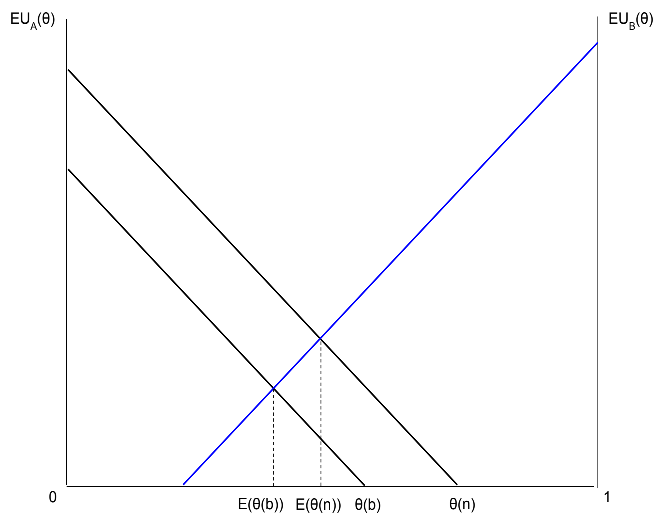


Figure 3: Competition uniformly reduces the marginal benefits of exerting effort.

Figure compares between the duopoly and monopoly scenarios. It contains all the information that firm A takes into account when making its effort decision in either setting. In the arguments below, I am keeping $\tilde{e_A}$ fixed. The vertical axes measure utility, while the horizontal axis is the [0,1] Hotelling line. The downward sloping lines are the expected utilities users receive by joining firm A after each of the two possible s_1^A realizations. In other words, they are the second period demand curves the firm at $A = 0$ faces and they determine θ_b and θ_n, the difference of which guides the monopolist's effort choice. The upward sloping line is the demand that firm A expects firm B to face in the second period.

The intersection of this upward sloping line with A's posterior demand curves determine $E(\theta_b)$ and $E(\theta_n)$, i.e. the expected market shares firm A achieves after each possible period 1 realization of its own outcome.

To explain the reduction in effort incentives under duopoly, it is sufficient to observe that, as Figure 1 suggests, the expected gain in market share from achieving a high reputation in period two is lower, i.e. $\theta_n - \theta_b \geq E(\theta_n) - E(\theta_b)$, with expectations taken with respect to s_1^B. Geometrically, whenever the competitor's posterior demand curve has slope above zero, this inequality is going to hold, hence the result holds for any parameter constellation that yields interior market shares in every state. I refer to the above as the "incremental location" effect of competition, which pushes e^A below e^M. Incentives are uniformly stronger under monopoly.

6 Regulation

Having solved the benchmark duopoly model, it is useful to examine how equilibrium investment is affected by two commonly employed policies: minimum security standards and penalties for data-breaches. In Appendix B, I also provide a very simple extension of the model to account for consumer *switching costs*; intuitively, a policy that tries to reduce those, in the spirit of the GDPR "data portability", will increase equilibrium investment.

6.1 Minimum security standards

I now examine the case in which the regulator can impose security standards e^{min}, such that every firm must have investment **at least** $e \geq e^{min}$ in each period. This immediately means that Normal types will choose $e_2 = e^{min}$. Thus, in the second period, given posterior reputations μ_2^A, μ_2^B the location of the indifferent consumer is given by:

$$\begin{aligned}\theta(\mu_2^A, \mu_2^B) &= \frac{1}{2} - \frac{\delta_2}{2\beta}\left[P(s_2^A = b) - P(s_2^B = b)\right] \\ &= \frac{1}{2} - \frac{\delta_2}{2\beta}(1 - e^{min})(1 - \zeta)(\mu_2^A - \mu_2^B) \qquad (29)\end{aligned}$$

so that the equivalent of first-order condition (28) for (interior) optimal investment is given by:

$$(1 - e^{min})(1 - \zeta)(\mu_n^A - \mu_b^B) = \gamma e \qquad (30)$$

where $\gamma := 2\beta c/p\delta_2$. Now assume that the first-period equilibrium investment is $e_1 > 0$ when there are no security standards imposed. Then, marginally increasing e^{min} from

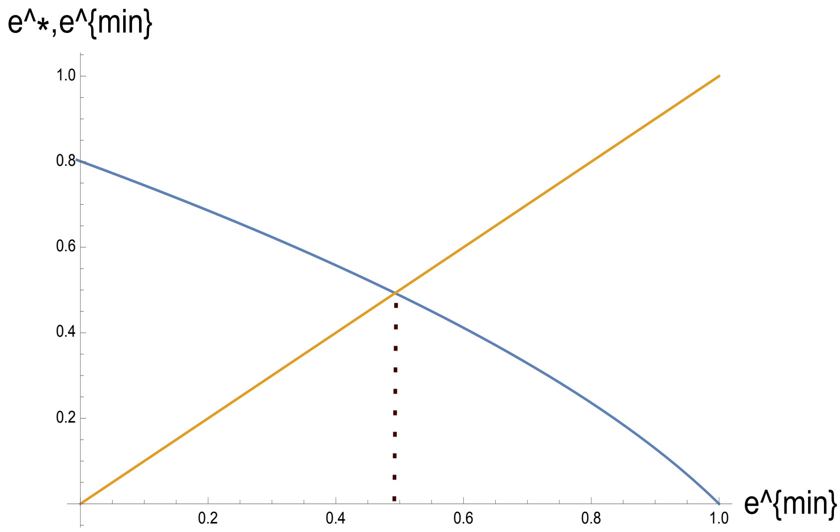


Figure 4: First-period investment is always given by the *maximum* of e^*, e^{min}. When the regulator cannot impose high-enough minimum standards, first-period investment can be **decreasing** in the level of standards.

its initial value of zero will mean that the security standards only bind the firm in the second period, so that in the new equilibrium $e_1(e^{min}) > e^{min}$. As seen in the first-order condition, an imposed security standard allows the firm to commit to a higher security level in the second period, hence **reduces** the harm from being exposed as a Normal type at the end of the first period. Of course, for sufficiently high e^{min}, the security standards will be binding in the first period too. The following Lemma and Figure 4 illustrate the effect of e^{min} on equilibrium first-period investment.

Lemma 8 *Suppose the regulator can impose minimum security standards at a level* $e^{min} \in [0, 1]$ *and when* $e^{min} = 0$*, equilibrium first-period investment is given by* $e(0)$*. For* $e^{min} > 0$ *the new first-period investment will only be higher than* $e(0)$ *if* $e^{min} > e(0)$*. Furthermore, there exists a unique value* e^{**}*, such that first-period investment is* ***decreasing*** *in* e^{min} *if and only if* $e^{min} < e^{**}$*.*

6.2 Penalties on breached firms

We saw that reputational dynamics imply that minimum security standards have commitment value to the firm and thus produce some potentially unpleasant reduction of equilibrium security investment. Another common policy prescription is fines for firms that suffer data breaches. Fines, by ensuring that breached firms are punished even in the second period, also have commitment value to firms which become less reliant on maintaining high reputations of being Commitment types. More concretely, assuming that the planner levies fine $F \geq 0$ to a firm that suffers a data breach, second period effort by a Normal type firm will be $e_2 = min\{1, (1-\zeta)F/c\}$. The firm's marginal benefit of investment in the first period for given consumers' conjectures is given by:

$$F + E(\Pi_n(F, e_2(F))) - E(\Pi_b(F, e_2(F))) \tag{31}$$

where the expectation is taken with respect to the competitor's outcome and I am being explicit about the dependence of second-period profits on F. A larger fine F has a direct, positive effect on investment incentives, and an indirect, negative one, via reducing reputational concerns, since users know that even a Normal type firm will invest in security in the second period[21]. As it turns out, the positive effect always dominates in this model:

Lemma 9 *When the planner can levy fine F on firms that suffer data breaches, second-period equilibrium investment by the Normal type is given by $e_2 = 1$, if $F > c/(1-\zeta)$, and by $F(1-\zeta)/c$, otherwise. In the first period, if $F > c$, then $e_1 = 1$, otherwise there exists unique equilibrium effort $e(F) \in (0, 1)$, which is always increasing in F.*

Note that the presence of both direct and reputation incentives in the first period means that there is a a lower threshold for the fine that makes that period's investment hit the corner solution $e^A = 1$.

7 Conclusion

In a two-period model, I examine the incentives of a digital service monopolist to invest in unobserved data security, when it charges no access fees but instead monetizes consumer data. Consumers suffer privacy-related disutility when data-breaches occur, and the firm wants to earn a reputation for protecting users' data to maintain high activity in period two. I analyse two regimes of endogenous data-sharing, depending on which side has ex-post control over it: if it is the firm, data-sharing requirements are chosen in every period to maximize current profits. If it is consumers, data-sharing is chosen to maximize consumer surplus, accounting for the firm's reputation. I ask whether a social planner can improve ex-ante consumer surplus by committing to different levels of data-sharing in

[21] The direct effect of F on Π_n and Π_b is the same since optimal e_2 is the same in every subgame.

period two, relative to the regulation-free equilibria, and I allow data-sharing to depend on the firm's posterior reputation.

Ex-ante commitment to data-sharing affects consumer surplus directly, but also via equilibrium investment. Starting at the firm-control equilibrium, the effects on investment are dominated, and the planner can improve total CS by reducing the amount of data that both high and low reputation firms collect. On the other hand, compared to the ex-post consumer optimum, committing to less data-sharing following a breach induces higher security; the ex-ante optimal level trades-off higher security and more "signal-jamming": greater investment impedes learning about the true levels of cyber-risk which harms consumers in the second period.

Currently, I am primarily working on further understanding the planner's problem in Section 5, in particular features of the global solution of ex-ante commitment to data-sharing levels of period two; the problem is not concave, hence this is not trivial. At the same time, I want to understand in what environments of endogenous data-sharing giving consumers ex-post control, as opt-in regulation does, will actually improve consumer surplus relative to the case of firm ex-post control. Regarding different policies, I am working on how breach-penalties that depend on the levels of data-sharing can improve welfare. Finally, and as shown in the final section and the Appendix, I want to understand what features of this analysis transfer to a duopoly setting, in which firms compete in their data-sharing "offers" to consumers, i.e. looking at the firm-control regime under duopoly. 'Price-like' competition in data should reduce the difference between the equilibrium levels of data sharing in each of the two-regimes.

References

Guy Aridor, Yeon-Koo Che, and Tobias Salz. The effect of privacy regulation on the data industry: empirical evidence from GDPR. *RAND Journal of Economics*, 54(4):695–730, December 2023. doi: 10.1111/1756-2171.12455. URL `https://ideas.repec.org/a/bla/randje/v54y2023i4p695-730.html`.

Shinichi Kamiya, Jun-Koo Kang, Jungmin Kim, Andreas Milidonis, and René M. Stulz. Risk management, firm reputation, and the impact of successful cyberattacks on target firms. *Journal of Financial Economics*, 139(3):719–749, 2021. URL `https://EconPapers.repec.org/RePEc:eee:jfinec:v:139:y:2021:i:3:p:719-749`.

Tesary Lin. Valuing intrinsic and instrumental preferences for privacy. *Marketing Science*, 41, 2022. URL `https://doi.org/10.1287/mksc.2022.1368`.

Bruno Jullien, Yassine Lefouili, and Michael Riordan. Privacy Protection, Security, and Consumer Retention. *TSE Working Papers*, August 2020. URL `https://ideas.repec.org/p/tse/wpaper/32902.html`.

Avi Goldfarb and Catherine Tucker. Introduction to The Economics of Privacy . May 2023. URL `https://ideas.repec.org/h/nbr/nberch/14786.html`.

Alessandro Acquisti, Curtis Taylor, and Liad Wagman. The economics of privacy. *Journal of Economic Literature*, 54(2):442–92, 2016. URL `https://EconPapers.repec.org/RePEc:aea:jeclit:v:54:y:2016:i:2:p:442-92`.

Alexandre de Cornière and Greg Taylor. A Model of Information Security and Competition. TSE Working Papers 21-1285, Toulouse School of Economics (TSE), December 2021. URL `https://ideas.repec.org/p/tse/wpaper/126391.html`.

Toni Ahnert, David Cimon, and Ryan Riordan. Cyber security and ransomware in financial markets. *CEPR Press Discussion Paper No. 17403.*, 2022. doi: https://cepr.org/publications/dp17403.

Itay P. Fainmesser, Andrea Galeotti, and Ruslan Momot. Digital Privacy. *Management Science*, 69(6):3157–3173, June 2023. doi: 10.1287/mnsc.2022.4513. URL `https://ideas.repec.org/a/inm/ormnsc/v69y2023i6p3157-3173.html`.

Sarit Markovich and Yaron Yehezkel. "for the public benefit": who should control our data? Working Papers 21-08, NET Institute, 2021. URL `https://EconPapers.repec.org/RePEc:net:wpaper:2108`.

Pantelis Koutroumpis, Farshad Ravasan, and Taheya Tarannum. (under) investment in cyber skills and data protection enforcement: Evidence from activity logs of the uk information commissioner's office. *Working Paper*, July 2022. URL `https://ssrn.com/abstract=4179601`.

Bengt Holmström. Managerial incentive problems: A dynamic perspective. *Review of Economic Studies*, 66(1):169–182, 1999.

A Benchmark Model

A.1 Proposition 1

Case 1: $\zeta = 0$. The monopoly equilibrium first-order condition is:

$$e^2 + \frac{\mu}{1-\mu}e - \frac{\delta_2 p}{c\beta}\frac{\mu}{1-\mu} = 0 \tag{32}$$

Define the LHS as $g(e)$, which has the same sign as $MC(e) - MB(e)$. Then, if $g(e) = 0$ has a unique solution which lies in (0,1), there exists a unique optimal effort level, which is interior. The constant term in the quadratic polynomial is negative, hence the product of solutions is negative, i.e. there is at most 1 positive solution. Thus, showing that a solution exists in (0,1) is sufficient and necessary for our purpose. Since g is continuous, I can appeal to the Intermediate Value Theorem. $g(0) = -\frac{p\delta_2}{c\beta}\frac{\mu}{1-\mu} < 0$, while $g(1) = 1 + \frac{\mu}{1-\mu} - \frac{p\delta_2}{c\beta}\frac{\mu}{1-\mu}$. Since $g'(e)$ is strictly positive for every non-negative e, $g(1) > 0$ is necessary and sufficient for the existence of an interior solution. The latter inequality requires $\frac{c}{p\delta_2} \geq \frac{\mu}{\beta}$. Thus, assuming high enough cost parameter relative to the payoff parameter is sufficient for a unique interior optimum. If the above condition is not satisfied, we get a corner solution $e^*_S = 1$, since that would imply $g(e) < 0$, $\forall e \in [0,1]$. The second-order condition for the Normal type's maximization problem is very easy to check; Holding $\tilde{e_A}$ fixed, the second derivative of expected profits is simply $\frac{\partial^2 E\pi^A}{(\partial e)^2} = -c$, indicating a maximum. This will be the case in every model presented with quadratic costs, so checking second order conditions will be omitted in the rest of the proofs.

Case 2: If $1 > \zeta > 0$, deriving the analytical solution is tricky; However, we have already discussed in the main text that for given beliefs, the marginal benefit of effort is decreasing in users' beliefs even when $\zeta > 0$; that is because higher beliefs both **decrease** the posterior following a no-breach outcome and **increase** the posterior following a breach outcome. Thus, in the equilibrium first-order condition:

$$\frac{\delta_2(1-\zeta)}{\beta}(\mu_n - \mu_b) = \frac{c}{p}e \tag{33}$$

the left hand side is decreasing in e, while the right-hand side is increasing. Further, at $e = 0$, the left-hand side is positive, $\mu^n = 1$, $\mu^b = \mu\zeta\big[\mu\zeta + (1-\mu)\big] < 1$. At $e = 1$, the left hand side is equal to zero, since the two types have equal chances of getting breached and there is no reputational gain from avoiding a breach, $\mu^n = \mu^b = \mu$, whereas the right-hand side is positive. By continuity, there must be a unique intersection point e^* in (0,1), verifying the claim in Proposition 1.

A.2 Lemma 8: Minimum security standards

Given that $e_2 = e^{min}$, if the security standard was **not** applied to the first-period investment, the first-order condition for optimal investment in the first period would be given by:

$$(1 - e^{min})(1-\zeta)(\mu^n - \mu^b) = \gamma e \tag{34}$$

where $\gamma := \beta c/\delta$ and p is normalized to 1. Proceeding in the same way as in Proposition 1, the FOC has, for every $e^{min} \in [0, 1)$, a unique solution $e^*(e^{min})$ within $[0, 1]$, and it is strictly between $(0, 1)$. If the minimum security standard only applied to the second period, it would be that $e_1 = e^*(e^{min})$ and the latter is clearly *decreasing* in e^{min}. When the standard is also applied to first-period investment, there are two candidate equilibrium values e_1 in this setting:

- $e_1 = e^*(e^{min}) > e^{min}$, where e^* is the unique solution that the FOC admits in $[0, 1]$. It would be a mistake to say that in this case security standards have no impact on firm behaviour in period 1, the solution e^* clearly depends on e^{min}.
- $e_1 = e^{min}$, in case $e^* < e^{min}$.

As the planner imposes a higher standard e^{min}, we get that both (1) the solution e^* decreases, and also (2) for given e^*, the security standard is more likely to bind in the first-period too; hence we are more likely to land in the second type of equilibrium. More formally:

Step 1: For every level of $e^{min} \in [0, 1)$, the equilibrium first-order condition has a unique solution $e^*(e^{min})$ within $[0, 1]$, which is strictly between $(0, 1)$.

Step 2: Defining $G(e, e^{min}) := (1 - e^{min})(1 - \zeta)(\mu_n - \mu_b) - \gamma e$, the solution $e^*(e^{min})$ is decreasing in e^{min} via the IFT. This is because the difference in posterior reputations is positive and decreasing in equilibrium investment.

Step 3: Define $H(e^{min}) := e^*(e^{min}) - e^{min}$ and we just argued that H is decreasing in its argument. From Step 1, we also know that $H(0) > 0$ and we also know that H is negative as $e^{min} \to 1$. Continuity and monotonicity imply there exists a unique e^{**}, such that $H(e^{**}) = 0$.

Thus we know that $e^* > e^{min}$ i.e. $H > 0$ iff $e^{min} < e^{**}$. The solution e^{**} is **smaller** than first-period investment in the absence of security standards, $e^*(0)$. First-period investment falls, while the planner increases security requirements from zero to e^{min} and start increasing afterwards. Starting from an equilibrium $e^*(0)$, the only way to increase first-period investment using security standards is by imposing security standards $e^{min} > e^*(0)$.

A.3 Lemma 9: Impact of a fine

In the second period, the firm maximizes $\pi_2 = P(b|N)(-F) - C(e_2)$, which has first order condition for interior optimum $(1 - \zeta)F = ce_2$, and the result follows. In the first period, the marginal benefit of investment for given users' beliefs is:

$$(1 - \zeta)F + \frac{\delta_2}{2\beta}(1 - e_2)(1 - \zeta)(\mu^n - \mu^b)$$

Case 1: If $e_2 = 1$, i.e. if $F > c/(1-\zeta)$, first-period marginal benefit is equal to $(1-\zeta)F > c$, so that a corner solution also arises for e_1.
Case 2: If $e_2 < 1$, i.e. $(1 - \zeta)F < c$, $e_1 = 1$ can only be an equilibrium if $c < F$. Thus, for $c \in [(1 - \zeta)F, F]$, we have a corner solution $e_1 = 1$.
Case 3: Finally, if $c < (1 - \zeta)F$, e_1 is given by the unique solution that the first-order condition admits in $[0, 1]$, $e(F)$, which we also know will be strictly within $(0, 1)$.

Define:

$$H(e,F) := (1-\zeta)F + \frac{\delta_2}{2\beta}(1 - \frac{F(1-\zeta)}{c})(1-\zeta)(\mu^n - \mu^b) - ce \tag{35}$$

with $H(e(F), F) = 0$ for all $F \in [0, c(1-\zeta))$. In equilibrium:

$$\tau(\mu^n - \mu^b) = (ce(F) - F)(1 - \frac{F(1-\zeta)}{c})^{-1}$$

where:

$$\tau := \frac{\delta_2(1-\zeta)}{\beta}$$

By the Implicit Function Theorem:

$$\frac{de(F)}{dF} = -\frac{H_F}{H_e}$$

where $H_e < 0$. So, the derivative is positive if and only if

$$\begin{aligned}
H_F > 0 &\iff \\
1 + \frac{\delta_2}{2\beta}(\frac{-(1-\zeta)}{c})(1-\zeta)(\mu^n - \mu^b) > 0 &\iff \\
c > \tau(1-\zeta)(\mu^n - \mu^b) &\iff \\
c > (1-\zeta)(ce(F) - F)(1 - \frac{F(1-\zeta)}{c})^{-1} &\iff \\
c - F(1-\zeta) > (1-\zeta)(ce(F) - F) &\iff \\
c > (1-\zeta)ce(F) &\iff \\
1 > (1-\zeta)e(F)
\end{aligned}$$

which always holds since both terms of the RHS are weakly smaller than 1.

B Additional Material

B.1 Switching Costs and the GDPR

In this subsection, we want to think about the *data-portability* requirement that the EU GDPR imposes on data controllers (firms that store user data, like the firms we model in this paper). Portability demands that a user that wants to *migrate* to a different platform, can ask for his current service provider to hand their data in machine-readable format that can be transferred to a different service provider. For instance, a user might want to transfer all their pictures from one-social media app to the other, or all their posts from one blog to the other, etc. In order to study the impact of this GDPR requirement, we must slightly modify this model to incorporate costs of switching from one platform to the other in the second period. We think of symmetric firms, meaning μ^A and μ^B are equal, and introduce a simple version of switching costs: a consumer that wants to *switch* websites in period two, must now suffer some *disutility* equal to χ.

We focus on symmetric equilibria, in which each firm's Normal type chooses the same effort level, thus period 1 market is split between the two firms. In this case of a symmetric equilibrium, the only case in which there might[22] be a positive mass of switching consumers from one firm to the other in period two is if the two firms have different outcome realizations in the first period. In that case, the equivalent first order condition of (9) for the symmetric duopoly with switching costs will be given by:

$$-\frac{\chi}{2\beta} + \frac{\delta_2(\mu_n^A - \mu_b^A)}{2\beta} = \frac{c}{p}e \tag{36}$$

Lemma 10 *In a symmetric equilibrium with $\zeta = 0$, $\mu^A = \mu^B$ and "small" switching costs, $\chi > 0$: equilibrium investments are weakly decreasing in χ.*

It should be easy to understand why positive switching costs always reduce the expected marginal benefit to investment provision and thus equilibrium investment incentives. The existence of $\chi > 0$ decreases both the potential expected gain from achieving high reputation and the potential expected loss from ending up with a low reputation as a result of a breach. Hence, policies like the EU GDPR which requires portability and facilitates switching between service providers will in this model also bring higher equilibrium investments. It is work in progress to examine the impact of switching costs and their elimination in asymmetric equilibria, so that we can see whether both dominant and low market share firms will increase investments and if so, which firm will see the biggest increase in investments following the implementation of a data portability policy.

C Duopoly with endogenous data

I extend the benchmark model to the case where two firms choose their δ_t^i values *simultaneously* in each period. In other words, I only examine on the firm-control regime. In this section, which is currently work in progress, I solve for the equilibrium of this game and

[22] If χ is large enough relative to β, no consumer ever switches. We focus on equilibria in which at least some consumers switch when the firms have unequal reputations in period two.

discuss the consumer-surplus impact of imposing data caps to breached firms in period 2, relying mostly on numerical output. The timing of the game is described below:

1. In $t = 1$, firms A and B simultaneously choose their observable data storage policies, δ_1^A, δ_1^B and consumers choose to join the firm that provides the highest expected utility.

2. Still in $t = 1$, firms A and B choose their unobserved levels of investment; at the end of the period, each firm suffers a breach or doesn't.

3. In $t = 2$, given their posterior reputations, firms simultaneously make their δ_2^A, δ_2^B choices to maximize expected profits and consumers join the firm that provides the highest expected utility.

4. Each Normal type firms suffers a data breach with probability 1 and consumers incur utility losses.

Again, since there are no adjustment or consumer switching costs, Normal and Commitment type firms have the same optimal δ choices at every period and state.

C.1 Regulation-free equilibrium

Given posterior reputations μ_2^A, μ_2^B and given the δ_2^B choice by the competitor, firm A chooses its own data storage policy to maximize $\Pi = (p_0 + p_1\delta^A) * \theta(\delta^A, \delta^B, \mu_2^A, \mu_2^B)$. As a reminder, the indifferent consumer location is given by:

$$\theta = \frac{1}{2} - \frac{1}{2\beta}\left[\delta^A(1-\mu^A) - \delta^B(1-\mu^B)\right]$$

Solving the first-order condition yields:

$$\delta^A = BR^A(\delta^B) = \frac{1}{2}\left[\frac{\beta + \delta^B(1-\mu_2^B)}{1-\mu_2^A} - \frac{p_0}{p_1}\right] \tag{37}$$

and the second-order condition is always satisfied. The comparative statics are intuitive; the best-response increases in the firm's own reputation, and decreases in competitor's. In addition, the firm increases its data requirement when p_1/p_0 is higher, i.e. the relative importance of data for per-user revenue rises. The interior equilibrium of the data-choice game is found at the intersection[23] of best-responses:

$$\delta^A(\mu_2^A, \mu_2^B) = \frac{3p_1\beta - p_0(1-\mu_2^B)}{3p_1(1-\mu_2^A)} - \frac{2p_0}{3p_1} \tag{38}$$

[23] Why is the best-response of A decreasing in μ^B but the equilibrium value increasing in that variable? A necessary but not sufficient condition is that data requirements are **strategic complements**. The total

and plugging $\delta^A(\mu_2^A, \mu_2^B), \delta^B(\mu_2^A, \mu_2^B)$ into the expression for the consumer cutoff, we find that:

$$\begin{aligned}\theta(\mu_2^A, \mu_2^B) &= \frac{1}{2} - \frac{1}{2\beta}\left[\delta^A(1-\mu_2^A) - \delta^B(1-\mu_2^B)\right] \\ &= \frac{1}{2} - \frac{1}{6\beta}\frac{p_0}{p_1}(\mu_2^A - \mu_2^B)\end{aligned}$$

We can thus define the value of reaching state (s^A, s^B) for firm A as $\Pi_{s^A,s^B} = (p_0 + p_1\delta^A(\mu_2^A, \mu_2^B))\theta(\mu_2^A, \mu_2^B)$. Effort beliefs determine posterior reputations in period two; for given consumer beliefs, equilibrium profits Π are pinned down for every state of period two. Given those, and given investment e^B by the competitor, firm A chooses its optimal investment level to maximize expected profit, given consumer beliefs about both firms' and given the competitor's investment level. The first-order condition for e_1 is $P^B(n)(\Pi_{nn} - \Pi_{bn}) + (1-P^B(n))(\Pi_{nb} - \Pi_{bb}) = C'(e_A)$, where $P^B(n) = \mu_1^B + (1-\mu_1^B)e^B$.

An *equilibrium* of this game, is a pair of investment levels such that **(a)** posterior reputations for firms that do not get breached in period 1 are given by (1), **(b)** data requirements in period 2 are given by 38 and **(c)** the investment focs are satisfied for both firms.

C.2 Equilibrium with data caps

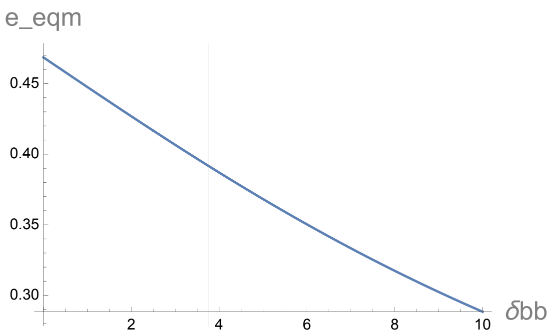
Our goal in this section, similar to the corresponding section for monopoly, is to understand whether a planner that cares about consumer surplus can benefit by imposing limits on the amount of data that firms can ask for in period 2, if they suffered a data breach at the end of period 1. A difference with our treatment of monopoly is that the planner could now also, in principle, condition the data limit on whether the *competitor* also suffered a breach or not.

We will only focus on games between symmetric firms and focus on *symmetric* equilibria of the game with second-period data-caps. The planner has two policy levers, $\delta_{bn} \geq 0$ and $\delta_{bb} \geq 0$. The first, is the second-period data requirement that is imposed on a firm which suffers a breach when the competitor did not suffer a breach, and the second is the data requirement imposed on both firms when they both suffer breaches in the first

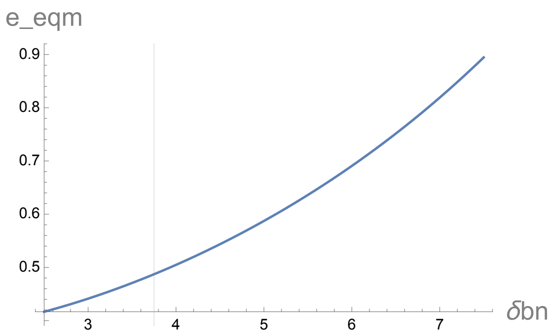
derivative, accounting for the opponent's reaction, is positive if:

$$\frac{\partial^2\Pi^B}{\partial\delta_B^2}\frac{\partial^2\Pi^A}{\partial\delta_A\partial\mu^B} - \frac{\partial^2\Pi^A}{\partial\delta_A\partial\delta_B}\frac{\partial^2\Pi^B}{\partial\delta_B\partial\mu^B} < 0$$

The above is the case whenever $p_0 > 0$, consistent with the solution in (38).



(a) Symmetric equilibrium investment is *decreasing* in δ_{bb}. Figure drawn for $\mu = 0.5, c = 3, \beta = 2.5, p_1 = 1, p_0 = 0$. Vertical line corresponds to $\delta_{bb} = \beta$, the equilibrium value in the game without regulation.



(b) Symmetric equilibrium investment is *increasing* in δ_{bn} for most parameter combinations. Figure drawn for $\mu = 0.5, c = 3, \beta = 2.5, p_1 = 1, p_0 = 0$. Vertical line corresponds to $\delta_{bn} = \beta$, the equilibrium value in the game without regulation.

period. Thus, the only state of period 2 in which both firms are able to choose their data requirements is $\{nn\}$. Removing the firm-superscirpts since we focus on symmetric equilibrium, the data requirements in equilibrium will be given by:

$$\delta_{nn} = \delta(\mu_n, \mu_n) = \frac{3p_1\beta - p_0(1-\mu_n)}{3p_1(1-\mu_n)} - \frac{2p_0}{3p_1} \tag{39}$$

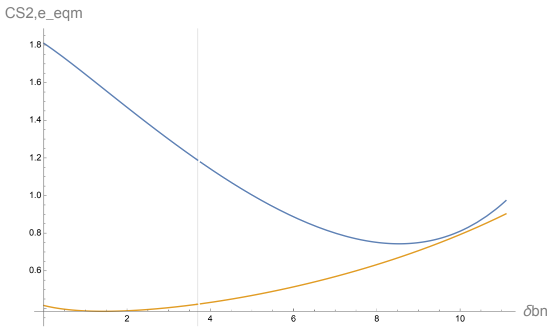
$$\delta_{nb} = BR(\delta_{bn}) = \frac{1}{2}\Big(\frac{\beta + \delta_{bn}}{1-\mu_2} - \frac{p_0}{p_1}\Big) \tag{40}$$

$$\delta_{bn} = \delta_{bn} \tag{41}$$

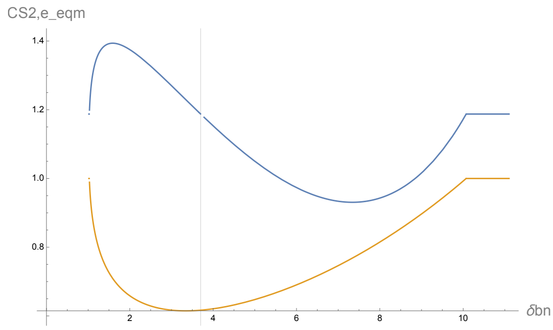
$$\delta_{bb} = \delta_{bb} \tag{42}$$

In other words, in state $\{nb\}$, firm A simply plays its best-response to firm B, which plays the data storage level δ_{bn} imposed by the planner. The investment first-order condition is derived in the way discussed already and equilibrium defined similarly.

In the duopoly setting, there is no equivalent of Lemma 5 to facilitate local arguments. But the comparative statics of investment in the symmetric equilibrium with respect to the planner's choices are simple to interpret. In any symmetric equilibrium, $\Pi_{bb} = p_1\delta_{bb}/2$. Hence, increasing that parameter will always increase expected profit in period two following a breach in period 1, and thus decrease the expected benefit to investment, holding e_B fixed. On the other hand, increasing δ_{nb} affects firm A via two channels: when it suffers a breach and B does not, and A is forced to play $\delta^A = \delta_{bn}$ in period 2; but also when the roles are reversed, and it is A that plays its best-response to $\delta^B = \delta_{bn}$. For the first case, Π_{bn} is initially increasing in δ_{bn} but eventually becomes decreasing. For the second



(a) Blue: second-period consumer surplus. Orange: investment at symmetric equilibrium. Figure drawn for $\mu = 0.5, c = 4.5, \beta = 3.7, p_1 = 1, p_0 = 0$ and $\delta_{bb} = \beta$.



(b) Blue: second-period consumer surplus. Orange: investment at symmetric equilibrium. Figure drawn for $\mu = 0.5, c = 3, \beta = 3.7, p_1 = 1, p_0 = 0$ and $\delta_{bb} = \beta$.

case, since data is always harmful to consumers, higher δ_{bn} always increases Π_{nb}. It's easy to see from the investment FOC, that when both profits Π_{bn}, Π_{bb}, i.e. the two possible profits following a breach of A, are decreasing in δ_{bn}, investment incentives are increasing in that parameter.

We now look at the impact of data caps on consumer surplus, focusing on the case[24] of $p_0 = 0$. Given the functional forms and under $p_0 = 0$, there is no impact on first-period consumer surplus, CS_1, from changes in either parameter; there are no direct effects, because the policies only take effect in period 2, and the induced changes in e which mean fewer breaches are exactly offset by the changes in equilibrium δ_1.

Below, we provide numerical evidence that holding δ_{bb} fixed at the no-regulation equilibrium value of β, CS_2 consumer surplus may be either maximized by $\delta_{bn} = 0$, or, for low c/p_1, by some *interior* value of $\delta_{bn} \in (0, \beta)$. In both cases, and robust across numerical simulations, CS_2 is always *decreasing* in the parameter at the no-regulation equilibrium value of $\delta_{bn} = \beta$. Thus, the regulator can always do better for consumers by imposing limits at a level lower than the laissez-faire equilibrium, similar to our finding for the monopoly case. Finally, I provide a numerical example in which the regulator varies $\delta_{bn} = \delta_{nb} = \delta_b$ as a single data cap for a firm that suffers a breach regardless of its competitor's outcome. For the vast majority of parameter combinations, CS_2 is decreasing in $\delta_b \in (0, \beta)$, thus the CS-maximizing value is again zero. Numerical simulations again confirm that decreasing

[24]When $p_0 = 0$, equilibrium values of δ_2 do not depend on the competitor's reputation in that period. Hence, $\delta_{bn} = \delta_{bb} = \beta$ at the symmetric equilibrium. Under $p_0 = 0$, p_1 does not affect δ choices and only enters the investment foc via c/p_1, thus can be normalized at 1.

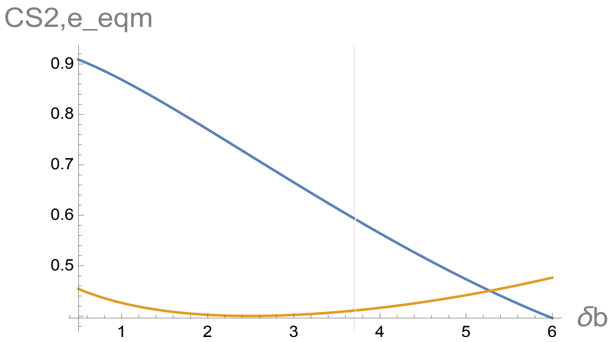


Figure 7: Blue: (scaled) second-period consumer surplus. Orange: investment at symmetric equilibrium. Figure drawn for $\mu = 0.5, c = 3.5, \beta = 3.7, p_1 = 1, p_0 = 0$. For most parameter combinations, δ_b maximizes CS when set to $\delta_b = 0$.

δ_b away from the laissez-faire equilibrium is *always* locally improving CS[25].

[25]Note that we are able to refer to a single laissez-faire equilibrium δ_b because of the $p_0 = 0$ assumption which makes equilibrium choices independent of the competitor's reputation, thus $\delta_{bn} = \delta_{bb} = \delta_b = \beta$ in equilibrium.