Incentivizing Firms to Protect Consumer Data: Can Reputation Play a (Bigger) Role?

Ying Lei Toh*

November 6, 2017

Abstract

Despite the growing threat of cyber-attacks, firms may lack incentives to invest in consumer data protection due to two market failures: imperfect information and externalities. I examine the mitigating role that reputation may play in a repeat-purchase setting—data breaches can damage a firm's reputation, leading to lost future sales. I consider policies for boosting investment and analyze their impact on the security level and consumer surplus. I show that interventions which directly address the market failures always lead to desirable outcomes, while indirect interventions aimed at enhancing the role of reputation may result in lower investment and consumer surplus.

*Toulouse School of Economics. Email: . First and foremost, I thank my thesis advisor, Bruno Jullien for his patient guidance and support. I would also like to extend special thanks Yassine Lefouili, Jacques Crémer and Robert Hunt for their very helpful discussions, comments and advice. Finally, my gratitude goes towards Rossella Argenziano, Alessandro Bonatti, Alexandre de Cornière, Miaomiao Dong, Simon Fuchs, Holger Herbst, Boqian Jiang, Zhaoxin Pu, Michael Riordan, Wilfried Sand-Zantman, Ananya Sen, Giridaran Subramaniam, the members of the applied theory group at TSE, seminar participants at Carnegie Mellon University, the University of Washington, the Federal Reserve Bank of Philadelphia and the University of Mannheim, conference participants at the 32nd Congress of the EEA (Lisbon), 3rd IO in the Digital Economy Workshop (Liège), the 43rd EARIE Conference (Lisbon), the 10th Bi-annual Postal Economics Conference (Toulouse) and the anonymous reviewers and participants at the 15th Workshop on the Economics of Information Security (Berkeley) for their feedback and comments. The financial support of the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation programme (grant agreement No 670494) is also gratefully acknowledged.

1 Introduction

> *"Today's organizational crown jewels are built of bytes. Back in 2003, physical security was most important to secure a company's most valuable information or goods. Today, everything is on a network—even medical records."*
>
> *—Ablon et al. (2014), p.34*

We live in an increasingly networked world. As the volume of digital data grows, so has the incidence of data thefts. In 2016 alone, there were 1,792 reported data breaches worldwide; these led to the compromise of close to 1.4 billion records (Gemalto, 2016).[1] With data breaches hitting major industry players such as Equifax, Target, Home Depot and Yahoo!, it is likely that most of us have been "pwned" at least once in recent years.[2,3] The rising number of breaches is a worrying trend. Data exposed in breaches—especially personally identifiable information—can be used to commit identity thefts and payment fraud; these may result in financial losses, inconvenience and distress for their victims.

The prevalence of data breaches underlines the difficulty of data protection in this digital age. Not only are firms faced with increasingly sophisticated cyber-threats, malicious parties can now easily gain access to cyber-crime tools, thanks to the emergence of dark-net marketplaces. That being said, technological solutions in themselves may not be enough. Data security is as much of an economic problem as a technical one—even computer scientists recognize that "security failure is caused at least as often by bad incentives as by bad design" (Anderson and Moore 2006, p.610). Surveys and anecdotal evidence suggest that investment incentives are indeed poor.[4] Therefore, a two-pronged approach to addressing the data security problem is necessary: developing better defense technology and creating stronger investment incentives. In order to formulate effective policies for encouraging investment, policy makers need to first understand the investment motives of firms. Yet, despite its importance, this economic aspect of the data security problem has received little attention in the academic literature. My work aims to fill this gap.

Consider the security investment problem of a firm. Its incentives to invest may be weak due to two market failures: imperfect information and investment externalities. It is clear that if the firm's state of security cannot be observed by consumers at the point of purchase (imperfect information) and if it bears neither direct costs

[1] Over 9 billion records have been lost or stolen in data breaches since 2013. For the latest figures, visit `http://breachlevelindex.com/`.

[2] The verb "pwn" is an informal term (that is used especially in the context of video gaming) which means to "utterly defeat (an opponent or rival); completely get the better of" (Source: Oxford dictionaries, s.v. "pwn", `https://en.oxforddictionaries.com/definition/pwn`).

[3] To check if you have been "pwned", visit `https://haveibeenpwned.com/`.

[4] For instance, in a recent survey of small and medium enterprises in the United Kingdom, 49% of respondents indicated that they intend to invest less than £1,000 in cyber-security in the next 12 months (Source: Josie Cox, "Small and medium-sized businesses are not investing in cyber protection despite spate of attacks", *The Independent*, July 31, 2017, `http://www.independent.co.uk/news/business/news/sme-cyber-protection-attacks-hackers-small-businesses-medium-sized-security-online-wannacry-a7868426.html`).

nor liability from a data breach (externalities), it will not invest in a one-shot interaction. However, in many markets, consumers and firms interact repeatedly. In these repeat-purchase settings, reputation concerns may provide incentives for a firm to invest. Because data security is initially unobserved, consumers often base their decision to buy (in part) on the firm's reputation for security. By not investing in data security, the firm exposes itself to a high risk of breaches. Its reputation suffers whenever consumers learn of these breaches, and it may lose future business as a result.[5] That said, can reputation concerns provide (enough) incentives for a firm to invest in security? And if not, what measures can be taken to improve its investment incentives? These are the questions that I seek to address.

To answer these questions, I develop a model of security investment that incorporates the key elements discussed above: imperfect information, investment externalities and customer turnover. I consider a two-period setting where a website (an on-line merchant) sells to a representative consumer. The website can make a one-time security investment at the start to protect its customer's payment data from potential thefts. Because these thefts can result in fraud losses, the consumer cares not only about the quality of the product sold when buying from the website, but also about the security of his payment data. The overall value of the website's offering is thus determined by both components.

Imperfect information. I focus on the case where product quality is exogenous (from the website's perspective) and is perfectly observed by the consumer. Data security, by contrast, is endogenously determined by the website's investment and initially unobserved by the consumer. The consumer has rational beliefs about the website's state of security; these beliefs constitute its *reputation for security*. The consumer's beliefs can be interpreted as the level of trust that he has in the firm with regards to the protection of his data.[6]

Investment externalities. Data breaches do not impose any direct costs on the website but may lead to fraud losses for the consumer. The incidence of fraud hinges on two factors: the website's state of security, which depends on its investment level, and the bank's (card issuer) fraud prevention ability, which I assume to be exogenous. I suppose that the website bears no liability for the fraud losses. These losses are fully borne by the consumer and his bank, with the bank insuring the consumer against a share of the losses.[7] Since the website faces no liability, it does not internalize these losses when deciding how much to invest.

Customer turnover. The consumer learns about the website's security over

[5]In fact, there is evidence suggesting that this may be so: a global survey in 2015 revealed that 64% of customers would terminate their relationship with a company upon learning that sensitive personal or financial data was stolen (Gemalto, 2015).

[6]Surveys have shown that consumer trust plays an important role in commercial transactions; this provides support that a problem of imperfect information indeed exists in reality. This infographic by Gemalto shows how data breaches affect consumer trust and loyalty: `https://safenet.gemalto.com/uploadedFiles/resources/Infographics/Data_Protection/customer-loyalty-data-breaches-infographic.pdf`.

[7]This is often the case in practice. See Julie Creswell, "As Online Data Theft Escalates, Banks Look to Retailers to Bear the Losses", *The New York Times*, September 28, 2015, `https://www.nytimes.com/2015/09/29/business/as-online-data-theft-escalates-banks-look-to-retailers-to-bear-the-losses.html?_r=0`.

time through the imperfect detection of data breaches; this occurs when he discovers fraudulent transactions on his bank statement. Data breaches serve as perfect bad news signals, and their detection leads to a decline in the website's reputation.[8] Consequently, the consumer may decide not to purchase from the website in the following period. The cost of lost business arising from customer turnover can be thought of as the *reputation cost* of a data breach.

I first characterize the Bayes-Nash equilibria of this game and examine when and how reputation concerns may impact the website's security investment. The website invests if and only if a bad reputation results in lost business in the second period; i.e., when the consumer is willing and able to punish the firm for poor data security by voting with his feet. Conditional on purchasing, punishment only occurs when the consumer's valuation for the website's product is not too high (i.e., lower than his maximum expected losses). Otherwise, he always finds it optimal to buy from the firm. Moreover, the higher his expected fraud losses, the higher his *willingness to punish* the firm. The impact of reputation concerns on the website's investment further depends on the consumer's *ability to punish* the firm. The consumer can only punish the website for poor security if he learns of a data breach; hence, his ability to do so is higher when the incidence and detection rate of fraud are higher. Given the low likelihood of detection, the high level of liability protection offered by major card brands, and the reduced risk of fraud due to the fraud prevention measures taken by banks, the consumer is likely to have little willingness and ability to punish firms for poor security in practice. This suggests that, in the absence of interventions, reputation concerns are unlikely to provide strong incentives for a firm to invest.

I consider various policy measures that may be introduced to improve the website's investment incentives and analyze their impact on the equilibrium security level and consumer surplus. Regulators can either intervene *indirectly* through measures that enhance the role of reputation—equivalently, the consumer's willingness or ability to punish the firm—or *directly*, by addressing the market failures of imperfect information and externalities. I first examine three indirect policy measures: mandatory breach notification (which affects both the consumer's willingness and ability to punish), active fraud monitoring by the bank (which raises only his ability), and the expulsion of breached merchants from payment networks (which raises only his willingness). I show that the website always invests (weakly) more under active fraud monitoring and the expulsion of breached merchants but may invest less under mandatory breach notification. This is because breach notification creates two opposing effects. On the one hand, it raises the consumer's ability to punish by increasing the likelihood that he learns of a breach; on the other hand, it reduces his willingness to do so by enabling him to mitigate more of his fraud losses. Further, I show consumer surplus is always (weakly) higher only in the case of active fraud monitoring and may be lower under the other two policies. Next, I discuss two direct policy measures. To alleviate the problem of imperfect information, I consider a policy that obliges the firm to reveal either its state of security or its amount of investment to the consumer; this may be implemented via certification. To address the issue of externalities, I examine a liability rule that shifts a part of

[8]This can be thought of as a loss of consumer trust.

the bank's fraud liability to the firm. I show that both of these policies increase the equilibrium investment level and consumer surplus. My analysis suggests that direct interventions may be more desirable from the consumer's perspective relative to indirect measures, particularly those that affect his willingness to punish (e.g., mandatory notification and the expulsion of merchants). It also demonstrates the tension that may exist between protecting the consumer against breach losses (*ex post*) and incentivizing the firm to invest in security (*ex ante*). This suggests that more than one policy instrument is needed if regulators want to achieve the twin objectives of raising investment and reducing consumer losses.

Finally, I present a few extensions to the model. I examine the case where the bank's fraud prevention ability is endogenously determined by its investment in counter-fraud technologies. Under this setting, the implications of policy interventions become more nuanced—the policies may alter the bank's investment decision and, consequently, their impact on the overall security level may be ambiguous. I also consider scenarios where there are multiple firms. I show the firm invests more in the presence of a competitor (selling a perfect substitute), provided that the data breaches are publicly announced. However, it invests less in the presence of a non-competing firm (selling an independent good). This arises because the consumer is unable to perfectly attribute a breach to its source when he uses his card at multiple firms.

My work stands at the intersection of various strands of literature. Closely related are two connected branches of literature: product safety and product quality. The data security investment problem is largely similar to that of product safety. In both cases, a firm can exert costly effort (resp. invest) to lower the probability of product (resp. security) failure, which results in losses for the consumers. The incentives of a firm to exert effort have been extensively studied in the product safety literature (see Daughety and Reinganum (2011) for an overview); however, the setting and focus of this literature differ from my work. The product safety literature studies the case of durable goods and focuses on how liability regimes affect a firm's effort, whereas I consider a repeat-purchase setting and focus on the impact of reputation concerns. In this regard, my paper draws from the product quality literature, which examines the role of reputation in the provision of quality (Allen, 1984; Board and Meyer-ter Vehn, 2013; Dybvig and Spatt, 1983; Klein and Leffler, 1981; Rogerson, 1983; Shapiro, 1982, 1983; Smallwood and Conlisk, 1979). Though similar in spirit, my work differs from the existing literature in several ways. In the quality literauture, the premium on high quality products arises endogenously;[9] in my work, however, the rent that the firm earns is fixed and independent of its security level. This is a significant point of departure for two reasons. First, it introduces a second source of market failure—investment externalities—that is not present in the literature. Second, and more importantly, it allows me to abstract away from the price setting problem and focus on the impact of consumer information and learning on the reputation mechanism (and hence the investment outcome); this has received relatively little attention in literature (Dybvig and Spatt (1983), Shapiro (1983) and Board and Meyer-ter Vehn (2013) are a few exceptions).

[9]In fact, much of the literature is focused on explaining how a premium on high quality products can arise at equilibrium despite free entry (Allen, 1984; Klein and Leffler, 1981; Shapiro, 1982).

Further, while prior studies find that an improvement in consumer learning always raises investment (Dybvig and Spatt, 1983; Shapiro, 1983), I show that this may not be the case.

My paper also contributes to the literature on information/cyber-security investment. In their seminal work, Gordon and Loeb (2002) develop a model in which a firm can invest to reduce the probability of costly breaches and characterize its optimal security investment using a cost-benefit approach. Their framework was later extended to settings where there are multiple firms. In these settings, a firm's investment may impose externalities on other firms, creating security interdependencies (Acemoglu et al., 2016; Grossklags et al., 2008; Kunreuther and Heal, 2003; Riordan, 2014; Varian, 2004). My work complements the existing literature by considering another form of investment externalities—those imposed on third parties (i.e., consumers and financial institutions) without the network or system. By embedding the reputation mechanism from the quality literature into the model of Gordon and Loeb (2002), I present a first analysis of how the consumer's behavior and the firm's investment decision interact as a result of these externalities. In doing so, my model also endogenizes the losses that firm incurs from a security breach, which prior works take as given.

Finally, my work forms part of the strand of literature examining a broad range of issues surrounding data breaches. These include the relation between data breaches and identity thefts and payment fraud (Roberds and Schreft, 2009; Sullivan, 2010), the optimal response by affected third parties (e.g. banks and financial institutions) to data breaches (Graves et al., 2014), the reaction of consumers to breaches and breach notification (Ablon et al., 2016; Greene and Stavins, 2017; Kwon and Johnson, 2015; Mikhed and Vogan, 2015, 2017), the impact of breach announcements on stock prices (Acquisti and Grossklags, 2005; Campbell et al., 2003; Cavusoglu et al., 2004) and so on. Most closely related to my paper is a study by Romanosky et al. (2010), which examines the theoretical impact of mandatory breach notification on a firm's investment incentives. The authors find that notification always increases a firm's investment. This lies in contrast with my finding that such a policy may at times reduce investment. The difference in our findings stems from the fact that the firm always incurs a higher loss under breach notification in their model, whereas this is not always the case in my framework.

The rest of this paper proceeds as follows. In the next section, I present the model set-up. I then characterize the equilibrium of the game and discuss the role reputation in security investment in Section 3. In Section 4, I examine why the reputation may not play a significant role in a firm's investment decision, from both theoretical and practical standpoints. I devote Section 5 to the analysis of the various policy measures that may help to improve a firm's investment incentives. In Section 6, I consider two extensions to my model: strategic bank and multiple firms. Section 7 concludes.

2 A Model of Data Security Investment

Players and actions. Consider a model with two strategic players, a website and a representative consumer, that interact across two periods. At every period $t \in \{1, 2\}$, the website offers one unit of its product for sale at a fixed price to the consumer.[10] The website only accepts card-based payments and the consumer is required to enter his payment card data on the website in order to complete a transaction. Cyber-criminals conduct attacks on the website at every period in order to steal the consumer's payment data. These cyber-attacks, when successful, result in data breaches. The stolen card data may then be used by the cyber-criminals to commit payment fraud.

Before the start of the first period, the website can make a once-and-for-all security investment to protect itself against these cyber-attacks. For a given amount of investment, $c(q)$, the website is *secure* against cyber-attacks with probability q and *vulnerable* with probability $1-q$. The website never experiences a breach when it is secure, but suffers from one with probability ρ when it is vulnerable.[11] Since the website's ex-ante vulnerability to cyber-attacks, $(1-q)\rho$, is decreasing in q, q can be broadly considered as a measure of the website's security level.[12] Throughout this paper, I refer to q as the *security level* of the website.

The consumer has to decide whether or not to purchase from the website at every period. Let $v \in \mathcal{R}^+$ denote the consumer's valuation for the website's product (net of the price paid).[13] When purchasing from the website, the consumer also exposes himself to the risk of data breaches and, consequently, payment fraud. Cyber-criminals succeed at committing fraud using stolen payment data with probability $1-\gamma$, where γ captures the fraud prevention ability of the consumer's payment card provider (henceforth, the consumer's bank). The bank's fraud prevention technology is assumed to be imperfect ($\gamma < 1$) and exogenous (i.e., the bank is non-strategic) in the baseline setting. In the event that fraud occurs, the consumer may incur losses of up to l.

Information. The amount of data security investment is privately known to the website, whereas the realized state of security (*secure* or *vulnerable*) is unobserved by both the website and the consumer. At the start of a period t, the consumer holds rational beliefs q_{t-1} about the security level; these beliefs constitute the website's reputation for security at that period. Data breaches serve as perfect bad news signals in my model, since breaches only occur when the firm is vulnerable. When the consumer's data is breached, he learns about it with probability $\lambda < 1$.

[10] The price of the product is assumed to be fixed exogenously so that it cannot serve as a signal of security. Such a scenario could arise in practice when the website is subject to a resale price maintenance policy for example.

[11] One can think of $1-q$ as the probability that the cyber-criminals successfully develop (or purchase on the dark net) tools to circumvent the security measures that the website has invested in. In this case, ρ can be interpreted as the probability that the hackers succeed in stealing the consumer's information with these tools.

[12] The precise measure of the website's ex-ante level of security is $1-(1-q)\rho$. This corresponds to the probability that no data breach is expected to occur for a given level of investment $c(q)$.

[13] The consumer's valuation v can alternatively be interpreted as a measure of product quality.

The breach detection rate λ can be interpreted the probability that the consumer (privately) notices fraudulent charges on his payment card statement.[14] Conditional on having purchased from the website at a given period, the consumer updates his beliefs at the end of the period using Bayes' rule.

Timing. The timing of the game is as follows:

- $t = 0$: The website decides the amount, $c(q)$, to invest in data security. The state of security is realized.
- $t = 1$: The consumer decides whether or not to purchase from the website given his valuation for its product and the website's (initial) reputation for security. Conditional on having purchased, the consumer updates his belief about its security level (i.e., its reputation) at the end of the period.
- $t = 2$: The consumer makes his purchase decision for the second period given the website's updated reputation.

Figure 1 illustrates how the game proceeds in the first period (the game in the second period is a replication of the first).

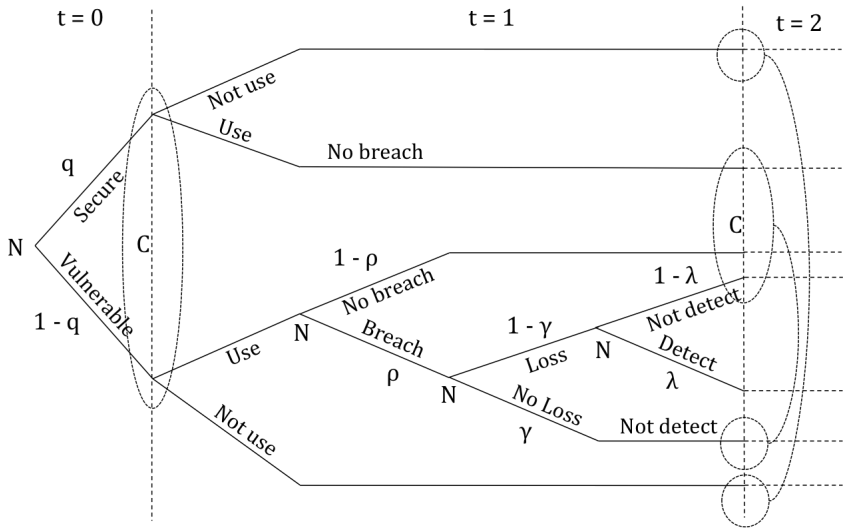

Figure 1: Investment Game with Non-Strategic Bank

Payoffs. At each period $t \in \{1, 2\}$, the consumer expected (within-period) utility from making a purchase is given by his net valuation v less any fraud losses that he expects to incur as a result of data breaches at the website. The fraudulent use of the consumer's card results in losses amounting (on average) to l. His bank's fraud liability protection policy insures him against a share α of these losses, provided

[14] I assume that fraud only arises when the consumer's payment data has been breached. Thus, consumer can directly infer from the detection of fraud that a data breach has occurred at the website. Throughout this paper, I will use the terms "loss detection" and "breach detection" interchangeably.

that the consumer discovers and reports them to the bank (within a reasonable time frame).[15] I assume that α is exogenously determined.[16] For a given rate of breach detection λ, the consumer's expected utility from using the website within a period is therefore

$$E(U_t) = v - (1 - q_{t-1})\tilde{\rho}(1 - \lambda\alpha)l, \tag{1}$$

where q_{t-1} denotes the consumer's belief about the website's state of security at (the beginning of) period t[17] and $\tilde{\rho} \equiv \rho(1-\gamma)$ gives the probability that the consumer experiences a fraud when using a vulnerable website (for a given fraud prevention rate).

I assume that data breaches do not impose any direct costs on the website.[18] The website's profit is given by the sum of its expected sales revenue R_t across the two periods, less its investment in data security at the beginning of the game ($t = 0$):

$$\pi(q; q_0, r) \equiv R_1(q_0, r) + \delta_f R_2(q; q_0, r) - c(q),$$

where δ_f is the discount factor of the website and r is the (net) revenue that it obtains when the consumer purchases its product at a given period.[19] As we will see shortly, the consumer's purchase decision at $t = 1$ depends only on the website's initial reputation q_0, while his decision at $t = 2$ may depend additionally on whether or not he has detected a breach during the first period. The website's security investment cost function $c(.)$ satisfies the following assumption.

Assumption 1 (Website's Investment Cost)**.**
The website's investment cost function $c : [0, 1] \mapsto \mathbb{R}^+$ *is continuously differentiable, increasing and convex. Further, it satisfies the boundary conditions* $c'(0) = c''(0) = 0$ *and* $\lim_{q\to 1} c'(q) = \infty$.

Assumption 1 captures several general characteristics of data security investments. The convexity of the cost function reflects how improving data security

[15]Depending on the type of card and the type of transaction, the banks are typically liable for the fraudulent charges incurred, provided that consumers detect and report the losses before a certain deadline. For more information on payment card liability, visit: http://www.consumer.ftc.gov/articles/0213-lost-or-stolen-credit-atm-and-debit-cards.

[16]For instance, the value of α may depend on the intensity of competition for consumers between card issuers.

[17]Unless stated otherwise, the consumer's beliefs at period t refers to those at the *beginning* of the period. These correspond to his updated beliefs at the end of period $t - 1$.

[18]Firms typically incur little direct costs from data breaches. First, as mentioned in the introduction, banks and financial institutions are largely liable for breach-related fraud losses. Moreover, many firms also receive insurance pay-outs and tax deductibles for breach-related expenses; these offset the direct costs that they incur. For instance, Target's and Home Depot's data breach expenses (net of insurance reimbursements and tax deductions) only amounted to 0.1% and 0.01% of their 2014 sales respectively. (See Benjamin Dean, "Why Companies Have Little Incentive to Invest in Cybersecurity", *The Conversation*, March 4, 2015, `http://theconversation.com/why-companies-have-little-incentive-to-invest-in-cybersecurity-37570`.)

[19]The sales revenue r is net of all costs that the website has incurred in sale of the product.

is more difficult (and costly) when the firm's level of security is already high.[20] Another aspect of data security investment—perfect security is not optimal—is captured by the boundary condition, $\lim_{q\to 1} c'(q) = \infty$. It is too costly, if not altogether impossible, for a firm to reduce the probability of experiencing a data breach to zero.[21]

3 Equilibrium Analysis: The Role of Reputation

3.1 Strategies

3.1.1 Consumer's Purchasing Decision

At every period, the consumer faces a consumption-security tradeoff: he derives a positive utility from consuming the website's product but exposes himself to potential fraud losses when purchasing from the website. His purchase decision thus hinges on his valuation for the product, which determines his consumption utility, and the website's reputation, which determines his expected fraud losses. The consumer purchases from the website whenever his valuation exceeds the expected fraud losses. Let $\overline{v}$ denote the maximum expected fraud losses that the consumer may incur; this arises when the website is vulnerable with certainty:

$$\overline{v} = \tilde{\rho}(1 - \lambda\alpha)l.$$

When the consumer's valuation is very high, such that it exceeds $\overline{v}$, he always buys from the website regardless of its reputation. When $v < \overline{v}$, however, the consumer's purchase decision is contingent on the website's reputation for security.

Consider the consumer's problem at $t = 2$. The website's reputation for security is given as follows:

$$q_1 = \begin{cases} q_0 & \text{if he did not purchase at } t = 1 \\ 0 & \text{if he purchased and a breach was detected} \\ \dfrac{q_0}{1 - \lambda(1 - q_0)\tilde{\rho}} & \text{if he purchased but no breach was detected.} \end{cases}$$

Notice that $0 \leq q_0 < \frac{q_0}{1-\lambda(1-q_0)\tilde{\rho}}$. Conditional on making a sale at $t = 1$, the website's reputation for security deteriorates following the detection of a breach but improves in the absence of detection. It is clear that the consumer will not purchase at the second period after detecting a breach—his expected losses when purchasing would be $\overline{v}$, which exceeds his valuation. If the consumer did not detect any breaches, he purchases if and only if the website's reputation is sufficiently high; more precisely,

$$q_0 \geq \frac{(1 - \lambda\tilde{\rho})\overline{v} - v}{\tilde{\rho}(\lambda v + (1 - \lambda\overline{v}))} = \hat{q}_0^{NB}.$$

[20] Indeed, a firm that is completely unsecured can easily raise its level of security by introducing measures that are relatively costless (such as stronger passwords and email encryption). By contrast, a firm that already possesses a strong security posture may have to purchase a more costly data protection software or engage security consultants to further improve its level of security.

[21] A similar assumption that perfect security cannot be attained with a finite amount of investment is also described in Gordon and Loeb (2002).

Similarly, if the consumer did not purchase at $t = 1$, it is optimal for him to purchase at $t = 2$ if and only if

$$q_0 \geq \frac{\tilde{\rho}(1-\lambda\alpha)l - v}{\tilde{\rho}(1-\lambda\alpha)l} = \hat{q}_0^{NP}.$$

Let us examine the consumer's problem at $t = 1$. His expected utility from purchasing is given by the sum of his within-period expected utility and the option value of learning.[22] This option value is given by:

$$J = \begin{cases} \delta_c(1-\lambda(1-q_0)\tilde{\rho})E(U_2|\text{No breach detected}) & \text{if } q_0 \in [\hat{q}_0^{NB}, \hat{q}_0^{NP}) \\ 0 & \text{if } q_0 \geq \hat{q}_0^{NP}, \end{cases}$$

where δ_c denotes the consumer's discount factor. When $q_0 \in [\hat{q}_0^{NB}, \hat{q}_0^{NP})$, learning presents the consumer with an option to purchase in the second period; he chooses to do so when he does not detect a breach. The consumer finds it optimal to purchase at $t = 1$ when

$$E(\tilde{U}_1) = E(U_1) + J > 0,$$

which is the case if

$$q_0 \geq \hat{q}_0 = \frac{(1+\delta_c(1-\lambda\tilde{\rho}))(\overline{v}-v)}{\tilde{\rho}\delta_c\lambda v + (1+\delta_c(1-\lambda))\overline{v}}.$$

Notice that $\hat{q}_0^{NB} < \hat{q}_0 < \hat{q}_0^{NP}$. The first inequality implies that whenever the consumer finds it optimal to purchase at the first period, he finds it optimal to continue doing so at the second if he did not detect any breaches. This is not surprising since the website's reputation for security is improved in that case. The second inequality shows how, due to the positive option value of learning, it may be optimal for a consumer to purchase at the first period even when his expected within-period utility ($E(U_1) = E(U_2|\text{No purchase})$) is negative.

Lemma 1 (Consumer's Purchasing Strategy)**.**
(i) When $v \geq \overline{v}$, always purchase.
(ii) When $v < \overline{v}$ and $q_0 \geq \hat{q}_0$, purchase at the first period and punish the firm by not purchasing at the second if a breach was detected.
(iii) When $v < \overline{v}$ and $q_0 < \hat{q}_0$, never purchase.

The above lemma illustrates how the consumer's purchase decision may depend upon the website's reputation for security. Specifically, when $v < \overline{v}$, the consumer only purchases at a period when the website has a sufficiently good reputation. The detection of a breach results in a decline in the website's reputation and may result in customer turnover (i.e., punishment) in this case.

3.1.2 Website's Investment Problem

At the beginning of the game, the website has to determine its profit-maximizing level of security. For a given initial reputation q_0, its profit is

$$\pi(q; q_0) = \begin{cases} (1+\delta_f)r - c(q) & \text{if } v \geq \overline{v} \\ (1+\delta_f(1-\lambda(1-q)\tilde{\rho}))r - c(q) & \text{if } v < \overline{v} \ \& \ q_0 \geq \hat{q}_0 \\ 0 - c(q) & \text{if } v < \overline{v} \ \& \ q_0 < \hat{q}_0. \end{cases}$$

[22] There exists an option value when $q_0 \geq \hat{q}_0^{NB}$.

The website's expected sales revenue depends on the consumer's purchasing strategy as described in Lemma 1; it obtains r when a sale is made.

Lemma 2 (Website's Investment Strategy)**.**
Let $q^{BR} : [0,1] \mapsto [0,1]$ denote the optimal investment level of the website for a given initial reputation q_0.[23] *We have that*

$$q^{BR}(q_0) = \begin{cases} c'^{-1}(\delta_f \lambda \tilde{\rho} r) & \text{if } v < \overline{v} \ \& \ q_0 \geq \hat{q}_0 \\ 0 & \text{otherwise.} \end{cases}$$

The website invests in data security if and only if breaches are costly. Since the firm incurs no direct costs from data breaches, it only invests when it faces an indirect (reputation) cost—the loss of future sales due to customer turnover. From Lemma 1, this occurs when the consumer adopts a *purchase and punish* strategy (i.e, when $v < \overline{v}$ and $q_0 \geq \hat{q}_0$). Its optimal level of security equates the reduction in expected reputation cost, $\delta_f \lambda \tilde{\rho} r$, to the marginal investment cost, $c'(q)$.[24]

3.2 Bayes-Nash Equilibrium with Rational Expectations

I now characterize the Bayes-Nash equilibria of this game. Let q^* denote the website's equilibrium security level and suppose that the consumer holds rational expectations at equilibrium; i.e., $q_0 = q^{BR}(q_0) \equiv q^*$.

It is easy to show that there always exists an equilibrium at which the website does not invest in security. At this equilibrium, the consumer believes the website to be vulnerable with certainty; there is no learning and, hence, no customer turnover. Therefore, it is indeed optimal for the website not to invest. There may also exist an equilibrium at which the website invests in security. From Lemma 2, we know that this occurs when $v < \overline{v}$ and $q_0 \geq \hat{q}_0$. The condition $q_0 \geq \hat{q}_0$ is satisfied in equilibrium if $q^{BR}(\hat{q}_0) \geq \hat{q}_0$, which is the case whenever

$$c'^{-1}(\delta_f \lambda \tilde{\rho} r) \geq -\frac{(1 + \delta_c(1 - \lambda\tilde{\rho}))(v - \overline{v})}{(1 + \delta_c)\overline{v} + \delta_c \lambda \tilde{\rho}(v - \overline{v})},$$

or equivalently, when

$$v \geq \frac{(1 - c'^{-1}(\delta_f \lambda \tilde{\rho} r))(1 + \delta_c(1 - \lambda\tilde{\rho}))}{1 + \delta_c(1 - \lambda(1 - c'^{-1}(\delta_f \lambda \tilde{\rho} r))\tilde{\rho})}\overline{v} = \underline{v}.$$

[23] The website's profit is decreasing and concave function in q when $v \geq \overline{v}$ and when $v < \overline{v}$ and $q_0 < \hat{q}_0$. This implies we have a corner solution to the website's maximization problem. By contrast, the website's profit is a positive and convex function of q when $q_0 \geq \hat{q}_0$. Therefore, a unique interior solution exists.

[24] The website's profit function when $v < \overline{v}$ and $q_0 \geq \hat{q}_0$ can be re-expressed as follows:

$$\pi(q, q_0) = (1 + \delta_f)r - \underbrace{\delta_f(1 - \lambda(1 - q)\tilde{\rho})r}_{\text{Expected reputation cost}} - c(q),$$

where the first term corresponds to the website's revenue in the absence of a breach and the second term gives the expected reputation cost of a breach. By investing in security, the website lowers the likelihood of a breach and hence the probability it experiences customer turnover.

The website's optimal level of security at this equilibrium, q^*, follows directly from Lemma 2.

Proposition 1 (Bayes Nash Equilibria with Rational Expections)**.**
(i) There always exists a no investment equilibrium where the website does not invest in data security; i.e., $q^* = 0$. The consumer always purchases at this equilibrium when $v \geq \overline{v}$ and never purchases otherwise.
(ii) There exists a positive investment equilibrium if and only if the consumer's valuation belongs to the interval $[\underline{v}, \overline{v})$. At this equilibrium, the website's level of security is

$$q^* = c'^{-1}(\delta_f \lambda \tilde{\rho} r).$$

The consumer purchases from the website at the first period and continues to do so at the second if and only if he did not detect a breach.

Proposition 1 illustrates how the presence of a reputation cost, which arises from customer turnover, can create incentives for a website to invest. Observe moreover that the website's security level q^* at the positive investment equilibrium is increasing the magnitude of the reputation cost.

Equilibrium selection. Both equilibria exist when $\underline{v} \leq v < \overline{v}$. In order to facilitate the analysis in the remainder of this paper, I perform a selection between the two equilibria based on the Pareto criterion. The total surplus generated at equilibrium is given by

$$W(q^*) = \pi(q^*) + CS(q^*),$$

where

$$CS(q^*) = (1 + \delta_c(1 - \lambda(1 - q^*)\tilde{\rho}))v - (1 - q^*)(1 + \delta_c(1 - \lambda\tilde{\rho}))\overline{v}.$$

It is straightforward to verify that the positive investment equilibrium is Pareto-dominant. At the no investment equilibrium, the website does not invest and the consumer never participates; therefore, both profit and consumer surplus are zero. By contrast, both profit and consumer surplus are positive at the positive investment equilibrium.[25] Hereafter, I assume that the website and the consumer will coordinate on the positive investment equilibrium whenever both equilibria exist.

4 Comparative Statics: When does Reputation Play a Bigger Role?

In the equilibrium analysis, we established that reputation concerns can help to provide the website with incentives to invest in security. However, how important

[25] At the positive investment equilibrium, $q^* = c'^{-1}(\delta\lambda\tilde{\rho}r)$, and the website's profit and the consumer surplus are respectively

$$\pi^*(c'^{-1}(\delta\lambda\tilde{\rho}r)) = (1 + \delta_f(1 - \lambda(1 - c'^{-1}(\delta\lambda\tilde{\rho}r))\tilde{\rho}))r - c(c'^{-1}(\delta\lambda\tilde{\rho}r)) > 0;$$

$$CS(c'^{-1}(\delta\lambda r\tilde{h}or)) = (1 + \delta_c(1 - \lambda(1 - c'^{-1}(\delta\lambda\tilde{\rho}r))\tilde{\rho}))v - (1 - c'^{-1}(\delta\lambda\tilde{\rho}r))(1 + \delta_c(1 - \lambda\tilde{\rho}))\overline{v} > 0.$$

are these reputation concerns to the firm and when do they matter more? To answer these questions, I examine a few factors that determine the reputation cost that data breaches impose on the website. I focus on three variables: the breach detection rate λ, the bank's share of liability α and the bank's fraud prevention ability γ. These factors are chosen because of their high policy relevance. To simplify the analysis, I restrict attention to the case of a myopic consumer (i.e., $\delta_c = 0$).[26] I denote the valuation threshold above which the myopic consumer purchases at $t = 1$ by $\underline{v}_M$ (which corresponds to $\underline{v}$ with $\delta_c = 0$).

Before proceeding with the analysis, I first define a few useful concepts: the consumer's *willingness to purchase* from the website, his *willingness* and *ability to punish* the website for data breaches, and the website's *willingness to invest*. The consumer's *willingness to purchase* at any period t given the website's reputation is captured by the valuation threshold above which he finds it optimal to buy the website's product. The *lower* the valuation threshold, the *higher* his willingness to purchase. The two thresholds that are relevant for the analysis are $\underline{v}_M$, that above which he purchases at the first period, and $\overline{v}$, that above which he purchases at the second following the detection of a data breach. The consumer's *willingness to punish* the website for data breaches is the contraposition of his willingness to purchase after learning the website has been breached—the *higher* the threshold $\overline{v}$, the *higher* his willingness to punish. The consumer's *ability to punish* the website for breaches is related to the probability of breach detection. Since the consumer can punish the website for breaches only if he discovers them, his ability to do so is increasing in the likelihood of learning. Finally, the website's *willingness to invest* is captured by the size of the interval over which it invests a positive amount in data security. Defining $\Delta v \equiv \overline{v} - \underline{v}_M$, the *larger* Δv, the *higher* the website's willingness to invest.

The impact of the breach detection rate, the bank's share of liability and its fraud prevention ability on the website's investment incentives depend on how they affect the consumer's purchasing behavior. I first suppose that the consumer is willing to punish the website and examine how these factors impact his ability to do so.

Corollary 1 (Consumer's Ability to Punish)**.**
The consumer's ability to punish the website for poor security is increasing in the breach detection rate λ but decreasing in the bank's fraud prevention ability γ.

The consumer learns about a breach when he detects the resulting fraud losses. This occurs with a higher probability when breaches are more likely to lead to fraud and when the detection rate of these losses is higher.

Corollary 2 (Website's Optimal Security Level)**.**
An increase in λ strengthens the reputation effect on the website's equilibrium security level q^*, while an increase in γ weakens it.

[26] A change in these variables, in particular the breach detection rate and the bank's fraud prevention ability, also affects the option value of learning. The impact of a change in these variables on the option value is *a priori* ambiguous, which further complicates the analysis.

An improvement in the consumer's ability to punish the website raises the cost and hence strengthens the *reputation effect* of data breaches. As such, the website finds it optimal to invest more in data security for any given initial reputation.

Consider now the consumer's willingness to purchase from the website and his willingness to punish it when he detects a breach. The consumer cares more about security, and is thus less willing to purchase from and more willing to punish a vulnerable firm, when his expected fraud losses from data breaches are higher. Holding the website's security level fixed, an increase in the breach detection rate λ, the bank's share of liability α or its fraud prevention ability γ all lower the fraud losses stemming from data breaches. This lowers the consumer's willingness to punish. However, a change in these variables may also have an indirect impact on the expected fraud losses, via a change in the website's optimal investment level. From Corollary 2, we know that an increase in λ raises q^*, which further reduces the consumer's expected losses, while an increase in γ has the opposite effect. Because the indirect effect of an increase in γ countervails its direct effect, fraud losses is reduced only if the indirect effect is relatively weak. This occurs when the website's investment cost function is sufficiently convex (see Appendix A for more details). I assume throughout the paper that this is the case.

Corollary 3 (Consumer's Willingness to Purchase and to Punish)**.**
The consumer's willingness to purchase at the first period is increasing in the breach detection rate λ, the bank's share of liability α and its fraud prevention ability γ; his willingness to punish the firm for breaches is decreasing in these variables.

As the above corollary illustrates, an increase in λ, α or γ creates two opposing effects on the website's investment incentives. The first effect is associated with the decline in the consumer's willingness to punish (i.e, a reduction in $\overline{v}$). This reduction reflects a consumer moral hazard problem, akin to that which arises in insurance markets. Better protection against fraud losses makes the consumer more willing to buy from a vulnerable website; consequently, the website is less willing to invest—a *crowding out effect*. The second effect results from the increase in the willingness to make an initial purchase (i.e., a decrease in $\underline{v}_M$)—a *market expansion effect*. This leads this website to invest at lower valuations. Because both $\overline{v}$ and $\underline{v}_M$ are simultaneously reduced, the overall impact of an increase in λ, α or γ on Δv is *a priori* ambiguous. The website's willingness to invest is higher (i.e., Δv is larger) when the *market expansion effect* is stronger than the *crowding out effect*.

Corollary 4 (Website's Willingness to Invest)**.**
The website's equilibrium willingness to invest (as indicated by Δv) is decreasing in the bank's share of liability α and its fraud prevention ability γ. It is increasing in the breach detection rate λ for all $\lambda \in [0,1]$ if the bank's share of liability is sufficiently small ($\alpha < \overline{\alpha}$).[27] *If $\alpha > \overline{\alpha}$, there exists a threshold $\overline{\lambda}$, below which Δv is increasing in λ; Δv is decreasing in λ otherwise.*

[27]The threshold level of liability is given by:

$$\overline{\alpha} = \frac{\delta_f \rho (1-\gamma) r}{\delta_f \rho (1-\gamma) r + c'^{-1}(\delta_f \rho (1-\gamma) r) c''(c'^{-1}(\delta_f \rho (1-\gamma) r))}.$$

Figures 2 to 5 provide graphical illustrations of the results presented in the above corollary.

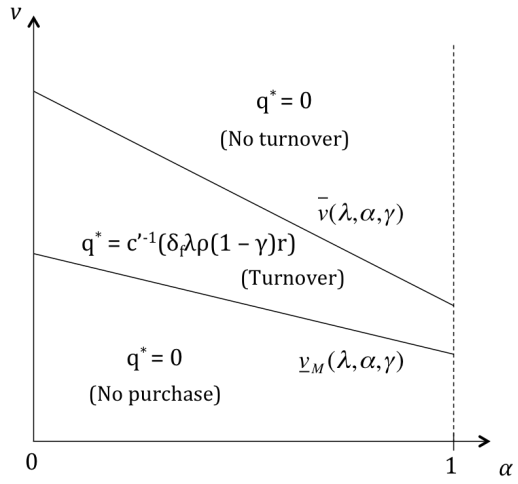

Figure 2: Impact of a change in α

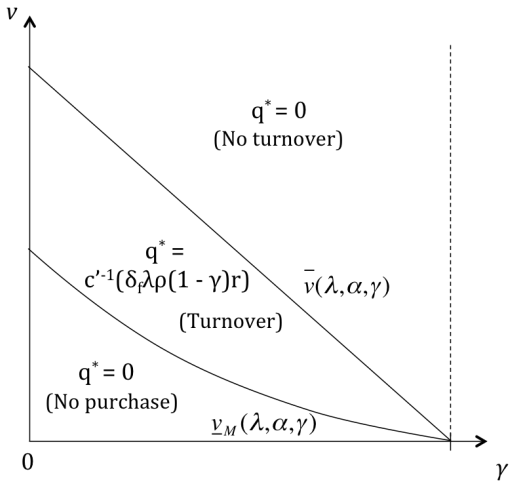

Figure 3: Impact of a change in γ

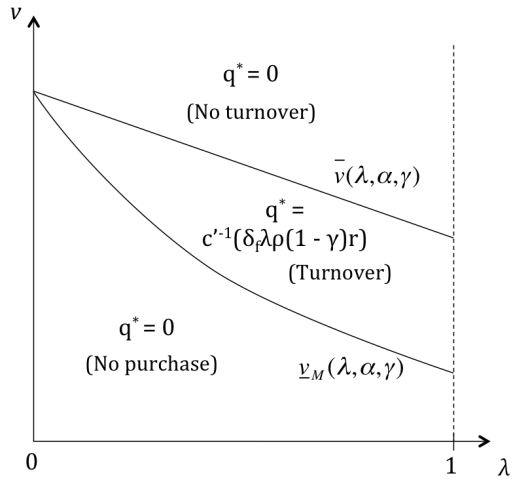

Figure 4: Impact of a change in λ ($\alpha \leq \overline{\alpha}$)

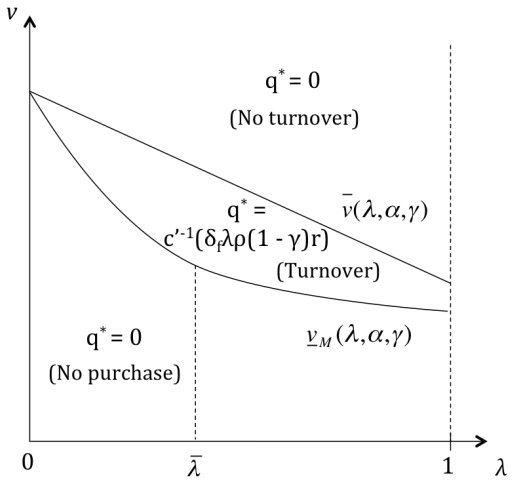

Figure 5: Impact of a change in λ ($\alpha > \overline{\alpha}$)

4.1 What Limits the Role of Reputation in Reality?

Having established the theoretical impact of the breach detection rate, the fraud liability protection level and the fraud prevention ability on the website's investment incentives, I now examine why the role of reputation may be limited in practice.

Low breach detection rate. The probability of breach detection may be low for several reasons. Many consumers do not go over their monthly payment card statements thoroughly. Even when they do, it is often difficult for them to identify where the charges to their cards came from. The billing descriptor, which provides

the name of the company with which the consumer conducted the transaction, is said to be "frustrating brief"—it is limited to between 26 and 28 characters (The New York Times, 2010, August 21). For example, a consumer who has used his credit card at a Shell gas station may see a vague billing descriptor,"SHO No. 15", on his statement.[28] This makes it hard for even the most meticulous of consumers to distinguish between legitimate and fraudulent transactions. Furthermore, fraudsters also employ various tactics in order to minimize the chances of detection by consumers. For instance, they would make small, inconspicuous charges to consumers' cards (e.g., charges of $9.84) instead of large ones; these are likely to go unnoticed by the consumers.[29] A low level of breach detection makes the consumer less willing to make an initial purchase from a firm (a *market contraction effect*) and also limits the consumer's ability to punish a firm for poor security.

Limited liability for fraud losses. In the United States, the Fair Credit Billing Act and the Electronic Fund Transfer Act set limits on a consumer's liability for the losses that result from the unauthorized use of his payment card. The consumer's maximum loss from fraudulent credit card charges is capped at US$50. For debit cards, the consumer's liability is limited to US$50 if he reports the loss within 2 days, up to US$500 if he reports it between 2-60 days and unlimited thereafter.[30] Likewise, in the European Union, the Payment Services Directive (Article 74) limits a consumer's liability for unauthorized payment transactions to €50 (except in the cases of gross negligence or payer fraud).[31] In fact, many major credit card issuers (such as VISA and Mastercard) go beyond these regulatory requirements, offering consumers full protection against fraudulent charges via their zero liability policies. In an interview study conducted by Cheney et al. (2012), merchants cited the high level of liability protection as the reason why consumers have little incentives to protect their data, which implies a low willingness to punish a firm for poor security. In other words, liability protection crowds out investment incentives.

Improvement in fraud prevention ability. Many banks are investing into better fraud prevention measures in order to combat the rise in payment card fraud. An improvement in the bank's ability to prevent fraud lowers the losses that the consumer expects to incur when he makes a purchase. He is therefore less willing to punish a breached firm by leaving, which leads to a *crowding out effect*. Further, the adoption of certain fraud prevention measures also reduces the probability that the consumer learns about a breach, weakening his ability to punish the firm. One such

[28]This example is taken from the article "Making sense of confusing credit card statements. Spotting fraud hard with so many unfamiliar, but legitimate, business names". See full article at `http://www.creditcards.com/credit-card-news/sense-confusing-card-statements-1282.php`.

[29]Cheryl Corley, "Lots of little credit charges add up to one big scam", *NPR*, February 3, 2014, `http://www.npr.org/2014/02/03/271027087/lots-of-little-credit-charges-add-up-to-one-big-scam`

[30]For more details, visit `https://www.consumer.ftc.gov/articles/0213-lost-or-stolen-credit-atm-and-debit-cards`.

[31]For more information on the Payment Services Directive, visit `http://europa.eu/rapid/press-release_MEMO-15-5793_en.htm?locale=en`.

measure is the introduction of EMV (or chip-and-PIN) cards. EMV cards are more difficult to counterfeit as compared to their magnetic-strip counterparts; hence, (counterfeit card) fraud is less likely to occur following a data breach. However, as the cyber-criminals' failure to commit fraud is not observable, neither the bank nor the consumer learns about the breach in this case. The implementation of (certain types of) multi-factor authentication to secure on-line transactions also generates the same effect. For instance, the bank could require (additionally) a token-generated one-time PIN. This would mean that criminals would not be able to make purchases with the stolen card data alone. In this case, the breach would once again not be observed by the consumer and the bank.

5 Policy Analysis

Although reputation concerns can provide incentives for a firm to invest in security, its impact on security investment may be limited for reasons outlined in the previous section. This suggests that further policy or regulatory interventions may be warranted. In this section, I consider two types of policy interventions—direct and indirect—that may help to increase the amount of investment made by the firm. Direct policy interventions address the root causes of the poor investment incentives: imperfect information and investment externalities. Indirect policy interventions, on the other hand, attempt to raise investment by increasing the reputation cost imposed by data breaches. To simplify the analysis, I continue to assume throughout that the consumer is myopic.[32]

5.1 Increasing the Reputation Cost of Data Breaches

5.1.1 Mandatory Data Breach Notification

Mandatory data breach notification is one of the most widely discussed and adopted policy measures targeted at increasing the amount of investment in data security. Many jurisdictions are currently considering or have recently implemented such a requirement. Under the European Union's General Data Protection Act (GDPR), which will come into force on 25 May 2018, data controllers are obliged to notify data subjects of a data breach whenever it is like to "result in a risk for the rights and freedoms of individuals".[33] The notification must be provided without undue delay (no later than 72 hours) after the data controller becomes aware of

[32]Whether the consumer is myopic or forward-looking has no implications when we analyze the direct policy interventions. The assumption is only maintained for reasons of consistency. The results for the direct policy interventions are qualitatively the same even when this assumption is relaxed.

[33]A 'controller' means the natural or legal person, public authority, agency or other body which, alone or jointly with others, determines the purposes and means of the processing of personal data; where the purposes and means of such processing are determined by Union or Member State law, the controller or the specific criteria for its nomination may be provided for by Union or Member State law (GDPR Article 4(7)). The data controller in my model is the website. A 'data subject' is an identified or identifiable natural person (GDPR Article 4(1)). The data subject corresponds to the consumer in my model.

the breach. Failure to comply with the regulation may subject the controller to an administrative fine of up to €10,000,000 or in case of an undertaking, up to 2% of the total worldwide annual turnover of the preceding financial year, whichever is higher. In the United States, the Data Security and Breach Notification Act of 2015 (a bill that was introduced in 2015) similarly obliges a breached entity to provide timely notification to affected individuals, "unless there is no reasonable risk that the breach of security has resulted in, or will result in, identity theft, economic loss or economic harm, or financial fraud to the individuals whose personal information was affected by the breach of security" (see section 3 of the Act). Any failure to comply with the notification requirement will be deemed as an unfair and deceptive act or practice in violation of the Federal Trade Commission Act, which may result in a maximum civil penalty of US$17,520,000 for first-time violation (see section 4 of the Act).

Consider a breach notification law that obliges the website to inform its customer whenever a data breach occurs. Assume further that the penalty associated with non-compliance is sufficiently high, such that the website always finds it optimal to disclose a data breach. Mandatory breach notification translates to an increase in the rate of breach detection in my model. In the absence of this regulation, the consumer detects a breach with probability λ; with breach notification, the consumer learns of a breach whenever it occurs (i.e., with probability 1). The increase in breach detection rate improves the consumer's ability to punish the website for data breaches, which (as we know from Corollary 2) increases the website's optimal security level.[34] This increase in investment is the *reputation effect* of mandatory notification.[35]

In addition to this *reputation effect*, there exists a second (often-cited) benefit to mandatory notification: the mitigation of fraud losses. Upon being notified of a breach, a consumer can, for instance, close any unused bank accounts and apply for credit freezes and fraud alerts (Romanosky et al., 2011). These actions would reduce the amount of fraud losses that the consumer incurs. In the context of my model, breach notification may bring to the consumer's attention fraudulent charges that he might have missed out otherwise, allowing him to obtain reimbursement from his bank. While the mitigation of losses would indeed be beneficial to the consumer holding all else constant, it may not necessarily be the case when we account for the strategic reaction of the website. The reduction in fraud losses (due to the loss mitigation benefit) creates two countervailing effects on the website's willingness to invest. On the one hand, the consumer is more willing to make an initial purchase, which creates a *market expansion effect*; on the other hand, he is less willing to punish a breached firm by leaving, which leads to a *crowding out effect*. As Corollary 4 illustrates, the overall impact of an increase in the breach detection rate λ on the website's willingness to invest is not always positive. Consequently, the consumer may not necessarily be better off under a breach notification law.

[34]The result in Corollary 2 correspond to that of a marginal increase in λ at a given point; however, since $q^*(\lambda, \gamma)$ is increasing in λ for all $\lambda \in [0, 1]$, $q^*(1, \gamma) > q^*(\lambda, \gamma)$ for all $\lambda < 1$.

[35]Holding the consumer's fraud losses from data breaches fixed, this increase in investment would further result in a higher initial willingness to purchase at equilibrium. This implies that the website would also be more willing to invest—a *market expansion effect*.

Proposition 2 (Mandatory Data Breach Notification)**.**
Consider a law that mandates the disclosure of data breaches to affected consumers. Let $\underline{v}_{M,BN}$ *and* $\overline{v}_{BN}$ *denote the valuations above which the consumer purchases from the website at* $t = 1$ *and following breach detection at* $t = 2$ *under the law respectively and let* $\Delta v_{BN} \equiv \overline{v}_{BN} - \underline{v}_{M,BN}$*.*
(i) The website invests weakly more under the law for $v < \overline{v}_{BN}$ *and weakly less otherwise; more specifically, it invests (strictly) more for* $v \in [\underline{v}_{M,BN}, \overline{v}_{BN})$ *and (strictly) less for* $v \in [max\{\underline{v}_M, \overline{v}_{BN}\}, \overline{v})$*.*
(ii) Let $\tilde{\Delta}v \equiv \overline{v} - max\{\underline{v}_M, \overline{v}_{BN}\}$*. The extent of investment crowding out, as captured by the ratio of* $\tilde{\Delta}v$ *to* Δv*, induced by the law is increasing in the bank's share of liability* α*. Moreover, full crowding out occurs for a bigger range of* α *when the initial breach detection rate* λ *is lower.*
(iii) Consumer surplus is reduced in the region of crowding out for α *sufficiently small.*

As highlighted in the above proposition, a breach notification law has an adverse impact on the website's investment incentives when the consumer's valuation lies in the interval $[max\{\underline{v}_M, \overline{v}_{BN}\}, \overline{v})$. In the absence of mandatory notification, a consumer whose valuation belongs to this interval would punish the firm by leaving following the detection of a breach. Under the law, however, the consumer continues purchasing from the breached firm (due to the loss mitigation benefit). Further, the higher the bank's share of liability, the larger the loss mitigation benefit. The extent of investment crowding out is therefore increasing in α. I show in the appendix that there exists a threshold $\tilde{\alpha}$ above which full crowding out occurs. Moreover, this threshold can be shown to be increasing in λ. This finding suggests that regulators may end up doing more harm than good to investment incentives when implementing a breach notification law, since a low breach detection rate and a high liability protection level are likely factors that limit the role of reputation in the first place. Figure 6 shows how the consumer's purchase decision and the website's incentives to invest are affected by the mandatory breach notification law for different levels of liability protection α. The two dashed lines correspond to the consumer's valuation thresholds ($\overline{v}(\lambda, \alpha, \gamma)$ and $\underline{v}_M(\lambda, \alpha, \gamma)$) in the unregulated market. The shaded region indicate the valuations over which the website's security investment is strictly lower under the breach notification law. Notice in addition that when the consumer bears no liability for fraudulent charges (i.e., $\alpha = 1$), the website never invests in data security under the law.

Finally, let us consider the impact of the policy on consumer surplus. Holding the level of investment fixed, the consumer's expected losses are lower, and therefore his surplus is higher, under the regulation due to the loss mitigation benefit. It is also clear that the consumer's surplus will be higher whenever the security level is increased under the regulation. However, the overall impact of the policy on consumer surplus is less obvious when the security level is reduced. Mandatory notification creates two opposing effects on consumer surplus when this is the case—it simultaneously lowers the losses faced by the consumer in the event of a breach and raises the likelihood of breaches. The policy lowers consumer surplus when α is small; i.e., when the benefit from loss mitigation is small relative to the harm arising from the reduction in the website's investment.

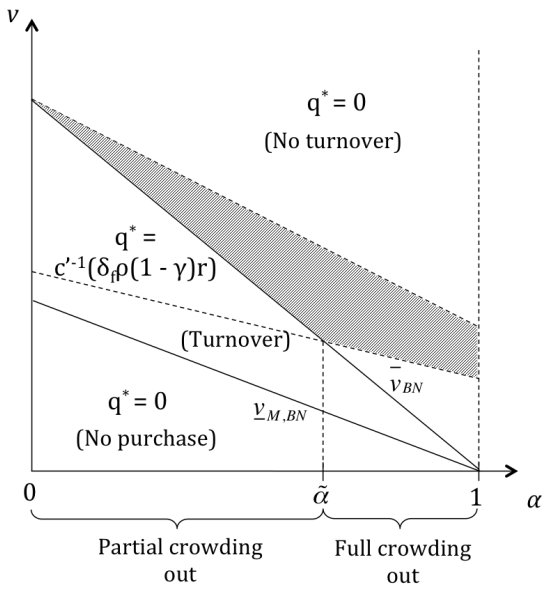

Figure 6: Impact of a mandatory notification law

5.1.2 Active Fraud Monitoring by the Bank

Fraud prevention by the consumer's bank can at times lower the reputation cost of data breaches to the website as it may weaken the consumer's ability to punish the firm. This occurs when the consumer does not learn about the breach whenever fraud is successfully prevented and, as a result, continues to purchase from the breached firm at the second period. That said, this is not an inevitable consequence of fraud prevention in general, but rather, an outcome specific to certain types of measures. One can classify the fraud prevention measures/technologies available to the bank into two categories: *passive prevention* and *active monitoring* measures. Passive prevention measures reduce the ease of committing fraud once they are implemented; they do not require further action on the part of the bank or the consumer. These measures are likely to inhibit the detection of data breaches. A good example is the adoption of EMV cards. The chip-and-PIN technology used on EMV cards reduces the incidence of fraud by making it more difficult for criminals to create counterfeit cards using stolen card data. The criminals' failure to commit counterfeit card fraud, however, is not observed by the bank and the consumer. Consequently, the consumer never learns about the data breach whenever fraud is prevented. Active monitoring measures, by contrast, require (active) intervention of the bank or the consumer; therefore, the consumer is made aware of the breach whenever fraud is successfully prevented. The use of behaviorial analytics is an example of such a measure. The bank can scrutinize transaction attempts and require additional authorization from the consumer for transactions that appear suspicious (e.g., made from an unusual location or device). Another example is the adoption of a multi-factor authentication process, where a one-time PIN is sent via SMS to the consumer whenever his card is being used for an on-line transaction.

Consider an active fraud monitoring measure that of equal effectiveness as its passive counterpart (i.e., the fraud prevention rate γ is the same). Assume further

that the cost of both types of measures to the bank is the same. For a given initial reputation, the consumer's willingness to purchase and his willingness to punish the website for breaches are unaffected by the form of fraud prevention measures chosen by the bank, since his expected losses are the same under both types of measures. However, the type of measures does affect the website's profit. The website's profit for when the bank actively monitors fraud is given by

$$\pi_A(q;q_0) = \begin{cases} (1+\delta_f)r - c(q) & \text{if } v \geq \overline{v} \\ (1+\delta_f(1-(\lambda(1-\gamma)+\gamma)(1-q)\rho))r - c(q) & \text{if } v < \overline{v} \ \& \ q \geq \hat{q}_0 \\ 0 - c(q) & \text{if } v < \overline{v} \ \& \ q < \hat{q}_0. \end{cases}$$

and its optimal security level is

$$q_A^{BR}(q_0) = \begin{array}{ll} c'^{-1}(\delta_f(\lambda(1-\gamma)+\gamma)\rho r) & \text{if } v < \overline{v} \ \& \ q \geq \hat{q}_0 \\ 0 - c(q) & \text{otherwise.} \end{array}$$

Observe that the website's optimal security level is higher when the bank engages in active fraud monitoring instead of passive fraud prevention. This is because active monitoring raises, rather than reduces, the probability of breach detection. As a result, the reputation cost of data breaches is higher. Further, notice that the optimal level of security chosen by the website is increasing in the effectiveness of the bank's monitoring; i.e., strategic complementarity exists.

Following the same steps as in the equilibrium analysis, we can establish that there exists a positive investment equilibrium for $v \in [\underline{v}_{M,A}, \overline{v})$, where

$$\underline{v}_{M,A} = (1 - c'^{-1}(\delta_f(\lambda(1-\gamma)+\gamma)\rho r))\overline{v}.$$

The website's security level at this equilibrium is given by

$$q_A^* = c'^{-1}(\delta_f(\lambda(1-\gamma)+\gamma)\rho r).$$

Notice that $\underline{v}_{M,A} < \underline{v}_M$, which implies that the consumer is more willing to participate at equilibrium when the bank actively monitors fraud. This is because the website invests more at equilibrium (i.e., $q_A^* > q^*$). Further, the consumer is always weakly better off under active monitoring relative to passive prevention. A consumer who purchases under both types of measures ($v \geq v_M$) obtains weakly higher surplus since he faces weakly lower fraud losses due to the increase in security level. Moreover, a consumer with $v \in [\underline{v}_{M,A}, \underline{v}_M)$ does not purchase from the website in the first period under passive prevention but purchases and obtains positive utility under active monitoring. The next proposition sums up the above discussion.

Proposition 3 (Active Fraud Monitoring)**.**
Consider the scenario where the bank engages in active monitoring for fraud prevention and suppose that the costs and the effectiveness of passive fraud prevention and active fraud monitoring are the same.
(i) The website invests weakly more for all $v \in \mathcal{R}^+$ under active fraud monitoring by the bank; more specifically, it invests (strictly) more when $v \in [\underline{v}_{M,A}, \overline{v})$.

(ii) The website's optimal level of security q_A^* is increasing in the bank's fraud prevention ability γ over $[\underline{v}_{M,A}, \overline{v})$.
(iii) Consumer surplus is higher for all $v \in \mathcal{R}$ under active fraud monitoring relative to passive fraud prevention.

Figure 7 illustrates the equilibrium outcomes under active fraud monitoring for different levels of fraud prevention effectiveness γ. The dotted line in the figure represents the valuation threshold above which the consumer purchases at $t = 1$ when the bank adopts passive fraud prevention measures.

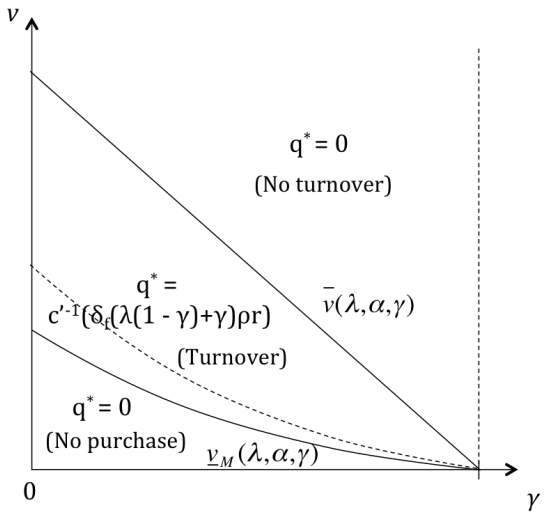

Figure 7: Equilibrium outcomes under active fraud monitoring

Several implications emerge from the above analysis. For any level of fraud prevention effectiveness, the probability that the consumer suffers a loss when purchasing from the website is lower under active fraud monitoring by the bank. In addition, the bank is always better off investing in an active fraud monitoring measure than in a passive prevention measure of the same cost effectiveness. This means that the bank would only choose passive fraud prevention over active fraud monitoring if it was less costly or more effective (or both). In the case where the lower cost effectiveness of active fraud monitoring stems from higher costs, regulators may find it optimal to either subsidize or mandate the adoption of active fraud monitoring measures.[36]

[36]Another reason that may the limit cost effectiveness of active monitoring is the difficultly in identifying the source of a fraud attempt. This may arise when the consumer interacts with multiple firms (which is not considered in my model), each of which has some probability of being breached. When this is the case, the consumer only learns with some probability whenever a fraud is prevented; this implies a smaller benefit to active monitoring. Consider the case where there are multiple (identical) banks serving (identical) consumers. Regulators may want to introduce a policy of information sharing between banks in this case. By sharing fraud intelligence, the banks will be able to look for patterns in fraud activities and use these patterns to help them identify the breached firm.

5.1.3 Suspension or Expulsion from Card Networks

The reputation cost of data breaches may also be affected by industry self-regulation. The Payment Card Industry Data Security Standard (PCI DSS) in the United States serves an example. The PCI DSS is a set of technical and operational standards that were developed jointly by American Express, Discover, JCB International, Visa and Mastercard, for safeguarding cardholder data. Under this self-regulation, data breaches may lead to adverse consequences for a merchant. Notably, a breached merchant that is found to be non-compliant with the PCI DSS at the time of the compromise may be suspended or expelled from the payment card network.

Suppose that the website is subject to such a policy. Following a data breach, the website loses its ability to accept card-based payments and will have to propose alternative, less convenient payment methods. Let τ denote the inconvenience cost that these payment methods imposes on consumers. First, notice that the policy has no impact on the website and the consumer at the first period. It only affects the consumer at the second period following the detection of a breach in the preceding period. In the case where the website is suspended or expelled from the card network, the consumer's expected utility from purchasing at $t = 2$ is given by

$$E(U_2|\text{Breach detected}) = v - \tau - \overline{v}$$

Because of the additional inconvenience cost, the consumer is less willing to purchase (or more willing to punish the website) following the detection of a data breach. He only continues to buy from the website if his valuation exceeds $\overline{v}+\tau$. Therefore, the policy induces the website to invest even when the consumer has a relatively high valuation for its product. Specifically, the website invests over the interval $[\overline{v}, \overline{v}+\tau)$ in the presence of the policy but would not have done so otherwise. Conditional on investing, however, the policy has no impact on the optimal security level chosen by the website since it does not affect the consumer's ability to detect and to punish the firm for data breaches.

The equilibrium outcomes of this game under a policy of potential suspension or expulsion is identical to that in the baseline set-up, with $\overline{v}$ being replaced with $\overline{v} + \tau$. The optimal security level chosen by the website at equilibrium is:

$$q_E^* = \begin{cases} c'^{-1}(\delta_f \lambda \tilde{\rho} r) & \text{if } v \in [\underline{v}_M, \overline{v} + \tau) \\ 0 & \text{otherwise.} \end{cases}$$

This brings us to the next proposition.

Proposition 4 (Suspension or Expulsion from Card Network)**.**
Consider an industry self-regulatory policy that suspends or expels breached merchants from the payment card network.
(i) The website invests weakly more for all $v \in \mathcal{R}^+$ under the policy; more specifically, it invests (strictly) more for $v \in [\overline{v}, \overline{v} + \tau)$.
(ii) The positive impact of the regulation, as captured by the size of the interval $[\overline{v}, \overline{v} + \tau)$, is increasing in the inconvenience cost τ and the breach detection rate λ but decreasing in the bank's fraud prevention ability γ.

(iii) Consumer surplus is weakly higher if and only if v sufficiently small; i.e., $v < \hat{v}$, where $\hat{v} \in (\overline{v}, \overline{v} + \tau]$.

The positive investment impact of the policy is bigger, the higher the inconvenience cost associated with switching to an alternative payment method. This suggests that the policy may be more effective for raising the investment incentives of on-line firms, which are heavily reliant on card-based payment methods, as compared to their brick-and-mortar counterparts, which can also easily accept cash and check payments. Additionally, the policy will only be effective insofar as data breaches are detected. Data breaches are more likely to be detected when they are more likely to generate fraud losses (lower γ) and when these losses are more likely to be discovered by the consumer (higher λ). For the above analysis, I have implicitly assumed that the card network learns about data breaches only when the consumer reports the associated fraudulent charges. In a more general setting, the card network (represented by the bank in my model) could also attempt to detect data breaches. For example, the bank could invest in active fraud monitoring measures as discussed in the previous section.[37] In fact, since the policy only raises the website's willingness to invest at the higher valuation levels and does not otherwise alter its investment incentives, imposing this policy on top of active fraud monitoring by the bank would unambiguously lead to even stronger investment incentives for the website.

Although this policy weakly improves the website's investment incentives, the consumer may be made worse off. This occurs when the consumer's valuation is very high (i.e., $v \geq \overline{v} + \tau$), such that he continues to always purchase from the website even when such a policy is implemented. When this is the case, the website continues not to invest but the consumer now incurs an extra inconvenience cost when purchasing from the website following the detection of a data breach. More generally, this analysis highlights the potential tradeoff between higher consumer surplus (*ex post*) and more security investment (*ex ante*) when relying on the reputation mechanism to provide investment incentives. The firm has stronger incentives to invest when the consumer is more willing to punish the firm; i.e., when the consumer faces higher losses and costs from data breaches.

5.2 Improving Consumer Information

So far, I have examined policy measures that aim to improve the website's investment incentives indirectly by raising the reputation cost of data breaches. I now consider policy measures that directly address a root cause of the lack of investment incentives—imperfect information.

[37]Note, however, that if the breach detection measures also reduces the incidence of fraud, it will affect the consumer's purchasing strategies and also the website's optimal level of investment.

5.2.1 Revealing the True State of Security

Suppose that the website can observe its state of security and consider a regime under which it is obliged to truthfully reveal this state to the consumer.[38] The timing of the game with the truthful revelation of security state is as follows:

- $t = 0$: The website decides the amount, $c(q)$, to invest in data security. It learns the outcome of its investment (secure or vulnerable) and reveals it to the consumer.
- $t = 1$: The consumer decides whether or not to purchase from the website.

Period $t = 2$ is a repetition of $t = 1$.[39]

Consider first the consumer's problem. The consumer purchases from a secure website for all $v \in \mathcal{R}_+$; he purchases from a vulnerable website if and only if his valuation lies above $\overline{v}$. Given the consumer's purchasing strategy, the website's expected profit under truthful revelation is

$$\pi_R(q) = \begin{array}{ll} q(1+\delta_f)r + (1-q)0 - c(q) & \text{if } v < \overline{v} \\ (1+\delta_f)r - c(q) & \text{otherwise.} \end{array}$$

Correspondingly, its optimal level of security is

$$q_R^* = \begin{array}{ll} c'^{-1}((1+\delta_f)r) & \text{if } v < \overline{v} \\ 0 & \text{otherwise.} \end{array}$$

For all $v < \overline{v}$, the website is penalized more heavily for poor security under a policy of truthful revelation. In the absence of such a policy, a vulnerable website is only punished at $t = 2$ (via customer turnover) if a breach occurs and is detected by the consumer. When its state is revealed, however, the consumer never purchases the website obtains no revenue at both periods. Therefore, the website has stronger incentives to invest to lower the probability that it is vulnerable to cyber-attacks. The following proposition sums up the above discussion.

Proposition 5 (Truthful Revelation of the State of Security)**.**
Suppose that the website can observe its state of security and consider a regime under which it is obliged to reveal this state to the consumer.
The website invests weakly more for all $v \in \mathcal{R}^+$ when it reveals its true state of security; more specifically, it invests (strictly) more for $v \in [0, \overline{v})$.

[38] In the model set-up, I have assumed that the website does not know its true state of security; it only knows how much it has invested into data security. This is a likely to be the case in reality, as most firms do not have the in-house expertise to evaluate whether it is indeed secure against cyber-attacks. In order to learn its true state of security, a firm may have to engage security experts to perform penetration testing on their system.

[39] Whether or not the consumer interacts with the website across more than one period has no strategic implications when the state of security is revealed to the consumer before he makes his initial purchase. This is because his interaction with the firm will not bring more information about the website's state of security. The additional period would simply be a repetition of period $t = 1$.

There are several challenges to the implementation of such a regime. First, while the website knows how much it has invested in security, it does not know if it is truly secure against data breach attacks. Second, even if the website could affirm its actual state of security (by means of penetration testing, for instance), the average consumer is unlikely to possess the knowledge and expertise to verify if the website is telling the truth. Suppose that the consumer always believes in the website's report. The website's dominant strategy would then be to always report that it is secure. This is clearly true when the website is secure. It is also easy to show that this is also the case when it is vulnerable—its expected revenue would be positive instead of zero in doing so.[40]

5.2.2 Revealing the Amount of Investment

The website's lack of ability to verify its true state of security poses a challenge to implementing the revelation regime previously discussed. As an alternative, the regulator could mandate the website to reveal its amount of security investment $c(q)$ (or equivalently, its choice of security level q). Although the consumer would remain imperfectly informed about the website's state of security under this regime, he would be better informed than in the absence of interventions.

Consider the following game:

- $t = 0$: The website decides on the amount, $c(q)$, to invest in data security and reveals its choice of security level q to the consumer.
- $t = 1$: The consumer decides whether or not to purchase from the website given its revelation. Conditional on purchasing, the consumer updates his beliefs at the end of the period depending on whether or not he has detected a breach.
- $t = 2$: The consumer decides whether or not to purchase from the website given his updated beliefs.

For any given level of valuation $v \in [0, \overline{v})$, the consumer purchases at $t = 1$ if the amount of security investment made by the website is sufficiently high, or equivalently, if q exceeds the following threshold

$$\underline{q}(v) = 1 - \frac{v}{\overline{v}}.$$

Observe that $\underline{q}$ is decreasing in v—the lower the valuation of the consumer, the higher the amount of security investment the website needs to make in order to induce the consumer to purchase.

[40] Its expected revenue when falsely reporting is given by

$$(1 + \delta_f)r - \delta_f \lambda \tilde{\rho} r > 0.$$

The second term of the above expression corresponds to the revenue in the case where the consumer detects a breach at $t = 1$ and consequently stops purchasing from it at $t = 2$.

Consider now the website's investment problem. The website's profit as a function of its security level is

$$\hat{\pi}_R(q) = \begin{cases} (1+\delta_f)r - c(q) & \text{if } v \geq \overline{v} \\ (1+\delta_f(1-\lambda(1-q)\tilde{\rho}))r - c(q) & \text{if } v < \overline{v} \ \& \ q \geq \underline{q}(v) \\ 0 & \text{otherwise.} \end{cases}$$

Let us focus on the case where $v < \overline{v}$. In the absence of any revelation, the website's optimal choice of q when $q \geq \underline{q}(v)$ is $q^* = c'^{-1}(\delta_f \lambda \rho(1-\gamma)r)$. There could be two scenarios at the website's investment stage of the game. In the first scenario, the website's unconstrained optimal choice of security lies above the threshold (i.e., $q^* \geq \underline{q}(v)$). In this case, the website's unconstrained and constrained choices of security level coincide (i.e., $\hat{q}_R^* = q^*$). In the second scenario, the consumer is not willing to purchase at $q = q^*$. Since the website's profit is decreasing in q for all $q > q^*$, its optimal level of security if it wants to induce the consumer to purchase is $q = \underline{q}(v)$. It is profitable for the website to serve a consumer if

$$(1+\delta_f(1-\lambda(1-\underline{q}(v)\tilde{\rho}))r - c(\underline{q}(v)) \geq 0,$$

or equivalently, if $v \geq \underline{v}_{M,R}$, where $\underline{v}_{M,R} \in [0, \overline{v})$ is the valuation threshold below which the website prefers not serving the consumer.[41] The website's optimal security level when it reveals its amount of investment is therefore given by

$$\hat{q}_R^* = \begin{cases} \underline{q}(v) & \text{if } v \in [\underline{v}_{M,R}, \underline{v}_M) \\ q^* & \text{if } v \in [\underline{v}_M, \overline{v}) \\ 0 & \text{otherwise.} \end{cases} \tag{2}$$

This leads us to the next proposition.

Proposition 6 (Revelation of the Amount of Security Investment)**.**
Consider a regime under which the website is obliged to reveal its amount of security investment by reporting its choice security level q to the consumer.
The website invests weakly more for all $v \in \mathcal{R}^+$ under the regime as compared to when there is no revelation, but it invests weakly less as compared to when it reveals its state of security; i.e. $q^* \leq \hat{q}_R^* \leq q_R^*$.

The relation between $\hat{q}_R^*$, q_R^* and q^* reflects how well informed the consumer about the website's state of security in each of the corresponding scenarios. The better informed the consumer is, the stronger the incentives for the website to invest in security.

[41] It can be shown such a valuation threshold does exist. First, observe that $\hat{\pi}_R$ is continuous in q and $\underline{q}$ is decreasing and continuous in v, which implies that $\hat{\pi}_R$ is a decreasing and continuous function of v. Moreover, $\hat{\pi}_R(0) < 0$ (since $\lim_{q \to 1} c(q) = \infty$) and $\hat{\pi}_R(\underline{v}) > 0$. Therefore, there exists a threshold, $\underline{v}_R$ below which $\hat{\pi}_R(v) < 0$. This threshold solves

$$(1+\delta(1-\lambda(1-\underline{q}(v)\tilde{\rho}))r - c(\underline{q}(v)) = 0.$$

5.2.3 Implementation via Certification

In practice, the revelation of either the state of security or the amount of investment can be implemented through certification. To reveal its state of security, the website can turn to a third-party security seal provider such as McAfee or Symantec. The seal provider will first perform an automated scan of the website for vulnerabilities, which will reveal its true state of security (assuming that the scan perfectly detects all vulnerabilities). If no vulnerabilities were discovered, the website will be able to display the security seal (at any desired location) on its page.[42] To reveal its amount of investment, the website can obtain certification against a set of information security standards. Examples of such security standards include the ISO/IEC 27001:2013, the ISAE 3402 and the PCI DSS. Under these certification programs, the website is typically assessed for its conformity with the best practices and/or compliance with the minimum security requirements as specified in the standards. Notice that this form of certification only indicates whether the firm's security investment has been sufficient for meeting the security standard but does not reveal the actual amount of investment.

There are several issues and challenges with the implementation of truthful revelation via certification. First, it may not always be in the best interest of a firm to participate in a certification program. The revelation of the state of security may lower the website's profit.[43] The website's equilibrium profit under a regime where its state is revealed is given by

$$\pi_R^* = \begin{cases} (1+\delta_f)r & \text{if } v \geq \overline{v} \\ q_R^*(1+\delta_f)r - c(q_R^*) & \text{otherwise,} \end{cases}$$

where $q_R^* = c'^{-1}((1+\delta_f)r)$. This is strictly lower than that in the absence of revelation when $v \in [\underline{v}_M, \overline{v})$. Thus, when the state of security is not known to the website (and therefore the consumer cannot make any inference about it from the website's certification decision), it will not voluntarily obtain certification when $v \in [\underline{v}_M, \overline{v})$. Moreover, even when certification does raise profit, the increase in profit may not be sufficient to cover its cost. Since the participation in most certification programs is voluntary,[44] it may be necessary for regulators to mandate certification. Second, the trustworthiness of these certification programs, in particular those of security seals, is questionable. Van Goethem et al. (2014) ran a vulnerable webshop experiment

[42] The security seal is typically provided to the firm in the form of a snippet of HTML code.

[43] By contrast, the website's equilibrium profit when it is obliged to reveal q, which is is given by

$$\hat{\pi}_R^* = \begin{cases} (1+\delta_f)r & \text{if } v \geq \overline{v} \\ (1+\delta_f(1-\lambda(1-\hat{q}_R^*)\tilde{\rho}))r - c(\hat{q}_R^*) & \text{if } v \in [\underline{v}_{M,R}, \overline{v}) \\ 0 & \text{otherwise,} \end{cases}$$

where $\hat{q}_R^*$ is as defined is in (2), is always weakly higher than that in the absence of revelation. Hence, the website will obtain certification even in the unregulated market, provided that the cost of certification is not too high.

[44] An exception to this is the PCI DSS. A merchant can only accept payment by cards if it is PCI certified. Note also that, in the B2B context, it is common for clients demand a firm to show that it has obtained a security certification (in particular, the ISO 27001 and the ISAE 3402) before they are willing to conduct business with the firm.

to analyze the certification methods of ten popular seal providers. They found that the best-performing provider only discovered less than half of the vulnerabilities on the webshop. The authors also compared the security hygiene of seal-bearing websites to that of non-certified websites and found no significant differences between the two groups. Both of these findings put to question the accuracy and effectiveness of the existing security seal programs. Third, certification against a set of standards will not provide meaningful information to the consumer if the standards are too demanding or too lax. Determining the appropriate standards to set may be difficult in practice. Finally, the consumer may have little or no awareness and understanding of third-party certification. Kim et al. (2008) found that 73.7 % of the participants in their study were unaware that the websites they have visited had been certified by third parties, despite being relatively active online consumers. This suggests that the impact third-party certification on security investment incentives is likely to be limited in the absence of consumer education.

5.3 Increasing the Direct Cost of Data Breaches

The website's investment in security lowers the probability of data breach and hence the resulting fraud losses. However, this benefit is not internalized by the website since it does not bear these losses. One means of alleviating this externality problem is therefore by making the website directly responsible for (part of) the losses resulting from data breaches. This could be achieved via the implementation of a liability rule or through the imposition of lump-sum penalties.

Consider first a liability rule under which the website is made responsible for a share β of the losses generated by a breach. This share β of losses could be either shifted from the consumer or the bank or both. Suppose that the liability is shifted from the bank, such that the policy has no direct impact on the consumer's purchasing decision.[45] The website's profit function under the liability rule is

$$\pi_L(q;q_0) = \begin{cases} (1+\delta_f)(r-\lambda(1-q)\tilde{\rho}\beta l) - c(q) & \text{if } v \geq \overline{v} \\ (1+\delta_f(1-\lambda(1-q)\tilde{\rho})r & \\ \quad -\lambda(1-q)\rho(1-\gamma)(1+\delta_f(1-\lambda\tilde{\rho}))\beta l - c(q) & \text{if } v < \overline{v} \;\&\; q_0 \geq \hat{q}_0 \\ 0 - c(q) & \text{if } v < \overline{v} \;\&\; q_0 < \hat{q}_0. \end{cases}$$

Its corresponding best-response function is

$$q_L^{BR}(q_0) = \begin{cases} c'^{-1}((1+\delta_f)\lambda\tilde{\rho}\beta l) & \text{if } v \geq \overline{v} \\ c'^{-1}(\delta_f\lambda\tilde{\rho}r + \lambda\tilde{\rho}(1+\delta_f(1-\lambda\tilde{\rho})\beta l)) & \text{if } v < \overline{v} \;\&\; q_0 \geq \hat{q}_0 \\ 0 & \text{if } v < \overline{v} \;\&\; q_0 < \hat{q}_0. \end{cases}$$

[45] Another reason for choosing to focus on the case where liability is reallocated from the bank rather than the consumer is that most of the fraud losses are incurred by the bank in reality (due to its fraud liability policy). This is also the reason why class-actions lawsuits—which can be considered as a market-based alternative to a liability rule—brought against a breached firm by consumers typically do not fare as well as those brought by financial institutions.

Assumption 2. *The website's sales revenue exceeds the breach liability it expects to incur when it is vulnerable; i.e.,*

$$r > \lambda\tilde{\rho}\beta l.$$

The website's best response in the case where $v \geq \overline{v}$ is lower than that in the case where $v < \overline{v}$ and $q_0 \geq \hat{q}_0$ whenever Assumption 2 is satisfied.

It is easy to verify that there exists an unique equilibrium in which the website chooses a security level of $q_L^* = c'^{-1}((1+\delta_f)\lambda\tilde{\rho}\beta l)$ and the consumer always purchases when $v \geq \overline{v}$.[46] When $v < \overline{v}$, there always exists an equilibrium where the website does not invest and the consumer does not purchase at any period (i.e., the no investment equilibrium). Further, when $v < \overline{v}$ and $q_0 \geq \hat{q}_0$, there may also exists a positive investment equilibrium, where the website's security level is given by $q_L^* = c'^{-1}(\delta_f\lambda\tilde{\rho}r + \lambda\tilde{\rho}(1 + \delta_f(1 - \lambda\tilde{\rho})\beta l)$. This equilibrium exists if $q_L^{BR}(\hat{q}_0) \geq \hat{q}_0$. This is the case whenever

$$v \geq (1 - c'^{-1}(\delta_f\lambda\tilde{\rho}r + \lambda\tilde{\rho}(1 + \delta_f(1 - \lambda\tilde{\rho})\beta l))\overline{v} = \underline{v}_{M,L}.$$

At this equilibrium, the consumer purchases at the first period and continues to do so in the second if and only if he did not detect a breach. Just as in the baseline model, it can be shown that the positive investment equilibrium Pareto dominates the no investment equilibrium. Therefore, I again assume that the website and the consumer coordinate on the positive investment equilibrium whenever both equilibria exist. The equilibrium security level of the website is given as follows:

$$q_L^* = \begin{cases} c'^{-1}((1+\delta_f)\lambda\tilde{\rho}\beta l) & \text{if } v \geq \overline{v} \\ c'^{-1}(\delta_f\lambda\tilde{\rho}r + \lambda\tilde{\rho}(1 + \delta_f(1 - \lambda\tilde{\rho})\beta l)) & \text{if } v \in [\underline{v}_{M,L}, \overline{v}) \\ 0 & \text{otherwise.} \end{cases}$$

This brings us to the next proposition.

Proposition 7 (Liability Rule)**.**
Consider a regulation under which the website is liable for a fraction $\beta < \alpha$ of the total fraud losses and suppose that these losses are reallocated away from the bank. The website invests weakly more under the liability rule for all $v \in \mathcal{R}^+$; more specifically, it invests (strictly) more for $v \geq \underline{v}_{M,L}$.

Under the rule, the website always incurs liability of βl whenever a breach occurs and is detected. Therefore, it invests even when data breaches do not lead to customer turnover (i.e., when $v \geq \overline{v}$). That said, the effectiveness of this policy once again depends on the likelihood of breach detection. The website's expected breach liability is small when the detection rate is low.

An alternative to implementing a liability rule is imposing lump-sum penalties for data breaches. This could take the form of fines or higher card processing fees

[46]First, notice that purchasing is a dominant strategy for the consumer whenever $v \geq \overline{v}$. Given that the consumer always purchases at equilibrium, the website's profit is a convex function of q; therefore, an unique maximizer exists.

(imposed by the payment card network).[47] Imposing a lump-sum fine amounting to βl will result in the same outcome as the liability rule. Increasing the card processing fees of a firm with a history of data breaches will also raise the amount of security investment. However, the impact of the increase in fees will be smaller as compared to a fine or liability of equal magnitude, since the firm only incurs the higher fees at $t = 2$.[48]

6 Extensions

6.1 Strategic Bank

In the baseline model, I assume the bank to be a non-strategic player and its fraud prevention ability γ to be exogenously given. In this extension, I endogenize the bank's decision to invest in fraud prevention. More specifically, I consider the scenario where the bank has some existing fraud prevention measures in place and has to decide whether to adopt a new technology, in face of the evolving threat landscape (due to digitization).[49]

6.1.1 Model Set-up and Equilibrium

Consider the following three-stage game. In the first stage, the bank decides whether to upgrade its fraud prevention technology. The fraud prevention rates of its existing and the new technologies are denoted γ_0 and γ_n respectively, where $0 < \gamma_0 < \gamma_n < 1$. The new technology comes at a fixed cost F. In the second stage, the website decides on its optimal level of security investment, having observed the choice made by the bank. In the third stage, the consumer decides whether to purchase at each period given the technology chosen by the bank and the website's reputation for security.

This game can be solved by backward induction. The sub-game following the bank's investment stage is identical to the website's investment game in the baseline model. The equilibrium outcomes are as described in Proposition 1, with γ taking on a value of either γ_0 or γ_n, depending on the bank's decision at the first stage of the game. As before, I suppose that the website and the consumer coordinate on the positive investment equilibrium when both equilibria exist. Consider now the bank's investment stage of the game. For simplicity, suppose that the bank's

[47] For instance, the payment card network could increase the interchange fee paid by the merchant's bank, which would then pass the increase on to the merchant by raising the merchant discount fee.

[48] Normalize the fees at $t = 1$ to zero and suppose the fee is raised by βl following the detection of a breach. The website's profit when the merchant fee is contingent on its breach history given by

$$\pi(q; q_0) = \begin{cases} (1+\delta_f)r - \delta_f\lambda(1-q)\tilde{\rho}\beta l - c(q) & \text{if } v \geq \overline{v} \\ (1+\delta_f(1-\lambda(1-q)\tilde{\rho})r & \\ \quad -\delta_f\lambda(1-q)\tilde{\rho}\beta l - c(q) & \text{if } v < \overline{v} \ \&\ q_0 \geq \hat{q}_0 \\ 0 - c(q) & \text{if } v < \overline{v} \ \&\ q_0 < \hat{q}_0. \end{cases}$$

[49] The decision to replace magnetic strip cards with chip-and-PIN cards constitutes one such example.

objective is to minimize the expected fraud liability arising from data breaches at the website. Let $\phi(\gamma)$ denote the bank's expected liability for a given choice of technology:

$$\phi(\gamma) = \begin{cases} (1+\delta_b)\rho(1-\gamma)\lambda\alpha l & \text{if } v \geq \overline{v} \\ \lambda(1-q^{BR}(\gamma))\rho(1-\gamma)(1+\delta_b(1-\lambda\rho(1-\gamma)))\alpha l & \text{if } v \in [\underline{v}, \overline{v}) \\ 0 & \text{if } v < \underline{v}, \end{cases}$$

where δ_b is the discount factor of the bank.

I focus on the case where the website invests in the sub-game following the bank's investment stage (i.e., when $\underline{v} \leq v < \overline{v}$).[50] Let $\Delta\gamma = \gamma_n - \gamma_0$ and $\Delta q^{BR} = q^{BR}(\gamma_n) - q^{BR}_{S,L}(\gamma_0)$. The change in the bank's expected liability from adopting the new technology in that case is

$$\begin{aligned} \Delta\phi =& \phi(\gamma_n) - \phi(\gamma_0) \\ =& \lambda(1-q^{BR}(\gamma_0))\rho((1+\delta_b)(1-\gamma_n-(1-\gamma_0)) - \delta_b\lambda\rho((1-\gamma_n)^2-(1-\gamma_0)^2)))\alpha l \\ & - \Delta q^{BR}(1-\gamma_n)(1+\delta_b(1-\lambda\rho(1-\gamma_n))\alpha l \\ =& \underbrace{-\lambda(1-q^{BR}(\gamma_0))\rho\Delta\gamma(1+\delta_b-\delta_b\lambda\rho(2-\gamma_0-\gamma_n)))\alpha l}_{\text{Direct effect}} \\ & \underbrace{-\lambda\Delta q^{BR}\rho(1-\gamma_n)(1+\delta_b(1-\lambda\rho(1-\gamma_n)))\alpha l.}_{\text{Indirect effect}} \end{aligned}$$

The direct effect captures the impact of its investment on its expected fraud liability, holding the website's investment level fixed. It is negative because the new technology reduces the success rate of fraud attempts. The indirect effect arises from the strategic reduction in the website's security investment—the bank's investment acts as a strategic substitute to the website's. This effect is positive because a lower level of data security investment implies a higher probability of data breaches. The bank finds it optimal to upgrade to the new technology if the "total cost" of investing (the sum of the fixed cost and the indirect effect) lies below the (direct) benefit:

$$\begin{aligned} & F - \lambda\Delta q^{BR}\rho(1-\gamma_n)(1+\delta_b(1-\lambda\rho(1-\gamma_n))\alpha l \\ \leq & \lambda(1-q^{BR}(\gamma_0))\rho\Delta\gamma(1+\delta_b-\delta_b\lambda\rho(2-\gamma_0-\gamma_n))\alpha l. \end{aligned} \tag{3}$$

The above condition is more likely to be satisfied when the fixed cost of investment is small or when the website's investment cost function is sufficiently convex at $q = \delta_f\lambda(1-\gamma_0)r$. Observe that it is not necessarily the case that the bank adopts the new technology when there is no costs to doing so (i.e., when $F = 0$). The bank only considers implementing the new technology when the the strategic effect is

[50] The bank's investment problem is straightforward to solve in the other two cases. When $v < \underline{v}$, not upgrading is a dominant strategy for the bank. When $v \geq \overline{v}$, the bank upgrades to the new technology if the fixed cost of investment F lies below

$$(1+\delta_b)\rho(\gamma_0-\gamma_n)\lambda\alpha l$$

and does not upgrade otherwise.

relatively small compared to the direct effect; more precisely, when

$$(1-q^{BR}(\gamma_0))\Delta\gamma > -\Delta q^{BR}(1-\gamma_n)\underbrace{\frac{1+\delta_b(1-\lambda\rho(1-\gamma_n))}{1+\delta_b(1-\lambda\rho(2-\gamma_n-\gamma_0))}}_{>1}. \tag{4}$$

For any given investment cost function of the website, the condition in (3) is equivalent to

$$F \leq -\Delta\phi(\gamma) = \hat{F},$$

where $\hat{F} \geq 0$ if and only if the condition in (4) holds.[51]

Before characterizing the equilibria of this game, it is helpful to introduce a more comprehensive measure of the level of security. Let $\mu(q,\gamma)$ denote the level of security in the payment system. It is jointly determined by the website's security level and the bank's fraud prevention ability:

$$\mu(q,\gamma) = 1-(1-q)\rho(1-\gamma).$$

Proposition 8 (Equilibrium Investment Outcomes with Strategic Bank)**.** *Consider the case where $v \in [\underline{v}_M, \overline{v})$.*
(i) If $F > \hat{F}$, only the website invests at equilibrium. The equilibrium bank's fraud prevention rate and website's level of security are respectively

$$\gamma^* = \gamma_0 \quad \textit{and} \quad q_S^* = c'^{-1}(\delta_f\lambda\rho(1-\gamma_0)r),$$

and the overall level of security in the system is

$$\mu_S^* = 1-(1-q_S^*)\rho(1-\gamma_0).$$

(ii) If $F \leq \hat{F}$, both the bank and the website invest at equilibrium. The equilibrium bank's fraud prevention rate and website's level of security are respectively

$$\gamma^{**} = \gamma_n \quad \textit{and} \quad q_S^{**} = c'^{-1}(\delta_f\lambda\rho(1-\gamma_n)r),$$

and the overall level of security in the system is

$$\mu_S^{**} = 1-(1-q_S^{**})\rho(1-\gamma_n).$$

The overall level of system security is higher if the bank upgrades its technology when the condition in (4) holds.

[51] The overall system security level is increased when the bank adopts the new technology whenever

$$(1-q^{BR}(\gamma_0))\Delta\gamma > -\Delta q^{BR}(1-\gamma_n),$$

which is a weaker condition than that stated in (4).

6.1.2 Policy Implications

Mandatory Breach Notification
The impact of mandatory breach notification on the level of security in the payment system is a priori ambiguous in most cases when the bank's decision to invest is endogenous. Let us focus on the case where the website is willing to invest for any fraud prevention rate $\gamma \in \{\gamma_0, \gamma_n\}$ both in the presence and absence of a breach notification law. For any given fraud prevention rate, the website's optimal level of security is higher under the law; i.e.,

$$q_{BN}^{BR}(\gamma) = c'^{-1}(\delta_f \rho(1-\gamma)r) > q^{BR}(\gamma).$$

The impact of the mandatory breach notification on the bank's investment is, however, *a priori* ambiguous. Under the law, the reduction in the bank's expected fraud liability from investing in the new technology (i.e., the benefit of investment) is given by

$$\begin{aligned} -\Delta\phi_{BN} =&(1-q_{BN}^{BR}(\gamma_0))\Delta\gamma\rho(1+\delta_b(1-\rho(2-\gamma_n-\gamma_0))\alpha l \\ &- \Delta q_{BN}^{BR}(1-\gamma_n)\rho(1+\delta_b(1-\rho(1-\gamma_n)))\alpha l, \end{aligned}$$

where $\Delta q_{BN}^{BR} = q_{BN}^{BR}(\gamma_n) - q_{BN}^{BR}(\gamma_0)$. Consider the change in the benefit of investment due to the implementation of the law (i.e., $-\Delta\phi_{BN} - (-\Delta\phi)$). This change can be decomposed into a direct effect (holding the website's security level fixed):

$$\begin{aligned} &(1-\lambda)((1-q^{BR}(\gamma_0))\Delta\gamma\rho(1+\delta_b(1-(1+\lambda)\rho(2-\gamma_n-\gamma_0)) \\ &- \Delta q^{BR}(1-\gamma_n)\rho(1+\delta_b(1-(1+\lambda)\rho(1-\gamma_n))))\alpha l, \end{aligned}$$

and an indirect effect arising from the change in the website's security level:

$$\begin{aligned} &- (q_{BN}^{BR}(\gamma_0) - q^{BR}(\gamma_0))(\Delta\gamma\rho(1+\delta_b(1-\rho(2-\gamma_n-\gamma_0)) - (1-\gamma_n)\rho(1+\delta_b(1-\rho(1-\gamma_n))))\alpha l \\ &- (q_{BN}^{BR}(\gamma_n) - q^{BR}(\gamma_n))\Delta\gamma\rho(1+\delta_b(1-\rho(1-\gamma_n))\alpha l. \end{aligned}$$

The signs of both effects are ambiguous.

Proposition 9 (Mandatory Breach Notification (Strategic Bank))**.**
Mandatory breach notification raises the overall security level of system when
(i) the bank's investment decision is unchanged by the law, or
(ii) the law induces the bank to adopt the new technology and if the initial breach detection rate is sufficiently low.[52]
The impact of the regulation is a priori ambiguous otherwise.

When the bank's investment decision is unaltered, the level of system security is higher because of the increase in security investment by the website. In the case when the bank invests in the new technology under regulation but would

[52]More precisely, this is the case when

$$\lambda < \frac{1-\gamma_n}{1-\gamma_0}.$$

not have done so otherwise, the level of security is also unambiguously higher if the website invests more despite the increase in the fraud prevention rate (i.e., $q^{BR}(\gamma_0) < q^{BR}_{BN}(\gamma_n)$). This occurs when the substitution effect from the increase in fraud prevention rate is weak relative to the reputation effect of breach notification, whose strength is decreasing in the initial breach detection rate. In all other scenarios, the law results in opposing effects on the website's and the bank's investment decisions; therefore, the overall impact on the system security level is ambiguous.

Liability Rule

Consider now the liability rule that shifts βl of the losses incurred by the bank to the website (as discussed in the policy analysis section earlier). The profit and loss functions of the website and the bank after the reallocation of liability are respectively

$$\begin{aligned}\pi_{S,L}(q, q_0, \gamma) =&(1 + \delta_f(1 - \lambda(1 - q)\rho(1 - \gamma))r \\ &- \underbrace{\lambda(1 - q)\rho(1 - \gamma)(1 + \delta_f(1 - \lambda\rho(1 - \gamma)))\beta l}_{\text{Expected fraud liability}} - c(q),\end{aligned}$$

and

$$\phi_L(\gamma) = \lambda(1 - q^{BR}_{S,L}(\gamma))\rho(1 - \gamma)(1 + \delta_b(1 - \lambda\rho(1 - \gamma)))(\alpha - \beta)l.$$

The impact of this shift in liability on the website's investment decision (for a given fraud prevention rate) is presented in Proposition 7—conditional on investing, the website invests more in data security. As for its impact on the bank's investment decision, there could be two scenarios. In the first scenario, the bank's decision to invest in the new technology is unaffected by the policy. This occurs when

$$\begin{aligned}F <& -\Delta\phi_L \\ =& -\lambda(1 - q^{BR}_L(\gamma_0))\rho\Delta\gamma(1 + \delta_b - \delta_b\lambda\rho(2 - \gamma_0 - \gamma_n))(\alpha - \beta)l \\ &- \lambda\Delta q^{BR}_L\rho(1 - \gamma_n)(1 + \delta_b(1 - \lambda\rho(1 - \gamma_n)))(\alpha - \beta)l \\ =&\hat{F}_L,\end{aligned}$$

where $\Delta q^{BR}_L = q^{BR}_L(\gamma_n) - q^{BR}_L(\gamma_0)$. In the second scenario, which arises when the above condition is violated, the bank does not upgrade its technology after the reallocation of liability. Note also that the shift in liability strengthens the strategic substitutability between the bank's fraud prevention investment and the website's security investment; therefore, $\hat{F}_L < \hat{F}$.[53]

Proposition 10 (Liability Rule with Strategic Bank)**.**
Consider a liability rule that transfers βl of the fraud losses incurred by the bank to the website and suppose that $F \leq \hat{F}$ such that the bank upgrades its technology in the absence of the liability rule.

[53]The reduction in the website's security investment from a marginal increase in γ is

$$\frac{\partial\widetilde{MB}(q,\gamma)}{\partial\gamma} = -\delta_f\lambda\rho r - \underbrace{\lambda\rho(1 + \delta_f - 2\delta_f\lambda\rho(1 - \gamma))\beta l}_{>0} > \frac{\partial MB(q,\gamma)}{\partial\gamma}.$$

(i) If $F > \hat{F}$, the bank does not invest regardless of whether a liability rule is imposed. The equilibrium bank's fraud prevention rate and website's security level are respectively

$$\gamma_L^* = \gamma_0 \quad \textit{and} \quad q_L^* = c'^{-1}(\delta_f \lambda \rho (1-\gamma_0) r + \lambda \rho (1-\gamma_0)(1 + \delta_f (1 - \lambda \rho (1-\gamma_0)) \beta l))$$

and the overall level of security in the system is

$$\mu_{S,L}^* = 1 - (1 - q_L^*)\rho(1 - \gamma_0).$$

(ii) If $\hat{F}_L < F \leq \hat{F}$, the bank invests in the absence but not in the presence of the liability rule. The equilibrium outcomes under the liability rule correspond to those in part (i).
(iii) If $F \leq \hat{F}_L < \hat{F}$, the bank invests regardless of whether a liability rule is imposed. The equilibrium bank's fraud prevention rate and website's security level are respectively

$$\gamma_L^{**} = \gamma_n \quad \textit{and} \quad q_L^{**} = c'^{-1}(\delta_f \lambda \rho (1-\gamma_n) r + \lambda \rho (1-\gamma_n)(1 + \delta_f (1 - \lambda \rho (1-\gamma_n)) \beta l)$$

and the overall level of security in the system is

$$\mu_{S,L}^{**} = 1 - (1 - q_L^{**})\rho(1 - \gamma_n).$$

Corollary 5 (System Security Level under Liability Rule (Strategic Bank))**.**
The equilibrium level of system security is higher when
(i) the bank's investment decision is unaffected by the rule;
(ii) the bank does not invest under the rule but would have done so otherwise if

$$(1 - q_L^{BR}(\gamma_0))\Delta\gamma < -\Delta q_L^{BR}(1 - \gamma_n),$$

and may be ambiguous otherwise.

When the bank's investment decision is unaltered by the rule (i.e., either the bank always invests or never invests), the shift in liability unambiguously increases the equilibrium level of security in the payment system. In these cases, the liability rule only has the direct effect of increasing the cost of data breaches to the website, leading it to invest more in security. The impact of the liability rule on the equilibrium level of system security may be ambiguous when the bank chooses not to upgrade its technology under the rule but would have done so otherwise. On the one hand, the bank's decision not to upgrade increases the fraud rate which lowers the overall security level. On the other hand, the liability rule induces the website to invest more in data security due to the direct effect described above, as well as the indirect effect arising from the reduction in the bank's investment.[54]. The liability rule unambiguously raises the overall level of security when the increase in the website's security level in response to the bank's decision not to upgrade is relatively large. This is always the case when the condition stated in part (ii) of the above proposition is satisfied. Notice that when this condition holds, the bank will not invest in the new technology even when its cost of adoption is zero.

[54]This implies that the increase in the website's security level in this case is larger than those in parts (i) and (iii) of the Lemma 10

6.2 Multiple Firms

In this extension, I examine a firm's incentives to invest in data security in settings where there are multiple firms. To simplify the analysis, I assume throughout this extension that the consumer is myopic (i.e., $\delta_c = 0$) and there is no bank (i.e., $\gamma = 0$ and $\alpha = 0$).

6.2.1 Competing Firms

Suppose that there are two identical firms, i and j, each serving a representative consumer.[55] Let $q_{i,0}$ denote firm i's initial reputation for security. Consider the case where $v < \overline{v}$; i.e., the consumers are willing to punish the firms for poor security.

I first examine the scenario where data breaches are privately detected by the customer of a firm. This implies that a consumer who has purchased from firm i would not update his belief about firm j's security level (i.e., $q_{j,1} = q_{j,0}$). Suppose that each firm's initial reputation is sufficiently high such that each consumer is willing to purchase from his respective firm at $t = 1$.

Conditional on investing, each firm's optimal amount of investment depends only on its expected revenue at $t = 2$. For a given level of security chosen by the rival firm, q_j, firm i's expected revenue in period two is given by

$$R_{i,2}(q_i) = \delta_f(\underbrace{(1 - \lambda(1 - q_i)\rho)}_{\substack{\text{probability of} \\ \text{retaining own} \\ \text{customer}}} + \underbrace{\lambda(1 - q_j)\rho}_{\substack{\text{probability of} \\ \text{gaining rival's} \\ \text{customer}}})r.$$

The second term in the above expression corresponds to the additional revenue obtained by firm i as a result of j's customer switching over to it following the detection of a breach at j. The optimal level of investment is given by

$$q_i^{BR}(q_j) = c'^{-1}(\delta_f \lambda \rho r).$$

Firm i's optimal investment is independent of firm j's investment level. At equilibrium, each firm sets $q_C^* = c'^{-1}(\delta_f \lambda \rho r)$, which corresponds to the equilibrium security level under the single firm set-up. In other words, competition has no impact on security investment when data breaches are private signals.[56]

Consider now the case where data breaches are publicly announced whenever they are detected. Each firm's revenue at $t = 2$ is now given by

$$\tilde{R}_{i,2}(q_i) = \delta_f(\underbrace{(1 - \lambda(1 - q_i)\rho)}_{\substack{\text{probability of retaining} \\ \text{own customer}}} + \underbrace{\lambda(1 - q_j)\rho(1 - \lambda(1 - q_i)\rho)}_{\substack{\text{probability of gaining} \\ \text{rival's customer}}})r.$$

[55] This assumption can be easily micro-founded. For instance, each of the two consumer may an ex-ante preference for a different firm due to the differences in their relative exposure to each firm's advertisements. However, these ex-ante preferences do not affect a consumer's valuation for the firms' products; his valuation for the products of both firms is the same since the products are identical.

[56] A corollary to this result is that if competition reduces the firm's market share relative to the monopoly setting, the level of security under the competitive setting will also be comparatively lower.

Firm i only gains its rival's consumer when a breach was detected at its rival's but no breach detected at the firm itself. Firm i's best response function when signals are public is

$$\tilde{q}_i^{BR}(q_j) = c'^{-1}(\delta_f \lambda \rho(1 + \lambda(1 - q_j)\rho)r).$$

Observe that the firms' security investments are strategic substitutes. The reaction functions of the firms are illustrated in Figure 8.

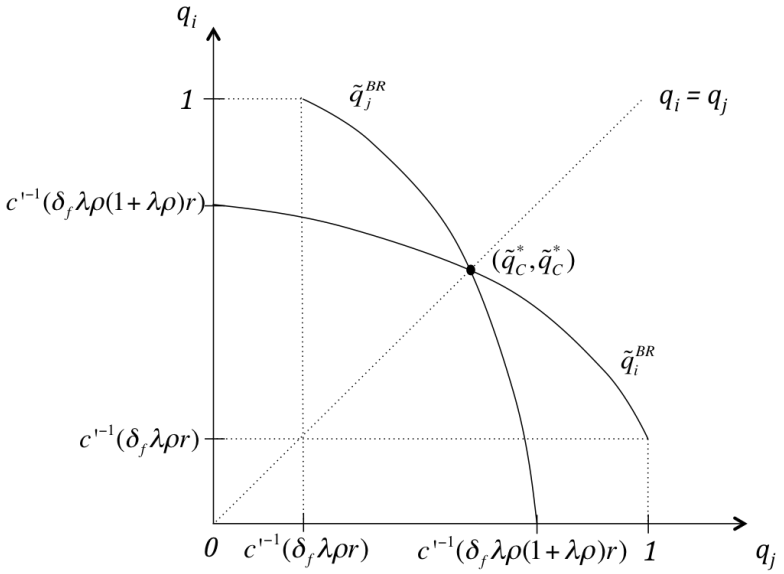

Figure 8: Best-response functions when signals are public

In equilibrium, the firms set $\tilde{q}_{i,C}^* = \tilde{q}_{j,C}^* = \tilde{q}_C^*$, where $\tilde{q}_C^*$ solves

$$\tilde{q}_C^* = c'^{-1}(\delta_f \lambda \rho(1 + \lambda(1 - \tilde{q}_C^*)\rho)r).$$

The level of security at the competitive equilibrium with public signals is higher than that under the monopoly setting. Observe further that the increase in security level due to competition is larger when the probability of breach detection is higher. This suggests that mandating the public disclosure of data breaches may be particularly useful in competitive markets. The following proposition sums up the above discussion.

Proposition 11 (Competing Firms)**.**
Consider a market with two identical firms and two representative consumers and suppose that each of the consumers is randomly assigned to a firm at $t = 1$.
(i) The equilibrium security level under competition is the same as that in the monopoly setting if data breaches are privately detected; i.e., $q_C^* = q^*$.
(ii) The equilibrium security level is weakly higher under competition if data breaches are made public whenever they are detected; i.e., $q_C^* \geq q^*$. Further, the increase security level is larger when the breach detection rate λ is higher.

6.2.2 Non-competing Firms

Consider now a simple example where there is one representative consumer and the two non-competing firms, i and j. The two firms are identical in all aspects and they offer products that are independent from the consumer's perspective. The consumer has valuations $v_i = v_j = v$ for the firms' product and decides whether to purchase from each firm at every period given the firms' reputations for security. All transactions are made using the same payment card.

I first examine the consumer's problem. Assume that $v < \rho l$ such that it is not optimal for a consumer to buy from a firm that he knows is vulnerable with certainty. For any given level of initial reputation of the firms $q_{i,0}$ and $q_{j,0}$, the consumer's purchase decision at $t = 1$ is unaffected by the presence of a non-competing firm—he purchases if his valuation is sufficiently high. The scenario of interest is the one where the consumer has purchased from both firms at $t = 1$. Under this scenario, his payment card would have been used at more than one location and, as such, he would not be able to perfectly identify the source of the breaches that he detects. The imperfect attribution of data breaches affects the updating of the consumer's beliefs. Firm i's updated reputation at $t = 1$, $q_{i,1}$, in this case is given by:

$$q_{i,1} = \begin{cases} \dfrac{q_{i,0}(1-q_{j,0})}{(1-q_{i,0})+(1-q_{j,0})-(1-q_{i,0})(1-q_{j,0})\rho} & \text{if a breach was detected} \\ \dfrac{q_{i,0}(1-\lambda(1-q_{j,0})\rho)}{1-\lambda((1-q_{i,0})+(1-q_{j,0})-(1-q_{i,0})(1-q_{j,0})\rho)\rho} & \text{if no breach was detected.} \end{cases}$$

Notice in particular how the firm's reputation following the detection of a breach differs from the monopoly setting. In the monopoly setting, the consumer knows that the firm is vulnerable with certainty whenever a breach is detected; therefore, $q_{i,1} = 0$. In the two-firm setting, however, a firm only receives part of the blame for the detected breach (imperfect attribution). Because the two firms share the blame for any detected breaches, the reputation damage suffered by the breached firm is reduced. Consequently, the consumer would be more willing to purchase from a breached firm (or equivalently, less willing to punish the firm) as compared to the monopoly setting. This means that the website would also be less willing to invest in the presence of a non-competing firm.

Consider the case when the consumer's valuations belong to the interval over which he is willing to punish a breached firm by leaving. How does the presence of a non-competing firm affect a firm's optimal security level? Given the firms' initial reputations $q_{i,0}$ and $q_{j,0}$ and firm j's level of investment q_j, firm i's expected revenue at $t = 2$ when it invests is

$$R_{i,2}(q_i) = (1 + \delta_f - \delta_f\lambda((1-q_i) + \underbrace{(1-q_j) - (1-q_i)(1-q_j)\rho}_{\text{Externality imposed by } j}\rho)r.$$

Due to the imperfect attribution of data breaches, firm i also incurs a reputation cost when firm j is breached. In other words, firm j investment imposes a positive externality on firm i. This size of this externality is larger when firm i's level of investment is higher. Firm i's best response is given by

$$q_i^{BR}(q_j) = c'^{-1}(\delta_f\lambda(1-(1-q_j)\rho)\rho r).$$

Observe that firm i's optimal investment level for a given reputation is increasing in firm j's investment—the firms' security investments are strategic complements. A unique symmetric equilibrium to this investment game exists if the investment cost function is sufficiently convex at $q = c'^{-1}(\delta_f \lambda \rho r)$; i.e.,

$$c''(c'^{-1}(\delta_f \lambda \rho r)) > \delta_f \lambda \rho^2 r.$$

The equilibrium level of security q^*_{NC} is the implicit solution to

$$q^*_{NC} = c'^{-1}(\delta_f \lambda (1 - (1 - q^*_{NC})\rho)\rho r).$$

Figure 9 provides a graphical illustration of the firm's best-response functions and the equilibrium to this game when the cost function is sufficiently convex.[57]

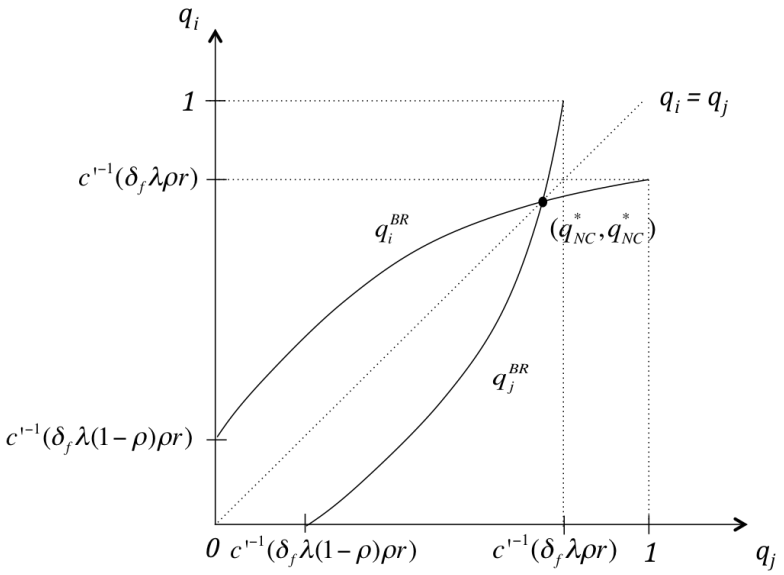

Figure 9: Best-response functions of non-competing firms

Proposition 12 (Non-Competing Firms)**.**
Consider an investment game with two identical firms selling independent products and suppose that the consumer purchases from both firms at $t = 1$.
The equilibrium security level in the non-competing firms setting is weakly lower as compared to the monopoly setting; i.e., $q^*_{NC} \leq q^*$.

The adverse impact of the presence of a non-competing firm arises from the imperfect attribution of data breaches. As discussed earlier, the imperfect attribution of breaches that arises in this setting implies that a firm may not experience a turnover when it is breached but may experience one when the other firm is breached (though it is not). Both of these factors weaken the firm's incentives to invest.

[57] In Appendix B, I prove that an equilibrium to the full game where both firms invest does exist under some restrictions on the set of parameter values.

7 Conclusion

In this paper, I examined the role of reputation in the data security investment decision of a firm. I consider a repeat-purchase setting where the firm can make a one-time investment in data security. The firm's state of security is initially unobserved but the consumer has rational beliefs about it; these beliefs constitute the firm's reputation for security. Further, the consumer learns about the state of security over time via the detection of data breaches. I show that reputation can indeed play a role in incentivizing a firm to invest in consumer data protection in this setting—the detection of data breaches leads to a decline in the firm's reputation, which may result in customer turnover. The role that reputation plays may, however, be limited in reality, as consumers may have little willingness or ability to punish a firm for poor security by leaving.

Policy makers can enhance the role of reputation by implementing measures that would raise the consumer's willingness or ability to punish the firm. However, they should exercise caution when doing so. I show in my analysis that security investment and consumer surplus can at times be reduced; specifically, this occurs when these measures affect the consumer's willingness to punish. For instance, when the consumer is initially unwilling to punish the firm (i.e., his gross utility from consumption exceeds his maximum expected losses), attempts to increase his willingness to punish can make him worse off. From the consumer's perspective, it is preferable for policy makers to adopt measures that improve investment incentives by directly addressing the market failures of imperfect information and externalities.

The ability of policy makers to improve investment incentives via the reputation channel may further be constrained by factors such as market power and the nature of the data that is compromised. When the firm is dominant in a market, the consumer may lack (good) outside options; this would be reflected by a high valuation level in my model. Consequently, he would continue patronizing the firm even after learning that it has been breached. The irreplaceable nature of certain types of data (e.g., date of birth and social security number) likewise weakens a consumer's incentives to terminate his relationship with a breached firm. Once compromised, there is little or no value to preventing future breaches of the same data by not choosing not to patronize the breached firm.

Finally, I conclude with a few suggestions for future work. In this paper, I have focused mainly on analyzing the indirect policy interventions—measures targeted at increasing the role of reputation. Future work may want to provide a more in-depth analysis of the direct measures, particularly the liability rule. A well-designed liability rule can be an effective and powerful instrument for incentivizing firms to invest in data security. One issue that warrant further investigation is therefore the optimal allocation of breach liability across the various parties: firms, consumers and financial institutions. Another aspect that may be explored in future work is the endogenization of attacks. Hackers may adjust their attack effort in response to a firm's investment level; this may be especially interesting in a framework with competing firms.

References

Ablon, L., P. Heaton, D. Lavery, and S. Romanosky (2016): "Consumer attitudes toward data breach notifications and loss of personal information," Tech. rep., RAND Corporation.

Ablon, L., M. C. Libicki, and A. A. Golay (2014): "Markets for Cybercrime Tools and Stolen Data: Hackers' Bazaar," Tech. rep., RAND Corporation.

Acemoglu, D., A. Malekian, and A. Ozdaglar (2016): "Network security and contagion," *Journal of Economic Theory*, 166, 536–585.

Acquisti, A. and J. Grossklags (2005): "Privacy and rationality in individual decision making," *IEEE Security & Privacy*, 2, 24–30.

Allen, F. (1984): "Reputation and product quality," *The RAND Journal of Economics*, 311–327.

Board, S. and M. Meyer-ter Vehn (2013): "Reputation for quality," *Econometrica*, 81, 2381–2462.

Campbell, K., L. A. Gordon, M. P. Loeb, and L. Zhou (2003): "The economic cost of publicly announced information security breaches: empirical evidence from the stock market," *Journal of Computer Security*, 11, 431–448.

Cavusoglu, H., B. Mishra, and S. Raghunathan (2004): "The effect of internet security breach announcements on market value: Capital market reactions for breached firms and internet security developers," *International Journal of Electronic Commerce*, 9, 70–104.

Cheney, J. S., R. M. Hunt, K. R. Jacob, R. D. Porter, and B. J. Summers (2012): "The efficiency and integrity of payment card systems: industry views on the risks posed by data breaches," *Economic Perspectives*, 36, 130–147.

Daughety, A. F. and J. F. Reinganum (2011): "Economic analysis of products liability: theory," .

Dybvig, P. H. and C. S. Spatt (1983): "Does it pay to maintain a reputation? Consumer information and product quality," *Financial Research Center Memorandum*.

Gemalto (2015): "Data breaches and customer loyalty report," Available: `https://safenet.gemalto.com/resources/data-protection/data-breaches-customer-loyalty-report/`.

——— (2016): "Breach level index report 2016," Available: `http://breachlevelindex.com/assets/Breach-Level-Index-Report-2016-Gemalto.pdf`.

Gordon, L. A. and M. P. Loeb (2002): "The economics of information security investment," *ACM Transactions on Information and System Security (TISSEC)*, 5, 438–457.

Graves, J. T., A. Acquisti, and N. Christin (2014): "Should payment card issuers reissue cards in response to a data breach," in *Proceedings of the 2014 Workshop on the Economics of Information Security*.

Greene, C. and J. Stavins (2017): "Did the Target data breach change consumer assessments of payment card security?" *Journal of Payments Strategy & Systems*, 11, 121–133.

Grossklags, J., N. Christin, and J. Chuang (2008): "Secure or insecure? A game-theoretical analysis of information security games," in *Proceedings of the 17th International World Wide Web Conference*, 209–218.

Kim, D. J., D. L. Ferrin, and H. R. Rao (2008): "A trust-based consumer decision-making model in electronic commerce: The role of trust, perceived risk, and their antecedents," *Decision Support Systems*, 44, 544 – 564.

Klein, B. and K. B. Leffler (1981): "The role of market forces in assuring contractual performance," *Journal of political Economy*, 89, 615–641.

Kunreuther, H. and G. Heal (2003): "Interdependent security," *Journal of risk and uncertainty*, 26, 231–249.

Kwon, J. and M. E. Johnson (2015): "The Market Effect of Healthcare Security: Do Patients Care about Data Breaches?" in *Proceedings of the 2015 Workshop on the Economics of Information Security*.

Mikhed, V. and M. Vogan (2015): "Out of sight, out of mind: Consumer reaction to news on data breaches and identity theft," Working Paper 15-42, Federal Reserve Bank of Philadelphia.

——— (2017): "How data breaches affect consumer credit," Working Paper 17-06, Federal Reserve Bank of Philadelphia.

Riordan, M. (2014): "Security in Partnerships," Working paper.

Roberds, W. and S. L. Schreft (2009): "Data breaches and identity theft," *Journal of Monetary Economics*, 56, 918–929.

Rogerson, W. P. (1983): "Reputation and product quality," *The Bell Journal of Economics*, 508–516.

Romanosky, S., R. Sharp, and A. Acquisti (2010): "Data Breaches and Identity Theft: When is Mandatory Disclosure Optimal?" .

Romanosky, S., R. Telang, and A. Acquisti (2011): "Do data breach disclosure laws reduce identity theft?" *Journal of Policy Analysis and Management*, 30, 256–286.

Shapiro, C. (1982): "Consumer Information, Product Quality, and Seller Reputation," *The Bell Journal of Economics*, 13, 20–35.

——— (1983): "Premiums for high quality products as returns to reputations," *The quarterly journal of economics*, 659–679.

Smallwood, D. E. and J. Conlisk (1979): "Product quality in markets where consumers are imperfectly informed," *The Quarterly Journal of Economics*, 93, 1–23.

Sullivan, R. J. (2010): "The changing nature of US card payment fraud: industry and public policy options," *Economic Review-Federal Reserve Bank of Kansas City*, 95, 101–133.

The New York Times (2010, August 21): "$9 Here, 20 Cents There and a Credit-Card Lawsuit," `http://www.nytimes.com/2010/08/22/business/22digi.html?_r=1`.

Van Goethem, T., F. Piessens, W. Joosen, and N. Nikiforakis (2014): "Clubbing seals: Exploring the ecosystem of third-party security seals," in *Proceedings of the 2014 ACM SIGSAC Conference on Computer and Communications Security*, ACM, 918–929.

Varian, H. (2004): "System reliability and free riding," in *Economics of information security*, Springer, 1–15.

Appendix A

Dominance of the Direct Effect of a Change in γ

The direct effect of a change in the bank's fraud prevention ability γ on the expected losses dominates the indirect effect arising from a change in the website's security level q whenever the following assumption is satisfied.

Assumption 3 (Convexity of Investment Cost Function)**.**
The convexity of the website's investment cost function at $q = \delta_f \lambda \rho (1-\gamma) r$ is sufficiently high:

$$c''(c'^{-1}(\delta_f \lambda \rho (1-\gamma) r)) > \frac{\delta_f \lambda \rho (1-\gamma) r}{1 - c'^{-1}(\delta_f \lambda \rho (1-\gamma) r)}.$$

Proof of Proposition 1

(i) It is clear from Lemma 1 that the consumer is playing his best-response given his belief that the website is not investing ($q_0 = q^* = 0$). We can also verify that the website is playing its best-response; i.e., it has no profitable deviation. Suppose that the website deviates to a positive level of security, $\tilde{q} \in (0, 1]$. In doing so, it incurs an investment cost of $c(\tilde{q})$. This deviation does not alter the consumer's belief ($q_0 = q^* = 0$); he continues not to purchase and the website's revenue remains at zero. Since $c(\tilde{q})$ is positive by Assumption 1, any positive deviation in security level lowers the website's profit. Therefore, it is optimal for website not to invest in equilibrium.

(ii) It can be verified that the website and the consumer are indeed playing best responses to each other in equilibrium when the consumer's valuation falls in the interval $[\underline{v}, \overline{v})$. Holding the website's strategy constant and imposing rational expectations (i.e., $q_0 = q^* = c'^{-1}(\delta \lambda \rho (1-\gamma) r)$), the consumer's expected utility from purchasing from the website in at $t = 1$ is positive if

$$v \geq \frac{(1 - c'^{-1}(\delta \lambda \rho (1-\gamma) r))(1 + \delta(1 - \lambda \rho (1-\gamma)))}{1 + \delta(1 - \lambda(1 - c'^{-1}(\delta \lambda \rho (1-\gamma) r)) \rho (1-\gamma)))} \rho (1-\gamma)(1-\lambda\alpha) l.$$

Thus, it is optimal for the consumer to participate at $t = 1$. If the consumer does not detect a breach at $t = 1$, he becomes more optimistic about the website's security level and his expected utility from participation at $t = 2$ is higher than at $t = 1$. This implies that it is optimal for the consumer to continue participating at $t = 2$ when he does not detect a breach. Finally, if the consumer detects a breach at $t = 1$, he knows that the website is vulnerable with certainty and his expected losses are $\rho(1-\gamma)(1-\lambda\alpha)l$. Since $v < \overline{v}$, the consumer is better off not using the website at $t = 2$. It is also straightforward to see that the website is playing its best response. Holding the consumer's strategy and his initial belief constant, the website's best response, $q^{BR}(q_0) = c'^{-1}(\delta \lambda \rho (1-\gamma) r) = q^*$.

Proof of Corollary 2

The equilibrium level of security is given by

$$q^* = c'^{-1}(\delta_f \lambda \rho (1-\gamma) r).$$

The partial derivatives w.r.t. λ and γ are respectively:

$$\frac{\delta q^*}{\delta \lambda}(\lambda, \gamma) = \frac{\delta_f \rho (1-\gamma) r}{c''(c'^{-1}(\delta_f \lambda \rho (1-\gamma) r))} > 0,$$

and

$$\frac{\delta q^*}{\delta \gamma}(\lambda, \gamma) = -\frac{\delta_f \lambda \rho r}{c''(c'^{-1}(\delta_f \lambda \rho (1-\gamma) r))} < 0.$$

Proof of Corollary 3

The upper bound on valuation is

$$\overline{v}(\lambda, \alpha, \gamma) = \rho(1-\gamma)(1-\lambda\alpha) l.$$

The first order partial derivatives w.r.t. λ, α and γ are respectively:

$$\frac{\partial \overline{v}}{\partial \lambda}(\lambda, \alpha, \gamma) = -\rho(1-\gamma)\alpha l < 0,$$

$$\frac{\partial \overline{v}}{\partial \alpha}(\lambda, \alpha, \gamma) = -\rho(1-\gamma)\lambda l < 0,$$

and

$$\frac{\partial \overline{v}}{\partial \gamma}(\lambda, \alpha, \gamma) = -\rho(1-\lambda\alpha) l < 0.$$

The lower bound on valuation (for the myopic consumer) is

$$\underline{v}_M(\lambda, \alpha, \gamma) = (1 - c'^{-1}(\delta_f \lambda \rho (1-\gamma) r)) \rho (1-\gamma)(1-\lambda\alpha) l.$$

The first order partial derivatives w.r.t. λ, α and γ are respectively:

$$\begin{aligned}\frac{\partial \underline{v}_M}{\partial \lambda}(\lambda, \alpha, \gamma) = &- \frac{\delta_f \rho (1-\gamma) r}{c''(c'^{-1}(\delta_f \lambda \rho (1-\gamma) r))} \rho (1-\gamma)(1-\lambda\alpha) l \\ &- (1 - c'^{-1}(\delta_f \lambda \rho (1-\gamma) r)) \rho (1-\gamma) \alpha l \\ &< 0,\end{aligned}$$

$$\frac{\partial \underline{v}_M}{\partial \alpha}(\lambda, \alpha, \gamma) = -(1 - c'^{-1}(\delta_f \lambda \rho (1-\gamma) r)) \rho (1-\gamma) \lambda l < 0,$$

and

$$\begin{aligned}\frac{\partial \underline{v}_M}{\partial \gamma}(\lambda, \alpha, \gamma) = &\frac{\delta_f \lambda \rho r}{c''(c'^{-1}(\delta_f \lambda \rho (1-\gamma) r))} \rho (1-\gamma)(1-\lambda\alpha) l \\ &- (1 - c'^{-1}(\delta_f \lambda \rho (1-\gamma) r)) \rho (1-\lambda\alpha) l.\end{aligned}$$

The first partial derivative of $\underline{v}$ w.r.t to γ is negative when Assumption 3 holds. It further be shown that $\underline{v}_M$ is convex in λ and γ.

Proof of Corollary 4

Using the results from Corollary 3, we obtain the following expressions for the first order derivatives of Δv w.r.t. to λ, α and γ respectively:

$$\frac{\partial \Delta v}{\partial \lambda}(\lambda,\alpha,\gamma) = \left((1-\lambda\alpha)\frac{\delta_f\rho(1-\gamma)r}{c''(c'^{-1}(\delta_f\lambda\rho(1-\gamma)r))} - \alpha c'^{-1}(\delta_f\lambda\rho(1-\gamma)r)\right)\rho(1-\gamma)l,$$

$$\frac{\partial \Delta v}{\partial \alpha}(\lambda,\alpha,\gamma) = -c'^{-1}(\delta_f\lambda\rho(1-\gamma)r)\rho(1-\gamma)\lambda l < 0,$$

and

$$\frac{\partial \Delta v}{\partial \gamma}(\lambda,\alpha,\gamma) = -\left((1-\gamma)\frac{\delta_f\lambda\rho r}{c''(c'^{-1}(\delta_f\lambda\rho(1-\gamma)r))} + c'^{-1}(\delta_f\lambda\rho(1-\gamma)r)\right)\rho(1-\lambda\alpha)l < 0.$$

The sign of $\frac{\partial \Delta v}{\partial \lambda}(\lambda,\alpha,\gamma)$ depends on α. For a given value of $\lambda \in [0,1]$, $\frac{\partial \Delta v}{\partial \lambda}(\lambda,\alpha,\gamma) > 0$ if

$$(1-\lambda\alpha)\delta_f\rho(1-\gamma)r - \alpha c'^{-1}(\delta_f\lambda\rho(1-\gamma)r)c''(c'^{-1}(\delta_f\lambda\rho(1-\gamma)r)) > 0$$

or equivalently,

$$\alpha < \frac{\delta_f\rho(1-\gamma)r}{\delta_f\lambda\rho(1-\gamma)r + c'^{-1}(\delta_f\lambda\rho(1-\gamma)r)c''(c'^{-1}(\delta_f\lambda\rho(1-\gamma)r))} = \hat{\alpha}(\lambda).$$

Replacing $c'^{-1}(\delta_f\lambda\rho(1-\gamma)r)$ by $q^*(\lambda,\gamma)$ in the expression of $\hat{\alpha}$ and differentiating w.r.t. λ, we obtain

$$\frac{\partial \hat{\alpha}(\lambda)}{\partial \lambda} = -\frac{\delta_f\rho(1-\gamma)r}{\left(\delta_f\lambda\rho(1-\gamma)r + q^*(\lambda,\gamma)c''(q^*(\lambda,\gamma))\right)^2}\left(\delta_f\rho(1-\gamma)r + \frac{\partial q^*}{\partial \lambda}(\lambda,\gamma)c''(q^*(\lambda,\gamma)) + q^*(\lambda,\gamma)\frac{\partial q^*}{\partial \lambda}(\lambda,\gamma)c'''(q^*(\lambda,\gamma))\right).$$

Using the result from Lemma 2, we can establish that the above derivative is negative. This implies that $\alpha \leq \hat{\alpha}(\lambda)$ for all λ if

$$\alpha \leq \hat{\alpha}(1) = \frac{\delta_f\rho(1-\gamma)r}{\delta_f\rho(1-\gamma)r + c'^{-1}(\delta_f\rho(1-\gamma)r)c''(c'^{-1}(\delta_f\rho(1-\gamma)r))} = \overline{\alpha}.$$

Consider now the case where $\alpha > \overline{\alpha}$, $\Delta v(\lambda,\alpha,\gamma)$. Replacing $c'^{-1}(\delta_f\lambda\rho(1-\gamma)r)$ by $q^*(\lambda,\gamma)$ in the first order partial derivative of $\Delta v(\lambda,\alpha,\gamma)$ w.r.t. λ, we have

$$\frac{\partial \Delta v}{\partial \lambda}(\lambda,\alpha,\gamma) = \left((1-\lambda\alpha)\frac{\partial q^*}{\partial \lambda}(\lambda,\gamma) - \alpha q^*(\lambda,\gamma)\right)\rho(1-\gamma)l.$$

The second order partial derivative of $\Delta v(\lambda,\alpha,\gamma)$ w.r.t. λ is

$$\frac{\partial^2 \Delta v}{\partial \lambda^2}(\lambda,\alpha,\gamma) = \left(-2\alpha\frac{\partial q^*}{\partial \lambda}(\lambda,\gamma) + (1-\lambda\alpha)\frac{\partial^2 q^*}{\partial \lambda^2}(\lambda,\gamma)\right)\rho(1-\gamma)l,$$

where

$$\frac{\partial^2 q^*}{\partial \lambda^2}(\lambda,\gamma) = -(\frac{\partial q^*}{\partial \lambda}(\lambda,\gamma))^2 \frac{c'''(q^*)}{c''(q^*)} < 0.$$

This implies that $\frac{\partial^2 \Delta v}{\partial \lambda^2}(\lambda,\alpha,\gamma) < 0$ and the function $\Delta v(\lambda,\alpha,\gamma)$ is concave in λ.

In addition, we have that

$$\left.\frac{\partial \Delta v}{\partial \lambda}\right|_{\lambda=0} = \frac{\delta_f \rho(1-\gamma)r}{c''(0)}\rho(1-\gamma)l > 0,$$

and

$$\left.\frac{\partial \Delta v}{\partial \lambda}\right|_{\lambda=1} < 0,$$

since $\alpha > \hat{\alpha}(1)$.

Therefore, there exists $\overline{\lambda}$ such that $\Delta v(\lambda,\alpha,\gamma)$ is increasing in λ for all $\lambda \in [0,\overline{\lambda}]$ and is decreasing otherwise. The threshold $\overline{\lambda}$ solves

$$\frac{\partial \Delta v}{\partial \lambda}(\lambda,\alpha,\gamma) = 0$$

or equivalently,

$$(1-\overline{\lambda}\alpha)\delta_f \rho(1-\gamma)r - \alpha c'^{-1}(\delta_f \overline{\lambda}\rho(1-\gamma)r)c''(c'^{-1}(\delta_f \overline{\lambda}\rho(1-\gamma)r)) = 0.$$

Proof of Proposition 2

(i) In the unregulated market, the website's equilibrium security level is given by

$$q^* = \begin{cases} 0 & \text{if } v < \underline{v}_M \text{ \& if } v \geq \overline{v} \\ c'^{-1}(\delta_f \lambda \rho(1-\gamma)r) & \text{if } v \in [\underline{v}_M, \overline{v}). \end{cases}$$

Under the breach notification law, its equilibrium security level is given by

$$q^*_{BN} = \begin{cases} 0 & \text{if } v < \underline{v}_{M,BN} \text{ \& if } v \geq \overline{v}_{BN} \\ c'^{-1}(\delta_f \rho(1-\gamma)r) & \text{if } v \in [\underline{v}_{M,BN}, \overline{v}_{BN}). \end{cases}$$

It is clear that $\underline{v}_{M,BN} < \underline{v}_M$ and $\overline{v}_{M,BN} < \overline{v}_M$. Further, it can be verified that $\overline{v}_{M,BN} \geq \underline{v}$ when $\alpha < \frac{c'^{-1}(\delta_f \lambda \rho(1-\gamma)r)}{1-\lambda(1-c'^{-1}(\delta_f \lambda \rho(1-\gamma)r))} = \tilde{\alpha}$ and $\overline{v}_{M,BN} < \underline{v}$ otherwise. Consider the case where $\alpha \leq \tilde{\alpha}$, the difference in the equilibrium security level with and without notification is

$$q^*_{BN} - q^* = \begin{cases} 0 & \text{if } v < \underline{v}_{M,BN} \\ c'^{-1}(\delta_f \rho(1-\gamma)r) & \text{if } v \in [\underline{v}_{M,BN}, \underline{v}_M) \\ c'^{-1}(\delta_f \rho(1-\gamma)r) - c'^{-1}(\delta_f \lambda \rho(1-\gamma)r) & \text{if } v \in [\underline{v}_M, \overline{v}_{BN}) \\ -c'^{-1}(\delta_f \lambda \rho(1-\gamma)r) & \text{if } v \in [\overline{v}_{BN}, \overline{v}) \\ 0 & \text{if } v \geq \overline{v}. \end{cases} \quad (5)$$

When $\alpha > \tilde{\alpha}$, the difference in the security levels is

$$q^*_{BN} - q^* = \begin{cases} 0 & \text{if } v < \underline{v}_{M,BN} \\ c'^{-1}(\delta_f \rho(1-\gamma)r) & \text{if } v \in [\underline{v}_{M,BN}, \overline{v}_{BN}) \\ 0 & \text{if } v \in [\overline{v}_{BN}, \underline{v}_M) \\ -c'^{-1}(\delta_f \lambda \rho(1-\gamma)r) & \text{if } v \in [\underline{v}_M, \overline{v}) \\ 0 & \text{if } v \geq \overline{v}. \end{cases} \tag{6}$$

It can be seen from (5) and (6) that for all $\alpha \in [0,1]$, the optimal security level under breach notification q^*_{BN} is weakly higher than q^* when $v < \underline{v}_{M,BN}$ and vice-versa. It can also be seen that $q^*_{BN} > q^*$ when $v \in [\underline{v}_{M,BN}, \overline{v}_{BN})$ in both scenarios. Finally, $q^*_{BN} < q^*$ for $v \in [\overline{v}_{BN}, \overline{v})$ when $\alpha < \tilde{\alpha}$ and for $v \in [\underline{v}_M, \overline{v})$ otherwise. Since $\overline{v}_{BN} > \underline{v}_M$ when $\alpha < \tilde{\alpha}$ and vice-versa, $q^*_{BN} < q^*$ when $v \in [max\{\overline{v}_{BN}, \underline{v}_M\}, \overline{v})$ for all α.

(ii) The range of valuation over which the website invests strictly less is given by

$$\tilde{\Delta}v = \begin{array}{ll} \overline{v} - \overline{v}_{BN} & \text{if } \alpha < \tilde{\alpha} \\ \overline{v} - \underline{v}_M & \text{if } \alpha \geq \tilde{\alpha}. \end{array}$$

The ratio of $\tilde{\Delta}v$ to Δv is strictly less than 1 when $\alpha < \tilde{\alpha}$ since $\tilde{\Delta}v < \Delta v$; i.e., there is "partial" crowding out. The ratio is 1 when $\alpha \geq \tilde{\alpha}$ since $\tilde{\Delta}v = \Delta v$; i.e., there is full crowding out.
Finally, it can be verified that the first order derivative of $\tilde{\alpha}$ w.r.t. λ is positive:

$$\frac{\partial \tilde{\alpha}}{\partial \lambda} = \frac{(1-\lambda)\frac{\delta_f \rho(1-\gamma)r)}{c''(c'^{-1}(\delta_f \lambda \rho(1-\gamma)r))}}{(1-\lambda(1-c'^{-1}(\delta_f \lambda \rho(1-\gamma)r)))^2} > 0.$$

(iii) In the region of crowding out, the consumer surplus at equilibrium in the absence and the presence of breach notification are respectively

$$CS(q^*, q^*) = q^*(1+\delta)v + (1-q^*)(1+\delta(1-\lambda\rho(1-\gamma))(v - \rho(1-\gamma)(1-\lambda\alpha)l);$$

where $q^* = c'^{-1}(\delta\lambda\rho(1-\gamma)r)$.

$$CS_{BN}(0,0) = (1+\delta)(v - \rho(1-\gamma)(1-\alpha)l).$$

The difference in consumer surplus is

$$\begin{aligned} \Delta CS_{BN} = & \underbrace{-q^*(1+\delta)\rho(1-\gamma)(1-\alpha)l}_{A} + \underbrace{(1-q^*)\delta\lambda\rho(1-\gamma)(v-\overline{v})}_{B} \\ & + \underbrace{(1-q^*)(1+\delta)\rho(1-\gamma)(1-\lambda)\alpha l}_{C}, \end{aligned}$$

where $A < 0$, $B < 0$, $C > 0$ and $B + C > 0$. A sufficient condition for ΔCS_{BN} to be negative is $A + C \leq 0$. This is the case when

$$\alpha \leq \frac{q^*}{1-\lambda(1-q^*)}.$$

Further, ΔCS_{BN} is increasing in α:

$$\begin{aligned}\frac{\partial \Delta CS_{BN}}{\partial \alpha} &= \frac{\partial CS_{BN}}{\partial \alpha} - \frac{\partial CS}{\partial \alpha} \\ &= (1+\delta)\rho(1-\gamma)l - (1-q^*)(1+\delta(1-\lambda\rho(1-\gamma))\lambda\rho(1-\gamma)l > 0,\end{aligned}$$

and

$$\Delta CS_{BN}|_{\alpha=1} = (1-q^*)(\delta\lambda\rho(1-\gamma)v + (1+\delta(1-\lambda\rho(1-\gamma))\overline{v} > 0.$$

Since ΔCS_{BN} is continuous in α, there must exist a threshold α' such that $\Delta CS_{BN} \leq 0$ for all $\alpha \leq \alpha'$ and $\Delta CS_{BN} > 0$ otherwise.

Proof of Proposition 4

Parts (i) and (ii) of the proposition are straightforward to show and their proofs are therefore omitted.

(iii) First, note that the level of investment at the positive equilibrium is unchanged by the policy. Consider the interval $v \in [0, \overline{v})$. A consumer with valuation in this interval is unaffected by the policy since the level of security is unchanged and he never purchases following a breach (and hence never incurs the inconvenience cost). It is also easy to see that a consumer with valuation $v > \overline{v} + \tau$ will be made worse off. The website continues not to invest in the presence of the policy—hence, the consumer's expected fraud losses remains the same—but the consumer now incurs an inconvenience cost when he purchases from the website following a data breach. Let us now examine the interval $v \in [\overline{v}, \overline{v} + \tau)$. In this region, the consumer's purchasing decision is affected by the policy; specifically, he no longer finds it optimal to purchase following the detection of the breach. The consumer surplus in the absence and in the presence of the policy are respectively

$$CS(0,0) = (1+\delta)(v - \overline{v}),$$

and

$$\begin{aligned}CS_E(q^*, q^*) =& (1+\delta(1-\lambda(1-q^*)\rho(1-\gamma)))v \\ &- (1-q^*)(1+\delta(1-\lambda\rho(1-\gamma)))\overline{v},\end{aligned}$$

where $q^* = c'^{-1}(\delta\lambda(1-\gamma)\rho r)$.
The difference in consumer surplus is

$$\Delta CS_E = -\delta\lambda(1-q^*)\rho(1-\gamma)v + (q*(1+\delta) + (1-q^*)\delta\lambda\rho(1-\gamma))\overline{v},$$

which is strictly positive if

$$v < min\left\{\left(1 + \frac{q*(1+\delta)}{(1-q^*)\delta\lambda\rho(1-\gamma)} \quad \overline{v}, \overline{v}+\tau\right\} = \hat{v}.$$

Therefore, consumer surplus is strictly higher when $v \in [\overline{v}, \hat{v})$, and weakly higher for all $v \in [0, \hat{v})$.

Proof of Proposition 5

The equilibrium levels of system security with and without the liability rule corresponding to case (ii) of Lemma 10 are respectively

$$\mu^*_{S,L} = 1 - (1-\gamma_0)(1 - q^{BR}_L(\gamma_0))$$

and

$$\mu^*_S = 1 - (1-\gamma_n)(1 - q^{BR}(\gamma_n)).$$

The change in equilibrium level of system security is given by

$$\begin{aligned}\mu^*_{S,L} - \mu^*_S &= -(1-\gamma_0)(1 - q^{BR}_L(\gamma_0)) + (1-\gamma_n)(1 - q^{BR}(\gamma_n)) \\ &= \underbrace{(1-\gamma_n)(\Delta\tilde{q}^{BR}(\gamma_n))}_{\substack{\text{Reduction in system} \\ \text{vulnerability level} \\ \text{if } \gamma \text{ is unchanged}}} - \underbrace{\Delta\gamma - (1-\gamma_0)q^{BR}_L(\gamma_0) + (1-\gamma_n)q^{BR}_L(\gamma_n)}_{\text{Change in system vulnerability due to change in } \gamma}\end{aligned}$$

The change in system security due to the change in γ can be re-expressed as follows:

$$\Delta\gamma - (1-\gamma_0)q^{BR}_L(\gamma_0) + (1-\gamma_n)q^{BR}_L(\gamma_n) = \Delta\gamma(1 - q^{BR}_L(\gamma_0)) - (1-\gamma_n)\Delta q^{BR}_L.$$

Therefore, the level of system security is unambiguously higher whenever

$$\Delta\gamma(1 - q^{BR}_L(\gamma_0)) + (1-\gamma_n)\Delta q^{BR}_L \leq 0. \tag{7}$$

8 Appendix B

Extension: Non-competing firms

Denote $q_{i,0}$ and $q_{j,0}$ the consumer's initial beliefs about the state of security at firms i and j respectively. Suppose that the consumer's valuations for the both firms' products are the same (i.e., $v_i = v_j = v$) and that $v < \rho l$ such that it is not optimal for the consumer to buy from a vulnerable firm. Consider now the consumer's problem. At $t = 1$, the consumer's expected utility when he purchases from firm i given his initial beliefs is

$$E(U_{i,1}(q_{i,0})) = v - (1 - q_{i,0})\rho l.$$

A consumer with valuation v finds it optimal to purchase if

$$q_{i,0} \geq 1 - \frac{v}{\rho l} = \hat{q}_{i,0}.$$

At $t = 2$, the consumer's consumer's beliefs about firm i's state of security are given as follows:

$$q_{i,1} = \begin{cases} q_{0,1} & \text{if he did not purchase} \\ \dfrac{q_{i,0}(1-q_{j,0})}{(1-q_{i,0}) + (1-q_{j,0}) - (1-q_{i,0})(1-q_{j,0})\rho} & \text{if a breach was detected} \\ \dfrac{q_{i,0}(1-\lambda(1-q_{j,0})\rho)}{1-\lambda((1-q_{i,0}) + (1-q_{j,0}) - (1-q_{i,0})(1-q_{j,0})\rho)\rho} & \text{if no breach was detected.} \end{cases}$$

As in the baseline set-up, the consumer's updated belief is higher when no breach is detected. Therefore, he continues to buy from the firm at $t = 2$. In the case where a breach was detected, the consumer's expected utility when he purchases at $t = 2$ is

$$E(U_{i,2}) = v - \frac{(1 - q_{i,0})(1 + (1 - q_{j,0})(1 - \rho))}{(1 - q_{i,0}) + (1 - q_{j,0}) - (1 - q_{i,0})(1 - q_{j,0})\rho} \rho l.$$

The consumer finds it optimal to purchase if

$$q_{i,0} \geq \frac{(1 + (1 - q_{j,0})(1 - \rho))(v - \rho l)}{(1 - (1 - q_{j,0})\rho)(v - \rho l) - (1 - q_{j,0})\rho l} = \hat{\hat{q}}_{i,0}(q_{j,0}).$$

The threshold $\hat{\hat{q}}_{i,0}$ is (strictly) smaller than 1 for all $q_{j,0} < 1$. This implies that while it is never optimal for a consumer to purchase following breach detection when there is a single firm, it may be optimal for him to do so when there are multiple firms. This is due to the imperfect attribution of data breaches in the presence of multiple firms. The reputation damage that a firm suffers when it is breached is smaller as the blame is shared by the other firm.

Let us examine firm i's investment problem. Taking the consumer's beliefs, $q_{i,0}$ and $q_{j,0}$, and firm j's security level, q_j, as given, firm i's profit is

$$\pi_i(q_i) \equiv R_i(q_i) - c(q_i),$$

where its revenue function corresponds to

$$R_i(q_i) = \begin{cases} 0 & \text{if } q_{i,0} < \hat{q}_{i,0} \\ (1 + \delta_f - \delta_f \lambda (1 - q_i)\rho) r & \text{if } q_{i,0} \in [\hat{q}_{i,0}, \hat{\hat{q}}_{i,0}) \\ & \quad \& \; q_{j,0} < \hat{q}_{j,0} \\ (1 + \delta_f - \delta_f \lambda ((1 - q_i) + (1 - q_j) - (1 - q_i)(1 - q_j)\rho)\rho) r & \text{if } q_{i,0} \in [\hat{q}_{i,0}, \hat{\hat{q}}_{i,0}) \\ & \quad q_{j,0} \in [\hat{q}_{j,0}, \hat{\hat{q}}_{j,0}) \\ (1 + \delta_f) r & \text{otherwise.} \end{cases}$$

Its best-response function to firm j's choice of security level for given consumer beliefs is

$$q_i^{BR}(q_j) = \begin{cases} c'^{-1}(\delta_f \lambda \rho r) & \text{if } q_{i,0} \in [\hat{q}_{i,0}, \hat{\hat{q}}_{i,0}) \;\&\; q_{j,0} < \hat{q}_{j,0} \\ c'^{-1}(\delta_f \lambda \rho (1 - (1 - q_j)\rho) r) & \text{if } q_{i,0} \in [\hat{q}_{i,0}, \hat{\hat{q}}_{i,0}) \;\&\; q_{j,0} \in [\hat{q}_{j,0}, \hat{\hat{q}}_{j,0}) \\ 0 & \text{otherwise.} \end{cases}$$

The case of interest here is the one where both firms are active in the market and where turnover occurs following breach detection (i.e., when $q_{i,0} \in [\hat{q}_{i,0}, \hat{\hat{q}}_{i,0})$ and $q_{j,0} \in [\hat{q}_{j,0}, \hat{\hat{q}}_{j,0})$). It can be shown that q_i^{BR} is increasing and concave in q_j when this is the case:

$$\frac{\partial q_i^{BR}}{\partial q_j}(q_j) = \frac{\delta_f \lambda \rho^2 r}{c''(c'^{-1}(\delta_f \lambda (1 - (1 - q_j)\rho)\rho r)} > 0,$$

and

$$\frac{\partial^2 q_i^{BR}}{\partial q_j^2}(q_j) = - \left(\frac{\delta_f \lambda \rho^2 r}{c''(c'^{-1}(\delta_f \lambda (1 - (1 - q_j)\rho)\rho r)} \right)^2 \frac{c'''(c'^{-1}(\delta_f \lambda (1 - (1 - q_j)\rho)\rho r))}{c''(c'^{-1}(\delta_f \lambda (1 - (1 - q_j)\rho)\rho r))} < 0.$$

Further, a unique solution exists whenever the investment cost function is sufficiently convex at $q =' c^{-1}(\delta\lambda\rho r)$; more precisely if

$$c''(c'^{-1}(\delta_f\lambda\rho r)) > \delta_f\lambda\rho^2 r.$$

The optimal security level (for given beliefs) solves

$$q^*_{NC} = c'^{-1}(\delta_f\lambda(1-(1-q^*_{NC})\rho)r).$$

Since $q^*_{NC} \in (0,1)$, we have the following relation:

$$c'^{-1}(\delta_f\lambda(1-\rho)\rho r) < q^*_{NC} < c'^{-1}(\delta_f\lambda\rho r).$$

There exists an equilibrium (with rational expectations) to the full game if the set of valuations for which the consumer is willing to purchase at $t=1$ but not at $t=2$ after detecting a breach is non-empty. The consumer hold consistent beliefs at equilibrium (i.e., $q_{i,0} = q_{j,0} = q^*_{NC}$). The consumer is willing to purchase from firm i at $t=1$ given his equilibrium beliefs if

$$q^*_{NC} \geq \hat{q}_{i,0}.$$

A sufficient condition for the above expression to hold is

$$c'^{-1}(\delta_f\lambda(1-\rho)\rho r) \geq \hat{q}_{i,0},$$

which gives

$$v \geq (1 - c'^{-1}(\delta_f\lambda(1-\rho)\rho r))\rho l = \underline{v}^{NC}.$$

The consumer is not willing to purchase from firm i following the detection of a breach if

$$q^*_{NC} < \hat{\hat{q}}_{i,0}(q^*_{NC}),$$

which gives us

$$v < \quad \frac{1+(1-q^*_{NC})(1-\rho)}{1+(1-q^*_{NC})(1-\rho)+q^*_{NC}} \quad \rho l.$$

The right hand side of the above inequality is decreasing in the valuation of q^*_{NC}. This implies that

$$\frac{1+(1-c'^{-1}(\delta_f\lambda\rho r))(1-\rho)}{1+(1-c'^{-1}(\delta_f\lambda\rho r))(1-\rho)+c'^{-1}(\delta_f\lambda\rho r)} < \frac{1+(1-q^*_{NC})(1-\rho)}{1+(1-q^*_{NC})(1-\rho)+q^*_{NC}}$$

and a sufficient condition for the consumer not to be willing to purchase after detecting a breach is thus

$$v < \quad \frac{1+(1-c'^{-1}(\delta_f\lambda\rho r))(1-\rho)}{1+(1-c'^{-1}(\delta_f\lambda\rho r))(1-\rho)+c'^{-1}(\delta_f\lambda\rho r)} \quad \rho l = \overline{v}^{NC}.$$

Observe that $\overline{v}^{NC} < \rho l$; i.e., the consumer is willing to purchase following a breach detection at a lower valuation compared to the case where there is a single firm. This is the result of imperfect attribution as mentioned earlier. It remains to be

verified that there exists parameter values such that the interval $[\underline{v}^{NC}, \overline{v}^{NC})$ is indeed non-empty. We have that

$$\overline{v}^{NC} - \underline{v}^{NC} = \frac{2 - c'^{-1}(\delta_f \lambda(1-\rho)\rho r) - 1 + (1 - c'^{-1}(\delta_f \lambda \rho r))(1 - c'^{-1}(\delta_f \lambda(1-\rho)\rho r)\rho)}{1 + (1 - c'^{-1}(\delta_f \lambda \rho r))(1-\rho) + c'^{-1}(\delta_f \lambda \rho r)},$$

which is (strictly) positive, for instance, when r is sufficiently large

$$r > \frac{c'\left(\frac{1}{2}\right)}{\delta_f \lambda (1-\rho)\rho}.$$

Therefore, we can conclude that there exists an equilibrium in which the firms invests in data security when $v \in [\underline{v}^{NC}, \overline{v}^{NC})$ (for some sets of parameter values).