

# Consumer Inattention and Bill-Shock Regulation\*

Michael D. Grubb<sup>†</sup>

May 1, 2012

## Abstract

For many goods and services, such as cellular-phone service and debit-card transactions, the price of the next unit of service depends on past usage. As a result, consumers who are inattentive to their past usage but are aware of contract terms may remain uncertain about the price of the next unit. I develop a model of inattentive consumption, derive equilibrium pricing when consumers are inattentive, and evaluate bill-shock regulation requiring firms to disclose information that substitutes for attention. When inattentive consumers are heterogeneous and unbiased, bill-shock regulation reduces social welfare in fairly-competitive markets, which may be the effect of the FCC's recent bill-shock agreement. If inattentive consumers underestimate their demand, however, then bill-shock regulation can lower market prices and protect consumers from exploitation. Hence the Federal Reserve's new opt-in rule for debit-card overdraft protection may substantially benefit consumers.

---

\*A previous version of this paper circulated under the title "Bill Shock: Inattention and Price-Posting Regulation". I thank Heski Bar-Isaac, Alessandro Bonatti, Glenn Ellison, Bob Gibbons, Ginger Jin, Bob Pindyck, and Tavneet Suri for careful reading and many helpful comments and suggestions.

<sup>†</sup>MIT Sloan School of Management. [mgrubb@mit.edu](mailto:mgrubb@mit.edu), [www.mit.edu/~mgrubb](http://www.mit.edu/~mgrubb).

# 1 Introduction

For a wide variety of services, including electricity, cellular-phone service, and debit and credit-card transactions, marginal prices increase sharply when consumers exceed specified usage thresholds. Consumers commonly cross such usage thresholds and accrue high fees without realizing it, resulting in *bill shock*, because they are inattentive and do not keep track of past usage.<sup>1</sup> For example, a cellular-phone user may not realize that the current call is charged a penalty (or *overage*) rate of 45 cents per minute, because he does not know that he has already used up his 500 included minutes. Similarly, a checking account holder may be unaware that her next debit transaction will incur a \$35 overdraft penalty because she does not realize her checking balance is negative.

In each example, firms have the ability to disclose whether a penalty fee is applicable at the point of sale. (Absent such disclosure I refer to the penalties as *surprise* penalty fees.) A mobile phone screen could flash “overage rate applies” before a call is made and a debit-card terminal could ask “Overdraft fee applies. Continue - Yes/No?” before processing transactions on an overdrawn account. That firms do not to make these disclosures and oppose<sup>2</sup> regulation requiring such disclosure suggests that firms benefit from bill shock. In contrast, consumer groups and regulators such as the FCC and the Fed believe that the lack of transparency is bad for consumers and bad for welfare,<sup>3</sup> which has led to new regulation. For instance, in late 2011, US President Barack Obama said that,

Far too many Americans know what it’s like to open up their cell-phone bill and be shocked by hundreds or even thousands of dollars in unexpected fees and charges. But we can put an end to that with a simple step: an alert warning consumers that they’re about to hit their limit before fees and charges add up.

Obama’s statement was made at the announcement of an agreement between cellular carriers and the FCC to begin providing such usage alerts by April 2013 (CTIA - The Wireless Association 2011). The Fed has been similarly concerned about overdraft fees on ATM and one-time debit-card

---

<sup>1</sup>Bill shock also arises without inattention when multiple family members consume from the same family-talk plan or joint checking-account but do not continually update each other about purchases. Shared usage will be an alternative interpretation to inattention throughout the paper.

<sup>2</sup>Prior to their recent bill-shock agreement with the FCC, the wireless industry trade group, C.T.I.A. - The Wireless Association, argued that the FCC’s proposed bill-shock regulation “violates carriers’ First Amendment protections. . . . against government compelled speech” (Altschul, Guttman-McCabe and Josef 2011). Similarly, banks opposed the Fed’s opt-in rule for overdraft protection (Federal Reserve Board 2009b).

<sup>3</sup>FCC Chairman Julius Genachowski said, “something is clearly wrong with a system that makes it possible for consumers to run up big bills without knowing it,” and a variety of consumer advocacy groups agree (Genachowski 2010, Deloney, Sherry, Grant, Desai, Riley, Wood, Breyault, Gonzalez and Lennett 2011).

transactions, fees which totalled \$20 billion in 2009 (Martin 2010). In response, since 2010, Fed rules prohibit such overdraft fees unless a consumer first opts in to overdraft service. Moreover, the Consumer Financial Protection Bureau (CFPB) is currently considering additional overdraft fee regulation, and one option the CFPB could consider is to require banks to issue zero-balance alerts.

Obama describes the new FCC bill-shock agreement as part of his Administration’s “ongoing efforts to protect American consumers by making sure financial transactions are fair, honest and transparent” (CTIA - The Wireless Association 2011). Holding pricing fixed, usage alerts should at least weakly benefit consumers, as Obama assumes, by giving them more information to make better choices. Presumably this is why bill-shock regulation has strong support from consumer groups and regulators. Nevertheless, firms will change prices in response to new disclosure requirements and therefore the net effect need not be beneficial.

This paper’s goal is to determine whether bill-shock regulation requiring firms to disclose information about transaction prices at the point of sale will benefit consumers or raise total welfare. To understand the effect of bill-shock regulation, however, I also answer a related question: Why do firms both charge *penalty fees* (so that high usage triggers high marginal charges) and make them a *surprise* by not alerting consumers when they cross the relevant threshold? I show that the answers to both questions depend on factors including (1) consumer heterogeneity, (2) consumer bias, (3) market power, and (4) firms’ ability to lower fixed fees.

I will assume that: (1) Once a consumer signs a cellular-phone contract or opens a bank account, two consumption opportunities arise sequentially. (2) All consumers are inattentive so that each decision to make an additional phone call or debit-card transaction is made without any recollection of prior usage.<sup>4</sup> (3) Consumers are aware of their own inattention when forecasting their own future consumption choices. Given these three assumptions, I show that, for any price schedule, an inattentive consumer’s optimal strategy is a threshold rule: she buys only those units valued above her expected marginal price.<sup>5</sup> In this setting, I analyze a disclosure requirement that is sufficient to make inattentive consumers attentive and, in the context of the model, is equivalent to requiring that firms alert consumers when penalty fees are applicable.

---

<sup>4</sup>Gabaix and Laibson (2006), Bubb and Kaufman (2011), and Armstrong and Vickers (2012) focus on the cross-subsidization of sophisticated consumers by unsophisticated consumers. By assuming that there are no attentive consumers, I abstract from the similar cross-subsidization of attentive consumers by inattentive consumers that would occur if both types were present.

<sup>5</sup>This provides a micro-foundation for the threshold labor supply rule used by Saez (2002) and the consumption rules used by Borenstein (2009) and Grubb and Osborne (2012). These papers use the threshold rules in demand or labor supply estimation, while I explore the supply-side ramifications of such behavior.

The benchmark model assumes that, while they may have heterogeneous brand preferences, consumers have the same demand (so there is no scope for price discrimination) and have correct beliefs (so there are no biases to exploit). The benchmark result is that firms have no incentive to charge surprise penalty fees and regulation will not affect firm profits or consumer surplus. This result is straightforward because, in the benchmark environment, a firm cannot do better than induce first-best consumption via marginal-cost pricing.

Moving beyond this initial benchmark, I show that firms can profit by charging surprise penalty fees when price discriminating between low- and high-demand consumers who are unbiased or when exploiting biased beliefs of consumers who underestimate their own demand. This leads to two main results: (1) Bill-shock regulation can lower social welfare and harm low-demand consumers when firms use surprise penalty fees to price discriminate between unbiased consumers by inducing them to choose different contracts. The result always holds in *fairly-competitive* markets (those that are highly but not perfectly competitive) because surprise penalty fees mitigate allocative distortions otherwise inherent in second-degree price discrimination. (2) When consumers underestimate demand, bill-shock regulation may help or hurt consumers if firms compete on positive fixed fees. Bill-shock regulation will benefit consumers, however, if bias is sufficiently severe and fixed fees have already been competed to zero (so that oligopolists profit only from penalty fees and other marginal charges). In the latter case, regulation stiffens competition by limiting penalty-fee revenues and thereby forcing firms to raise and compete on fees to which consumers are more price-sensitive.

These results suggest that the CFPB should require banks to issue low-balance alerts. Banks price discriminate but typically do not vary overdraft fees across checking accounts to do so, as would be required if the fees were used to influence consumers' account choices. Thus the first main result does not apply to overdraft fees. However, there is growing evidence that consumers underestimate future spending and borrowing (e.g. Ausubel (1991), Skiba and Tobacman (2008), and DellaVigna (2009)). Moreover, *free checking* is common, being the industry norm prior to recent Fed regulation. Thus the second main result implies low-balance alerts would benefit checking account holders. Finally, if consumers are aware of their own inattention, then the Fed's new opt-in rule should already be providing some of this benefit.

Unfortunately, an assessment of the FCC's recent bill-shock agreement is less clear cut. In contrast to checking fees, cellular companies' overage charges are clearly used to help sort consumers into low and high included-minute contracts and their monthly fees are positive. One cannot conclude from the first main result that the FCC's bill-shock agreement will reduce social welfare because the result assumed a fairly-competitive market and unbiased consumers, whereas in fact cellular carriers do have substantial market power (Justice 2011) and evidence shows that their

consumers are biased (Grubb 2009, Grubb and Osborne 2012). Nevertheless, insights from both main results are relevant, indicating that the FCC’s recent bill-shock regulation may not live up to expectations and could lower welfare and harm some consumers.

Section 4 develops the first main result. It enriches the benchmark model by incorporating two ex ante types, with low and high expectations of future demand. Given such heterogeneity, the surprising result is that the combination of surprise penalty fees and consumer inattention can be socially valuable (as well as privately valuable to firms) and benefit low-demand consumers. Surprise penalty fees can be socially valuable because they can reduce allocative distortions imposed by price discriminating firms. Thus bill-shock regulation can lower social welfare and harm some consumers. Moreover, this is always the case in fairly-competitive markets. The intuition is that, when consumers are inattentive, both surprise penalty fees and quantity distortions are useful tools for price discrimination because both relax incentive constraints. By substituting for attention and eliminating surprise, bill-shock regulation removes surprise penalty fees from the price-discrimination toolbox. This limits firms’ ability to price discriminate, explaining their aversion to the regulation. Moreover, if surprise penalty fees and quantity distortions are substitutes, regulation leads to an increase in quantity distortions and reduces social welfare. In contrast, if the two tools are complements, regulation has the opposite effect. While, in general, surprise penalty fees and quantity distortions could be either substitutes or complements, they are always substitutes in fairly-competitive markets. This follows because (1) competition limits the additional markup firms would like to charge high-demand consumers; and (2) surprise penalty fees render quantity distortion unnecessary if differences in markups are small.

Section 5 develops the second main result. It enriches the benchmark model in a second direction by assuming that consumers underestimate their own future demand and hence the chance of paying penalties (while remaining aware of their own inattention). In this case firms use surprise penalty fees to exploit consumers’ bias. The degree to which firms can profit from consumer bias is limited by incentive constraints which are tightened when there are bill-shock alerts that help consumers avoid paying penalties. Under monopoly this means that bill-shock regulation serves its intended role of consumer protection by limiting monopoly rents. In an oligopoly, however, if firms compete on positive fixed fees then bill-shock regulation will not effect equilibrium markups and consumers are residual claimants of social surplus. This follows because revenues from penalty fees are always rebated through lower fixed fees, a result which matches claims made by critics of bill-shock regulation (Federal Reserve Board 2009b).<sup>6</sup> Moreover, whether the regulation will increase

---

<sup>6</sup>Jamie Dimon, CEO of JPMorgan Chase said, “If you’re a restaurant and you can’t charge for the soda, you’re going to charge more for the burger. Over time, it will all be repriced into the business” (Dash and Schwartz 2010).

or decrease total welfare (and hence help or hurt consumers) will vary with the level of marginal costs and the nature of consumer bias. However, if fixed fees have already been competed down to zero (so that oligopolists profit only from penalty fees and other marginal charges, as in the case of free checking) then an extension in Section 6 shows that bill-shock regulation will benefit consumers if bias is sufficiently severe. By limiting penalty fee revenues, bill-shock regulation forces firms to raise and compete on more salient fees, thereby intensifying competition and lowering equilibrium markups.

The paper proceeds as follows. Section 2 discusses related literature. Section 3 introduces the benchmark model, derives an inattentive consumer’s consumption rule, and shows the benchmark result. Section 4 analyzes the model enriched with ex ante heterogeneity, in which firms price discriminate by offering multiple contracts and inattention may reduce typical distortions. Section 5 makes the alternative extension to biased consumer beliefs, for which inattention can increase the scope for exploitation. Section 6 reconsiders biased beliefs assuming nonnegative pricing. Section 7 discusses policy implications for the FCC’s bill-shock agreement and the Fed’s overdraft opt-in regulation, and Section 8 concludes. Proofs are in the online appendix ([www.mit.edu/~mgrubb](http://www.mit.edu/~mgrubb)).

## 2 Related Literature

Standard models of consumer choice from multi-part tariffs are static and assume that individuals make a single quantity choice, tailored to the ex post marginal price relevant at the chosen quantity. This implicitly assumes perfect consumer foresight and is empirically rejected by the lack of bunching at tariff kink points in electricity (Borenstein 2009) and cellular-phone-service (Grubb and Osborne 2012) consumption. Relaxing the perfect foresight assumption, attentive consumers will reduce consumption following periods of high usage that increase the chance of triggering penalty fees. Using call-level data, Grubb and Osborne (2012) find no evidence of this behavior among US cellular-phone users, suggesting that they are inattentive to their own past usage. (In contrast, Yao, Mela, Chiang and Chen (2011) show evidence that Chinese phone consumers are attentive, which may be due to higher financial stakes.)

Stango and Zinman (2009) find other evidence of inattention: the median US checking-account holder could avoid more than 60% of overdraft charges by using alternative cards with available liquidity. Using different US data, Stango and Zinman (2012) find that at least 30 percent of overdraft fees are avoidable and that in survey responses “60% of overdrafters reported overdrafting

because they “thought there was enough money in my account.”<sup>7</sup> Survey data also show that UK checking-account holders accrue overdraft fees due to inattention (Armstrong and Vickers 2012).

Liebman and Zeckhauser (2004) analyze optimal nonlinear pricing given *ironing* or *spotlighting* decision errors. Consumers who iron have perfect foresight but confuse average price with marginal price. Consumers who spotlight myopically consider only on the transaction price of the current unit. In contrast, I assume consumers make choices optimally conditional on their limited memory. Aron-Dine, Einav, Finkelstein and Cullen (2011) provide evidence for partial spotlighting in healthcare consumption but reject complete spotlighting. There is empirical evidence of ironing for individual labor choices (Liebman and Zeckhauser 2004) and electricity consumption choices (Ito 2010). This may reflect the fact that a typical consumer does not realize the tax code or electricity pricing are nonlinear, in which case average price is a good estimate of marginal price. In settings I focus on, however, ironing is unlikely because consumers are fully aware that contracts include an allowance of ‘free’ units.

In this paper, inattentive consumers are aware of prices when signing a contract, but are uncertain about marginal prices at the point of sale. Many models of add-on pricing examine the opposite situation, by assuming that consumers are aware of transaction prices at the time of purchase, but are unaware of marginal prices or hidden fees at the time they make an ex ante decision to visit a store or purchase a base product (Diamond 1971, Ellison 2005, Gabaix and Laibson 2006, Bubb and Kaufman 2011). As a result, add-ons are sold at monopoly prices in spite of competition or the use of two-part tariffs, either of which would normally lead to marginal cost pricing. Much of this work focuses on firms’ incentives to unshroud hidden fees (Gabaix and Laibson 2006, Heidhues, Köszegi and Murooka 2012).

A common finding between the biased belief model with attentive consumers (Section 5.2) and the existing behavioral industrial organization literature is that demand underestimation, due either to biased beliefs (Eliaz and Spiegel 2008, Grubb 2009), myopia (Gabaix and Laibson 2006, Miao 2010), or naive quasi-hyperbolic-discounting (DellaVigna and Malmendier 2004, Eliaz and Spiegel 2006), leads to high marginal prices above marginal cost. (See Spiegel (2011) for a survey.) In competitive markets, models typically predict that firms do not benefit because they offset high marginal fees with low fixed fees. In such markets, Gabaix and Laibson (2006), Bubb and Kaufman (2011), and Armstrong and Vickers (2012) focus on the cross-subsidization of unbiased consumers by biased consumers. Armstrong and Vickers (2012) argue that such cross-subsidization

---

<sup>7</sup>Stango and Zinman (2012) also show that individuals who are reminded about overdraft fees by answering an online survey with related (but uninformative) questions such as “Do you have overdraft protection?” are substantially less likely to overdraft. This is similar to Agarwal, Driscoll, Gabaix and Laibson’s (2011) finding that accruing one credit card late penalty fee reduces the likelihood of incurring one in the following month.

is particularly troubling in retail banking, where those paying the most in fees are below average income, and discuss five regulatory interventions including point-of-sale alerts. I abstract from heterogeneity in sophistication and the resulting cross-subsidization to focus on heterogeneity in add-on demand and the interaction between inattention and bias.

Miao (2010) and Heidhues et al. (2012) show that profits from aftermarket sales are not necessarily competed away in primary market competition because firms cannot set negative prices for primary goods.<sup>8</sup> Requiring total prices be nonnegative (the *no-free-lunch* constraint), I also find that biased beliefs soften price competition. Moreover, I show that inattention can exacerbate the effect by shifting competition from base marginal charges to less-salient penalty fees.

Competition is often sufficient to protect biased but attentive consumers from exploitation. In contrast, I show that biased and inattentive consumers can be exploited even in fairly-competitive markets if the product is socially wasteful or there is a lower bound on prices. In complementary work, Heidhues et al. (2012) also highlight the importance of socially wasteful products and non-negative pricing for consumer exploitation but focus on firms' incentives for innovation in hidden fees and product quality.

Finally, the price-discrimination model with attentive consumers is a competitive sequential-screening model and hence closely related to the literatures on monopoly sequential-screening (surveyed by Rochet and Stole ((2003), Section 8)) and competitive static-screening (surveyed by Stole (2007)), as discussed in Section 4.

### 3 Single Contract Benchmark

This section develops the model structure used throughout the paper. The benchmark assumptions that are relaxed later are that all consumers have correct beliefs and the same demand at the time of contracting. After describing the model, I derive optimal strategies of attentive and inattentive consumers. Attentive consumers solve a dynamic programming problem and buy all units valued above a critical threshold which is a function of the date and past consumption. Inattentive consumers cannot condition on past usage, so implement a constant threshold. I define bill-shock regulation as a requirement for firms to disclose information that perfectly substitutes for attention, which in the context of the model is equivalent to requiring firms post transaction prices at the point of sale. Comparing equilibrium pricing with inattentive consumers to that with attentive consumers thus illuminates the effect of bill-shock regulation. I show an equivalence result: neither

---

<sup>8</sup>Ellison (2005) shows that shrouded add-on fees can soften price competition without biased beliefs, if the consumers most price-sensitive to cuts in fixed fees are those least likely to purchase add-ons.

inattention nor bill-shock regulation affects substantive market outcomes.

### 3.1 Model

There are mass 1 of consumers and  $N \geq 1$  firms. Each consumer privately learns a vector  $\mathbf{x}$ , describing his or her nonnegative idiosyncratic costs of doing business with each of the  $N$  firms. At the contracting stage,  $t = 0$ , firms simultaneously offer contracts, and each consumer either signs a contract or receives her outside option (normalized to zero). At each later period,  $t \in \{1, 2\}$ , consumers privately learn a taste shock  $v_t$  that measures a consumer's value for a unit of add-on service. Taste shocks  $v_t$  are drawn independently with cumulative distribution  $F$  that is atomless and has full support on  $[0, 1]$ . Then consumers (who have accepted a contract) make a binary quantity choice,  $q_t \in \{0, 1\}$ , by choosing whether or not to consume a unit of service. In the final period, consumers contracted with firm  $i$  make a payment  $P(\mathbf{q}, \mathbf{p}^i)$  to firm  $i$ , as a function of past quantity choices  $\mathbf{q} = (q_1, q_2)$ . Firm  $i$ 's offered contract can be any deterministic price schedule.<sup>9</sup> A contract is characterized by a vector of prices  $\mathbf{p}^i = (p_0^i, p_1^i, p_2^i, p_3^i)$  that includes a fixed fee  $p_0^i$ , base marginal charges  $p_1^i$  and  $p_2^i$  charged for purchasing a unit in either period 1 or 2 respectively, and an additional penalty fee  $p_3^i$  charged if both units are purchased:

$$P(\mathbf{q}, \mathbf{p}^i) = p_0^i + p_1^i q_1 + p_2^i q_2 + p_3^i q_1 q_2. \quad (1)$$

A consumer's net utility is his gross utility less an idiosyncratic cost of doing business with the firm, such as a transportation cost. A consumer's gross utility  $u$  from contracting with firm  $i$  is a function of the value of the base good  $v_0$ , add-on quantity choices  $q_t$ , private taste shocks  $v_t$ , and payment to the firm:

$$u(\mathbf{q}, \mathbf{v}, \mathbf{p}^i) = v_0 + q_1 v_1 + q_2 v_2 - P(\mathbf{q}, \mathbf{p}^i). \quad (2)$$

Conditional on contract prices  $\mathbf{p}$ , a consumer's optimal consumption strategy can be described by a function mapping valuations to quantity choices:  $\mathbf{q}(\mathbf{v}; \mathbf{p})$ . A consumer's expected gross utility from contracting with firm  $i$  and making optimal consumption choices thereafter is

$$U^i = E [u(\mathbf{q}(\mathbf{v}; \mathbf{p}^i), \mathbf{v}, \mathbf{p}^i)].$$

Similarly, let

$$S^i = v_0 + E \left[ \sum_{t=1}^2 (v_t - c) q_t(\mathbf{v}; \mathbf{p}^i) \right]$$

---

<sup>9</sup>See Rochet and Stole (2002) for an insightful discussion of this assumption.

be the expected gross social surplus (excluding transportation costs) generated by a consumer contracting with firm  $i$  and making optimal consumption choices.

A consumer's expected net utility,

$$U^i - x^i,$$

includes the consumer's idiosyncratic cost  $x^i$  of doing business with firm  $i$ . If firm  $i$  is a monopoly,  $x^i$  can be interpreted as the consumer's outside option, which when subtracted from gross utility normalizes the outside option to zero. In a market with competitors,  $x^i$  can be interpreted as a transportation cost or brand taste. Thus, fraction  $G(U^i; U^{-i})$  of consumers buy from firm  $i$  if  $i$  offers expected gross utility of  $U^i$ , while competitors offer  $U^{-i}$ :

$$G(U^i; U^{-i}) \equiv \Pr(U^i - x^i \geq \max_{j \neq i} \{U^j - x^j, 0\}).$$

For a monopolist,  $G(U^i) = \Pr(x^i \leq U^i)$  is the exogenous distribution of a consumer's outside option,  $x^i$ .

Firm profits per consumer equal payments less fixed costs (normalized to zero) and marginal cost  $c \geq 0$  per unit served. Thus firm  $i$ 's expected profits are

$$\Pi^i = G(U^i; U^{-i}) E [P(\mathbf{q}(\mathbf{v}; \mathbf{p}^i), \mathbf{p}^i) - c(q_1(\mathbf{v}; \mathbf{p}^i) + q_2(\mathbf{v}; \mathbf{p}^i))],$$

which can always be rewritten in terms of expected gross social surplus and consumer utility:

$$\Pi^i = G(U^i; U^{-i}) (S^i - U^i).$$

### 3.2 Consumer Strategies

The optimal consumption rule for an attentive consumer who signs a contract would be to consume a unit of service at time  $t$  if and only if her value for the unit,  $v_t$ , exceeds a threshold  $v^*(\mathbf{q}^{t-1}, t)$  that is a function of the date  $t$  and the vector of past usage choices  $\mathbf{q}^{t-1}$ . Let the period one and two thresholds be  $v_1^*$  and  $v_2^*(q_1)$  respectively. Then, suppressing firm  $i$  superscripts from prices,

$$v_2^*(q_1) = p_2 + p_3 q_1, \tag{3}$$

and  $v_1^*$  (derived in Appendix D.1) depends on the distribution of taste shocks:

$$v_1^* = p_1 + (1 - F(p_2 + p_3)) p_3 + \int_{p_2}^{p_2 + p_3} (v - p_2) f(v) dv. \tag{4}$$

The intuition is that  $v_1^*$  equals the expected marginal price conditional on purchase,  $p_1 + (1 - F(p_2 + p_3))p_3$ , plus the expected opportunity cost of foregone second-period purchases,  $\int_{p_2}^{p_2+p_3} (v_2 - p_2) f(v_2) dv_2$ . Integrating by parts, equation (4) simplifies to

$$v_1^* = p_1 + \int_{p_2}^{p_2+p_3} (1 - F(v)) dv. \quad (5)$$

An inattentive consumer cannot condition her strategy on past usage  $\mathbf{q}^{t-1}$  because she does not keep track of past usage. She exhibits imperfect recall. Moreover, for tractability, I assume that inattentive consumers are not aware of the time period  $t$  within the billing cycle.<sup>10</sup> Otherwise, I assume that inattentive consumers are entirely rational and, in particular, are aware of their own inattention and plan accordingly.<sup>11</sup> Formally, the consumer's decision problem exhibits Piccione and Rubinstein's (1997) *absentmindedness*. Unlike optimal strategies for Piccione and Rubinstein's (1997) absent-minded driver, I show that an inattentive consumer's optimal strategy is time consistent and hence Bayesian Nash Equilibrium is an appropriate solution concept.<sup>12</sup> Proposition 1 describes an inattentive consumer's optimal strategy.

**Proposition 1** *An inattentive consumer's optimal strategy is a constant threshold strategy, to buy if and only if  $v_t$  exceeds  $v^*$ . The optimal threshold  $v^*$  is equal to the expected marginal price conditional on purchasing in the current period and satisfies:*

$$v^* = \frac{p_1 + p_2}{2} + (1 - F(v^*))p_3. \quad (6)$$

*Equation (6) is necessary up to the fact that all thresholds above one are equivalent and all thresholds below zero are equivalent. For all  $p_3 \geq 0$ , equation (6) has a unique solution and is sufficient as*

---

<sup>10</sup>Assuming that consumers do not attend to the date makes the model more tractable but does not qualitatively change the primary welfare results. If inattentive consumers did know  $t$ , then sufficiently high penalty fees would lead them to choose different calling thresholds in each period. For example, suppose a firm offered a million dollar subsidy for the first unit ( $p_1 = p_2 = -10^6$ ) but charged a multi-million dollar penalty for a second unit ( $p_3 = 10^7$ ). An inattentive consumer who knew the date would always buy in period 1 but never buy in period 2 to ensure collection of the subsidy and avoidance of the penalty. This consumer response makes such large penalty fees futile as they are never collected. Hence assuming inattentive consumers know the date  $t$  would endogenously limit the size of penalty fees but would not qualitatively affect the primary pricing or welfare predictions.

<sup>11</sup>Inattentive consumers are unaware of past shocks  $v^{t-1}$ , usage  $q^{t-1}$ , or the current date  $t$ . They are aware of this limitation, the distribution of their taste shocks  $F$ , and can remember their chosen consumption thresholds  $v^*$ .

<sup>12</sup>I assume consumers choose a consumption strategy when they sign a contract. This rules out suboptimal equilibria that exist in the game modeled between multiple selves. An alternate interpretation of the game is that the decision makers at times one and two are distinct family members who share a joint account but do not communicate purchases to each other between model periods 1 and 2. By ruling out suboptimal equilibria I implicitly assume that they can communicate ex ante and coordinate on the best equilibrium, which seems reasonable for family members who choose to setup joint accounts.

well as necessary for  $v^*$  to be the optimal threshold. A consumer's choice of  $v^*$  is time consistent, she will find it optimal to follow through and implement her chosen  $v^*$  in periods one and two.

The threshold strategy described by Proposition 1 is intuitive. It says that a consumer will buy a unit if and only if her value exceeds its expected marginal price, which is given in equation (6). The first term in equation (6) captures the expected base marginal charge, which may be  $p_1$  or  $p_2$  depending on the date. The second term captures the expected penalty fee. Considering a purchase at time 1, the penalty fee  $p_3$  applies if the consumer also purchases at time 2, which happens with probability  $(1 - F(v^*))$  given the threshold strategy. Similarly, considering a purchase at time 2, the penalty fee applies if the consumer already purchased at time 1. An inattentive consumer cannot remember if she purchased last period, but given her own threshold strategy, she knows that the probability she made the purchase is  $(1 - F(v^*))$ . From either perspective,  $(1 - F(v^*))p_3$  is the expected penalty fee.

Note that given fixed prices and a positive penalty fee, equation (6) implies that  $v^*$  and  $(1 - F(v^*))$  both increase as the distribution of values  $F$  increases in a first-order stochastic dominance sense. Thus as anticipated demand increases, the likelihood of incurring a penalty fee and the expected marginal price both increase, leading consumers to be more selective in their consumption choices.

### 3.3 Bill-Shock Regulation

Suppose that a firm faced some inattentive consumers and had the option to disclose information at the point of sale that would be a perfect substitute for attention. In the context of the model this would mean disclosing the date and, in period 2, whether or not the penalty fee applies.

**Definition 1** *Bill-shock regulation requires firms to disclose information at the point of sale that is a perfect substitute for attention.*

Within the model, which only incorporates two consumption opportunities, my definition of bill-shock regulation is equivalent to a *price-posting* requirement that firms disclose the transaction price applicable at the point of sale, which is similar to the requirement in the FCC's recent bill-shock agreement. Note that in a richer model with more than two purchase opportunities, a perfect substitute for attention would in general require reporting the full purchase history  $\mathbf{q}^{t-1}$ , which could be cumbersome relative to price-posting. However, in practice firms commonly set prices only as a function of total purchases  $\sum_{t=1}^T q_t$  and, in this case, disclosing total purchases to date is sufficient to make inattentive consumers attentive. For instance, in the case of cellular phones,

bill-shock regulation might require disclosures on the phone screen of the simple form “107 included minutes and 10 days remaining in billing cycle.”<sup>13</sup>

An alternative regulation that could be considered would prohibit the use of penalty fees:

**Definition 2** *Banning penalty fees is the requirement that firms charge a constant marginal price as a function of usage:  $p_1 = p_2$  and  $p_3 = 0$ .*

In the benchmark model (as well as the model of biased beliefs in Section 5) it will be a result that firms optimally offer attentive consumers two-part tariffs with zero penalty fees. In this case, the two forms of regulation have the same effect on market outcomes, since inattentive consumers behave as attentive consumers do when penalty fees are zero. Moreover, although the formal results in Sections 4 and 6 are shown only for bill-shock regulation, the two regulations would have qualitatively similar effects (see Online Appendix C).

### 3.4 Benchmark Result

When consumers have homogeneous unbiased beliefs ex ante, firms do best by setting marginal charges to implement the first-best allocation and extracting surplus through the fixed fee (balancing the trade-off between mark-up and volume in the standard way). This is made transparent by writing firm  $i$ 's profits in terms of expected gross social surplus and consumer utility:  $\Pi^i = G(U^i; U^{-i})(S^i - U^i)$ . For any fixed utility offer  $U^i$ , firm profits are maximized by choosing marginal prices  $p_1^i$ ,  $p_2^i$ , and  $p_3^i$  to achieve first-best surplus, while adjusting the fixed fee  $p_0^i$  to keep  $U^i$  constant. This is true independent of regulation.

If consumers are attentive, achieving first-best allocations requires setting the marginal price of all units equal to marginal cost. If consumers are inattentive, however, achieving first-best allocations only requires that the expected marginal price equal marginal cost. As a result, inattention allows for contracts with positive surprise penalty fees in equilibrium but bill-shock regulation restricts equilibrium pricing and eliminates such penalty fees. Nevertheless, bill-shock regulation does not affect allocations or the division of surplus.

**Proposition 2** *If consumers have homogeneous unbiased beliefs,  $v_t \sim F(v_t)$ , then equilibrium allocations are efficient. If at least some consumers are attentive then there is marginal cost pricing*

---

<sup>13</sup>Cellular bills are typically only a function of total calling within each calling category (peak, off-peak, etc.), and do not depend on when during the billing cycle calls occurred. Note that I find that it is optimal for firms to deviate from such simple pricing when consumers are attentive. However, it is reasonable to believe that in practice firms are restricted to price as a function only of total usage because contract complexity is inherently expensive. Adding such a restriction to the model would not qualitatively change the main predictions about the consequences of regulation in Propositions 2 and 9-11 or Corollaries 2 and 4.

( $p_1 = p_2 = c$  and  $p_3 = 0$ ). If all consumers are inattentive then the set of possible equilibrium prices is larger and includes all three-part tariffs with  $p_1 = p_2 = p$  and  $p_3 = \frac{c-p}{1-F(c)}$  for  $p \in [0, c]$ . Both bill-shock regulation and banning penalty fees would restrict equilibrium prices but have no effect on allocations, firm profits, or consumer surplus.

One might have thought that surprise penalty fees could be used to profitably exploit inattentive consumers. However, while Proposition 2 shows that surprise penalty fees can be weakly optimal if all consumers are inattentive, it also shows that firms cannot do strictly better than marginal cost pricing. This is because inattentive consumers who are aware of their own inattention and have unbiased beliefs about their future value for consumption cannot be exploited. Thus, under benchmark assumptions, the firm cannot do better than maximizing total surplus.

Assuming all consumers are inattentive, the equivalence result in Proposition 2 appears to capture an argument of some critics of bill-shock regulation: that it would only cause firms to recoup lost penalty fees through fixed fees and other charges (Federal Reserve Board 2009b). However, the predictions of Proposition 2 are fragile and implausible. First, the equivalence result is fragile because it relies on the joint assumptions of homogeneous demand and correct beliefs. Second, the predictions of Proposition 2 are hard to reconcile with firm behavior. In particular, Proposition 2 cannot explain firms' expressed aversion to bill-shock regulation (see footnote 2 and Section 7).

## 4 Price Discrimination via Multiple Contracts

In this section, I relax the assumption of homogeneous demand imposed in the benchmark model and show that heterogeneous demand and the resulting incentive for firms to price discriminate can explain why consumer inattention is strictly profitable for firms. In this alternative setting, the equivalence result fails and bill-shock regulation does affect substantive market-outcomes. In particular, bill-shock regulation will be socially harmful in fairly-competitive markets. To analyze the effect of bill-shock regulation, I analyze equilibrium with and without the regulation and compare. I begin by characterizing equilibrium with bill-shock regulation, which is simpler and corresponds to all consumers being attentive.

### 4.1 Model

Consider a model with two types of consumers. Prior to choosing a contract, each consumer privately receives one of two private signals  $s \in \{L, H\}$ , where  $\Pr(s = H) = \beta$ . As a result, each firm  $i$  simultaneously offers a menu with a choice of two contracts,  $s \in \{L, H\}$ . Each contract is characterized by the vector of prices  $\mathbf{p}_s^i = (p_{0s}^i, p_{1s}^i, p_{2s}^i, p_{3s}^i)$ , which specifies payments as in

equation (1). Each consumer either signs a contract,  $\hat{s} \in \{L, H\}$ , from one of the firms or receives her outside option (normalized to zero).

As before, at each later period,  $t \in \{1, 2\}$ , a consumer privately learns her value  $v_t$  for a unit of add-on service. Conditional on receiving signal  $s$ , a consumer's values  $v_t$  are drawn independently with cumulative conditional distribution  $F_s$ , which is atomless and has full support on  $[0, 1]$ . The conditional value distributions are ranked by first-order stochastic dominance,  $F_L(v) \geq F_H(v)$ , and the ranking is strict at  $v = c$ . Marginal cost is assumed to be less than 1 so that the service is socially valuable:  $c \in [0, 1)$ .

As before, a consumer's net utility equals her gross utility less a transportation or brand cost, where gross utility  $u(\mathbf{q}, \mathbf{v}, \mathbf{p}_s^i)$  is given by equation (2). The expected gross utility of a consumer of type  $s$  who chooses contract  $\hat{s}$  from firm  $i$  at time zero and makes optimal consumption choices thereafter is

$$U_{s\hat{s}}^i = E[u(\mathbf{q}_s(\mathbf{v}; \mathbf{p}_s^i), \mathbf{v}, \mathbf{p}_s^i) | s],$$

where  $\mathbf{q}_s(\mathbf{v}; \mathbf{p})$  is the optimal consumption rule for type  $s$  given prices  $\mathbf{p}$ . Define  $U_s^i \equiv U_{ss}^i$  to be the expected gross utility of a consumer who chooses the intended contract from firm  $i$ . Similarly, let

$$S_s^i = v_0 + E\left[\sum_{t=1}^2 (v_t - c) q_{t,s}(\mathbf{v}; \mathbf{p}_s^i) | s\right]$$

be the expected gross social surplus (excluding transportation or brand costs) from a consumer of type  $s$  who chooses contract  $s$  from firm  $i$  and makes optimal consumption choices at  $t \in \{1, 2\}$ .

A consumer's expected net utility,  $U_s^i - x^i$ , includes transportation or brand cost  $x^i$ . Fraction  $G_s(U_s^i; U_s^{-i})$  of consumers of type  $s$  buy from firm  $i$  if firm  $i$  offers contract  $s$  with expected gross utility of  $U_s^i$ , while competitors offer  $U_s^{-i}$ :

$$G_s(U_s^i; U_s^{-i}) = \Pr(U_s^i - x^i \geq \max_{j \neq i} \{U_s^j - x^j, 0\}).$$

In the case of monopoly,  $x^i$  can be interpreted as a consumer's exogenous outside option, with cumulative distribution  $G(U^i) = \Pr(x^i \leq U^i)$ .

Suppressing competitors' offers  $U_s^{-i}$  and firm  $i$  superscripts from the notation, the firm's expected profit maximization problem is:

$$\begin{aligned} & \max_{\mathbf{p}_L, \mathbf{p}_H} ((1 - \beta) G_L(U_L) (S_L - U_L) + \beta G_H(U_H) (S_H - U_H)) \\ & \text{s.t. } U_H \geq U_{HL} \text{ (downward IC) and } U_L \geq U_{LH} \text{ (upward IC)}. \end{aligned}$$

This initial statement of the firm’s problem encompasses both attentive and inattentive cases. The two cases only differ by the consumers’ optimal consumption rule  $\mathbf{q}_s(\mathbf{v}; \mathbf{p})$ , which is a constant threshold strategy for inattentive consumers but conditions calling thresholds on past consumption for attentive consumers (Section 3.2). The constraint that type  $H$  not choose contract  $L$  ( $U_H \geq U_{HL}$ ) is the *downward* incentive constraint. The constraint that type  $L$  not choose contract  $H$  ( $U_L \geq U_{LH}$ ) is the *upward* incentive constraint.

Conceptually, the firm’s pricing problem can be broken into two parts. First, the firm’s choice of marginal prices determines contract allocations and hence expected social surpluses from serving each type,  $S_L$  and  $S_H$ . Second, the firm’s choice of fixed fees then determines the utilities offered to each type,  $U_L$  and  $U_H$ . The differences  $\mu_s \equiv (S_s - U_s)$  are the firm’s expected markups on each contract and the profits per customer served. Absent ex ante incentive constraints, the choice of markups would be a standard monopoly pricing problem.

With or without regulation, I begin by solving the monopoly pricing problem, where  $G_s(U_s)$  is an exogenous distribution of outside options. For the monopoly case, I make one of two assumptions: (1) Zero outside option monopoly (ZOOM):  $G_s(U_s)$  is one if  $U_s \geq 0$  and zero otherwise, which captures a monopolist serving horizontally-homogeneous customers ( $x = 0$  for all).<sup>14</sup> (2) Heterogeneous outside options (HOO):  $G_s(U_s)$  is differentiable and  $U_s + \frac{G_s(U_s)}{g_s(U_s)}$  is strictly increasing, which corresponds to a decreasing marginal revenue assumption, guaranteeing the simple monopoly pricing problem has a uniquely optimal markup.

**Definition 3** *Unconstrained optimal markup  $\mu_s^*$  is the optimal expected markup for type  $s$  given first-best allocations and ignoring ex ante incentive constraints:  $\mu_s^* = S_s^{FB} - \hat{U}_s$  where  $\hat{U}_s \equiv \arg \max_U G_s(U) (S_s^{FB} - U)$ .*<sup>15</sup>

Unconstrained optimal markups are those that would be charged under third-degree price discrimination. Given ZOOM,  $\mu_H^* > \mu_L^*$ , while given HOO, demand will satisfy one of three cases: (1)  $\mu_L^* = \mu_H^*$ , (2)  $\mu_H^* > \mu_L^*$ , or (3)  $\mu_H^* < \mu_L^*$ . The relative ranking of unconstrained optimal markups is important because it determines which market segment (if either) the firm would like to offer a discounted markup to. I characterize optimal contracts in each case, but I will often focus on

---

<sup>14</sup>I assume there are  $T = 2$  sub-periods when quantity choices are made after a contract is signed. Given attentive consumers and  $T = 1$ , ZOOM coincides with Courty and Li (2000), which models airline-ticket refund-contracts. When consumers are attentive and  $T \geq 1$ , ZOOM is nearly a special case of the problem studied by Pavan, Segal and Toikka (2011). However, because I assume period-zero types are discrete rather than continuous, Pavan et al.’s (2011) results do not apply, and conditional independence of values does not lead to a repetition of the Courty and Li (2000) solution. Moreover, I allow for heterogeneous outside-options so that I can move beyond monopoly pricing and analyze imperfect competition.

<sup>15</sup>Given ZOOM,  $\hat{U}_s = 0$  and  $\mu_s^* = S_s^{FB}$ . Given HOO,  $\mu_s^* = \frac{G_s(\hat{U}_s)}{g_s(\hat{U}_s)}$  where  $\hat{U}_s$  uniquely satisfies  $S_s^{FB} = \hat{U}_s + \frac{G_s(\hat{U}_s)}{g_s(\hat{U}_s)}$ .

the case in which  $\mu_H^* > \mu_L^*$ . This is a natural assumption if high-average-value customers are high-income customers who have a low marginal-value of money.

## 4.2 Pricing with bill-shock regulation (Attentive Case)

I first characterize equilibrium pricing when consumers are attentive because this will be the outcome if bill-shock regulation is imposed. (Recall that disclosures mandated by bill-shock regulation compensate for inattentive consumers' limited memory, enabling them to operate like attentive consumers.) The main result will be that equilibrium contracts induce inefficient allocations except in knife-edge circumstances. This will be for the standard reason in second-degree price-discrimination models: to give one group a discounted markup relative to another, the discount must be accompanied by a distorted allocation to prevent everyone choosing the discounted markup.

Let  $v_{s\hat{s}}$  be the optimal first-period consumption-threshold of an attentive consumer of type  $s$  who chooses contract  $\hat{s}$  and let  $v_s = v_{ss}$ . The expression for  $v_{s\hat{s}}$  is an extension of equation (5):

$$v_{s\hat{s}} = p_{1\hat{s}} + \int_{p_{2\hat{s}}}^{p_{2\hat{s}}+p_{3\hat{s}}} (1 - F_s(v)) dv. \quad (7)$$

An attentive consumer  $s$  who chooses contract  $\hat{s}$  earns expected gross utility

$$\begin{aligned} U_{s\hat{s}} = & v_0 - p_{0\hat{s}} + \int_{v_{s\hat{s}}}^1 (v - p_{1\hat{s}}) dF_s(v) \\ & + F_s(v_{s\hat{s}}) \int_{p_{2\hat{s}}}^1 (v - p_{2\hat{s}}) dF_s(v) + (1 - F_s(v_{s\hat{s}})) \int_{p_{2\hat{s}}+p_{3\hat{s}}}^1 (v - p_{2\hat{s}} - p_{3\hat{s}}) dF_s(v), \end{aligned} \quad (8)$$

and for  $\hat{s} = s$  earns  $U_s = U_{ss}$  and generates expected gross social surplus

$$S_s = v_0 + \int_{v_s}^1 (v - c) dF_s(v) + \int_{p_{2s}+p_{3s}}^1 (v - c) dF_s(v) + F_s(v_s) \int_{p_{2s}}^{p_{2s}+p_{3s}} (v - c) dF_s(v). \quad (9)$$

It is useful to reframe the firm's problem in two ways. First, think of the firm choosing offered utility levels  $U_s$  so that fixed fees  $p_{0s}$  are determined by equation (8) evaluated at  $\hat{s} = s$  as function of  $U_s$ . Second, think of the firm choosing a consumer's first-period threshold  $v_s$  rather than marginal price  $p_{1s}$ . Given a choice of  $v_s$ , it is necessary for  $p_{1s}$  to satisfy equation (7) evaluated at  $\hat{s} = s$ . The firm's problem can then be written as:

$$\begin{aligned} \max_{\substack{U_L, v_L, p_{2L}, p_{3L} \\ U_H, v_H, p_{2H}, p_{3H}}} & (1 - \beta) G_L(U_L) (S_L(v_L, p_{2L}, p_{3L}) - U_L) + \beta G_H(U_H) (S_H(v_H, p_{2H}, p_{3H}) - U_H) \\ \text{s.t.} & U_H \geq U_{HL} \text{ (downward IC) and } U_L \geq U_{LH} \text{ (upward IC),} \end{aligned}$$

where  $U_{HL}$ ,  $U_{LH}$ ,  $S_L$ , and  $S_H$  are given by equations (8) and (9) and  $p_{1s}$  and  $p_{0s}$  are given by

equations (7) and (8) evaluated at  $\hat{s} = s$  for  $s \in \{L, H\}$ .

Proposition 3 characterizes the solution to a single firm's problem, treating residual demand  $G_s(U_s)$  as exogenous. Proposition 4 applies the result to a Hotelling duopoly, where firm  $i$ 's residual demand  $G_s(U_s^i, U_s^j)$  depends endogenously on firm  $j$ 's equilibrium offer  $U_s^j$ .

The solution to the firm's problem varies depending on which incentive constraints bind. This, in turn, depends on how unconstrained optimal markups (Definition 3) are ranked across low and high market-segments. When there is no reason to price discriminate ( $\mu_L^* = \mu_H^*$ ) neither ex ante incentive constraint binds and the firm offers a single first-best contract. When market segment  $L$  would receive a discounted markup under third-degree price discrimination ( $\mu_L^* < \mu_H^*$ ) the downward incentive constraint is binding, contract  $H$  is first best, and marginal prices on contract  $L$  are above marginal cost, distorting allocations downwards. When market segment  $H$  would receive a discounted markup under third-degree price discrimination ( $\mu_L^* > \mu_H^*$ ) the reverse is true: the upward incentive constraint is binding, contract  $L$  is first best, and marginal prices on contract  $H$  are below marginal cost, distorting allocations upwards.

As outlined above, demand curves fall into one of three categories. Proposition 3 characterizes optimal monopoly contracts for each case and shows that penalty fees are strictly positive on distortionary contracts, irrespective of the direction of the distortion.

**Proposition 3** Monopoly Best Response (with bill-shock regulation): Assume demand curves  $G_L(U_L)$  and  $G_H(U_H)$  satisfy ZOOM or HOO. Optimal monopoly contracts satisfy the following:

1. If  $\mu_L^* = \mu_H^*$ , then a single marginal-cost contract with markup  $\mu_L^*$  gives both types first-best allocations.
2. If  $\mu_H^* > \mu_L^*$ , then  $H$ 's allocation is first best via marginal-cost pricing but  $L$ 's allocation is distorted downwards:  $v_L, p_{2L}, p_{2L} + p_{3L} > c$ . Penalty fee  $p_{3L}$  is strictly positive. The triple  $\{v_L, p_{2L}, p_{3L}\}$  satisfies equations (17)-(19) in Appendix A.
3. If  $\mu_H^* < \mu_L^*$ , then  $L$ 's allocation is first best via marginal-cost pricing but  $H$ 's allocation is distorted upwards:  $v_H, p_{2H}, p_{2H} + p_{3H} < c$ . Penalty fee  $p_{3H}$  is strictly positive. The triple  $\{v_H, p_{2H}, p_{3H}\}$  satisfies equations (20)-(22) in Appendix A.

To understand Proposition 3, begin with case (2). The downward incentive constraint binds because the firm would like to offer the low segment a discounted markup ( $\mu_H^* > \mu_L^*$ ). The fact that marginal prices are distorted above marginal cost on the low contract follows from standard price discrimination logic. High types find increases in marginal prices more costly than do low types because high types make more purchases. Thus raising marginal prices on the low contract relaxes

the downward incentive constraint (discouraging the high type from choosing the low contract) at the cost of distorting low-types' allocations downwards. The positive penalty fee  $p_{3L}$  makes the second-period marginal price larger after an initial purchase. This is optimal because a deviating high type is more likely to purchase in the first period than a low type. (A more detailed intuition for optimal pricing accompanies the first-order conditions in Appendix A.)

If  $\mu_H^* = \mu_L^*$  then firms have no desire to price discriminate, which means that a single marginal cost contract is optimal. If  $\mu_H^* < \mu_L^*$  then optimal pricing follows a similar intuition to that for the case  $\mu_H^* > \mu_L^*$ , but distortions are reversed because high-types receive a discount and hence the upward incentive constraint binds rather than the downward constraint.

Proposition 3 can explain the use of penalty fees but not the use of *surprise* penalty fees. Importantly, Proposition 3 shows that allocations are first best only when unconstrained optimal markups are identical for both types. As Proposition 4 shows, this implies that allocations are only efficient in a Hotelling duopoly when both market segments have identical transportation costs.

Turning to oligopoly, consider a *simple duopoly*:

**Definition 4** *In a simple duopoly, firms with marginal costs  $c > 0$  compete on a uniform Hotelling line. Transport costs are  $\tau_H$  and  $\tau_L > 0$  for high and low types respectively and are sufficiently small for strict full-market-coverage.<sup>16</sup>*

Proposition 4 qualitatively characterizes the simple duopoly equilibrium according to whether  $\tau_H$  is larger than, smaller than, or equal to  $\tau_L$ . The main step in proving Proposition 4 is to show that, in equilibrium, unconstrained optimal markups  $\mu_L^*$  and  $\mu_H^*$  have the same relative ranking as transport costs  $\tau_L$  and  $\tau_H$ . (This is intuitive, recalling that with a single market segment equilibrium markups would equal the transport costs.) Given this result, Proposition 4 follows from a firm's best response characterized in Proposition 3.

**Proposition 4** *Simple Duopoly Equilibrium (with bill-shock regulation):*

1. If  $\tau_H = \tau_L = \tau$  then the unique equilibrium is for firms to split the market and each offer a single marginal-cost contract with fixed-fee markup of  $\tau$ .
2. If  $\tau_H \neq \tau_L$  then all equilibria are inefficient.

---

<sup>16</sup>Strict full-market-coverage requires that every consumer *strictly* prefer the best offer to her outside option. Results generalize to oligopolies with more than two firms. The important simplifying assumption is full market coverage, which implies that regulation affects allocations on the intensive margin but not the extensive margin.

3. If  $\tau_H > \tau_L$  then, in all symmetric equilibria, high types receive first-best allocations, while low types' allocation is distorted downwards. For  $\tau_H < \tau_L$ , low types receive first best, while high types' allocation is distorted upwards.

The knife-edge efficiency-result in Proposition 3 and Proposition 4 is analogous to findings by Armstrong and Vickers (2001) and Rochet and Stole (2002) in a static rather than sequential screening context. Moreover it is very intuitive: If unconstrained optimal markups are equal, firms can implement first-best allocations with marginal-cost pricing and charge both groups the same fixed fee. If  $\mu_L^* < \mu_H^*$ , however, a firm would like to maintain first-best allocations but offer low types a discount relative to high types. This is not incentive compatible, as high types would always pool with low types and choose the discount. As a result, firms are forced to distort the allocation of the low type downwards to maintain incentive compatibility.

### 4.3 Pricing without bill-shock regulation (Inattentive case)

I now characterize equilibrium pricing when consumers are inattentive and respond only to the expected marginal price because they do not keep track of past usage. I first solve the firm's problem assuming that the firm keeps penalty fees a surprise and then later show that this nondisclosure is optimal in fairly-competitive markets. It is striking that, in contrast to the attentive case, firms can charge different markups to different market segments without distorting allocations. This leads to the result that bill-shock regulation will reduce welfare in fairly-competitive markets.

Define  $\bar{p}_s = (p_{1s} + p_{2s})/2$ . When consumers are inattentive, any pair  $\{p_{1s}, p_{2s}\}$  which have the same average are equivalent, both in terms of allocations and surplus division. I focus on symmetric pricing,  $p_{1s} = p_{2s}$ , for which the firm's problem reduces to the choice of  $p_{0s}$ ,  $\bar{p}_s$ , and  $p_{3s}$  for  $s \in \{L, H\}$ .

Let  $v_{s\hat{s}}$  be the optimal consumption threshold of an inattentive consumer of type  $s$  who chooses contract  $\hat{s}$ , and let  $v_s = v_{ss}$ . The first-order condition for  $v_{s\hat{s}}$  is a natural extension of equation (6):

$$v_{s\hat{s}} = \bar{p}_s + p_{3\hat{s}}(1 - F_s(v_{s\hat{s}})). \quad (10)$$

An inattentive consumer  $s$  who chooses contract  $\hat{s}$  earns expected gross utility

$$U_{s\hat{s}} = v_0 - p_{0\hat{s}} + 2 \int_{v_{s\hat{s}}}^1 v dF_s(v) - 2\bar{p}_s(1 - F_s(v_{s\hat{s}})) - p_{3\hat{s}}(1 - F_s(v_{s\hat{s}}))^2, \quad (11)$$

and for  $\hat{s} = s$  earns  $U_s = U_{ss}$  and generates expected gross surplus

$$S_s = v_0 + 2 \int_{v_s}^1 (v - c) dF_s(v). \quad (12)$$

It is useful to reframe the firm's problem in two ways. First, think of the firm choosing offered utility levels  $U_s$  rather than fixed fees  $p_{0s}$ , which are then determined by equation (11) evaluated at  $\hat{s} = s$ . Second, think of the firm first choosing consumer threshold  $v_s$ , so that  $\bar{p}_s$  is determined by equation (10), and then choosing the best penalty fee  $p_{3s}$  which makes  $v_s^*$  globally (rather than just locally) incentive compatible. Then the firm's problem can then be written as:

$$\begin{aligned} & \max_{\substack{U_L, v_L, p_{3L} \\ U_H, v_H, p_{3H}}} (1 - \beta) G_L(U_L) (S_L(v_L) - U_L) + \beta G_H(U_H) (S_H(v_H) - U_H) \\ \text{s.t. } & U_H \geq U_{HL} \text{ (downward IC), } U_L \geq U_{LH} \text{ (upward IC),} \\ & \text{and } v_L \text{ and } v_H \text{ are incentive compatible,}^{17} \end{aligned}$$

where  $U_{HL}$ ,  $U_{LH}$ ,  $S_L$ , and  $S_H$  are given by equations (11) and (12) and  $p_{1s}$  and  $p_{0s}$  are given by equations (10) and (11) evaluated at  $\hat{s} = s$  for  $s \in \{L, H\}$ .

Notice that only offered utilities  $U_s$  and consumer thresholds  $v_s$  enter the objective function directly. Penalty fees  $p_{3s}$  only affect profits via the incentive constraints. By Proposition 1, choosing  $\bar{p}_s$  to satisfy the first-order condition in equation (10) is sufficient for  $v_s$  to be incentive compatible for all  $p_{3s} \geq 0$ . Moreover, increasing  $p_{3s}$  weakly relaxes both upward and downward ex ante incentive constraints, from which it follows that it is weakly optimal to set  $p_{3s}$  as large as possible.

**Proposition 5** *Increasing  $p_{3L}$  weakly relaxes the downward incentive constraint without affecting the upward incentive constraint. Increasing  $p_{3H}$  weakly relaxes the upward incentive constraint without affecting the downward incentive constraint. It is weakly optimal to choose nonnegative penalties  $p_{3s}$  as large as possible.*

For intuition behind Proposition 5, consider what happens if penalty  $p_{3s}$  is increased by one dollar. First, following equation (10), base marginal charge  $\bar{p}_s$  must be reduced by  $(1 - F_s(v_s))$  to keep expected marginal price  $v_s$  constant. This reduces expected variable payments by  $(1 - F_s(v_s))^2$  because the extra dollar in the penalty fee is paid with probability  $(1 - F_s(v_s))^2$  but the  $(1 - F_s(v_s))$  discount on the base marginal charge is paid with probability  $2(1 - F_s(v_s))$ . Thus, following equation (11), a second change is that the fixed fee is increased by  $(1 - F_s(v_s))^2$  to keep the offered gross utility  $U_s$  constant. By construction, these changes leave type  $s$  indifferent. Any other type  $\hat{s}$

---

<sup>17</sup> $v_s \in \arg \max_x \left\{ 2 \int_x^1 v f_s(v) dv - 2\bar{p}_s (1 - F_s(x)) - p_{3s} (1 - F_s(x))^2 \right\}$ .

that chooses contract  $s$  would pay the same increase in the fixed fee but receive a smaller reduction in expected variable payments. If type  $\hat{s}$  buys with probability  $\pi = 1 - F_{\hat{s}}(v_{\hat{s}s})$ , her expected variable payments are reduced by  $2\pi(1 - F_s(v_s)) - \pi^2$ . Notice that this reduction is maximized at  $\pi = (1 - F_s(v_s))$ . Thus regardless of whether type  $\hat{s}$  is a higher type and  $\pi > (1 - F_s(v_s))$  or a lower type and  $\pi < (1 - F_s(v_s))$ , increasing the penalty  $p_{3s}$  increases total expected payments for her on contract  $s$ .

Proposition 5 suggests that optimal penalty fees could be unreasonably high. In practice, however, they would be restricted by a variety of forces.

**Remark 1** *As penalty fees grow large, the remaining profit increase from increasing them all the way to infinity becomes arbitrarily small because profits are bounded (strictly) below first-best surplus. Hence an arbitrarily small cost of raising penalty fees would be sufficient to endogenously limit penalty fees to be finite. Economic forces that would endogenously restrict penalty fees include: (1) limited liability, (2) mild risk aversion, (3) regulatory threat, (4) a small fraction of consumers who are attentive, (5) rationally inattentive consumers who could invest effort  $k > 0$  to be attentive if it were worth doing so, and (6) consumers who attend to the date.*

For simplicity, I exogenously impose the upper bound  $p_{3s} \leq h_s(v_s)$  stated in Condition 1. This upper bound corresponds either to a cap on penalty fees in the case  $h_s(v_s) = p^{\max} > 0$ , or to the restriction to nonnegative marginal prices in the case  $h_s(v_s) = v_s / (1 - F_s(v_s))$ . Notice that all prior results and statements remain true with this addition to the problem.<sup>18</sup>

**Condition 1** *Penalty fees are bounded by  $p_{3s} \leq h_s(v_s) \in \{p^{\max}, v_s / (1 - F_s(v_s))\}$ .*

Proposition 6 characterizes the solution to a single firm's problem, treating residual demand  $G_s(U_s)$  as exogenous. Proposition 7 applies the result to a fairly-competitive Hotelling duopoly, where firm  $i$ 's residual demand  $G_s(U_s^i, U_s^j)$  depends endogenously on firm  $j$ 's equilibrium offer  $U_s^j$ .

As in the attentive case, the solution to a single firm's problem varies depending on which incentive constraints bind. When the downward incentive constraint is binding, contract  $H$  is first best ( $v_H = c$ ) but contract  $L$  allocations are distorted downwards ( $v_L > c$ ). When the upward incentive constraint is binding, contract  $L$  is first best ( $v_L = c$ ) but contract  $H$  allocations are distorted upwards ( $v_H < c$ ). The crucial difference relative to the attentive case is that both constraints may be slack even when unconstrained optimal markups differ.

Given strictly positive constants  $X_L$  and  $X_H$ , defined by equations (26)-(27) in Appendix A, demand will fall into one of three categories, depending on how unconstrained optimal markups

---

<sup>18</sup>In particular, the constraint is symmetric so that any  $\{p_{1s}, p_{2s}\}$  which have the same mean are still equivalent.

(Definition 3) differ across market-segments. Proposition 6 characterizes optimal monopoly contracts in each case.

**Proposition 6** *Monopoly Best Response (without bill-shock regulation): Assume (1) demand curves  $\{G_L(U_L), G_H(U_H)\}$  satisfy ZOOM or HOO, (2) the firm chooses to keep penalty fees a surprise, and (3) penalty fees are restricted by Condition 1. Optimal monopoly contracts satisfy the following:*

1. If  $\mu_H^* - \mu_L^* \in [-X_L, X_H]$ , then both types receive first-best allocations ( $v_L = v_H = c$ ) and contract mark-ups are  $\mu_L^*$  and  $\mu_H^*$  respectively.
2. If  $\mu_H^* - \mu_L^* > X_H$ , then  $H$ 's allocation is first best ( $v_H = c$ ) but  $L$ 's allocation is distorted downwards ( $v_L > c$ ). Threshold  $v_L$  satisfies equation (24) in Appendix A.  $L$  pays the maximum penalty fee.
3. If  $\mu_H^* - \mu_L^* < -X_L$ , then  $L$ 's allocation is first best ( $v_L = c$ ) but  $H$ 's allocation is distorted upwards ( $v_H < c$ ). Threshold  $v_H$  satisfies equation (25) in Appendix A.  $H$  pays the maximum penalty fee.

Note that while Proposition 6 exogenously assumes nondisclosure of penalty fees, it is clear by comparison to Proposition 3 that if unconstrained optimal markups differ but satisfy  $(\mu_H^* - \mu_L^*) \in [-X_L, X_H]$  then nondisclosure is strictly optimal.

Comparing Propositions 3 and 6 shows the underlying insight of the first main result: the combination of surprise penalty fees and consumer inattention can be both profitable and socially valuable by reducing allocative distortions due to price discrimination when unconstrained optimal markups differ across consumer segments but are not too different. In the attentive problem, contracts implement first-best allocations only for the knife-edge case  $\mu_L^* = \mu_H^*$ . With inattentive consumers this is no longer true. Slack ex ante incentive constraints and first-best allocations are a feature for  $(\mu_H^* - \mu_L^*)$  in an interval around zero because penalty fees relax incentive constraints when consumers are inattentive.

For intuition, suppose that  $\mu_H^* > \mu_L^*$ . If consumers are attentive, firms cannot induce first-best allocations and charge low types a discounted markup. First-best allocations require marginal-cost pricing for every unit on every contract. With identical marginal prices on all contracts, all consumers would choose the lowest fixed fee and pay the same markup. To discount low types' markup, firms must combine a discounted fixed fee with higher marginal prices that distort quantity choices. The striking result for inattentive consumers is that this is no longer the case for small discounts. First-best allocations only require that expected marginal prices equal marginal cost and could, for instance, be implemented by offering  $\bar{p}_s = 0$  and  $p_{3s} = c/(1 - F_s(c))$ . As high types

pay penalty fees more often than low types, these contracts involve lower penalty fees on the high contract to achieve the same expected marginal price. Moreover, high types are willing to pay a higher increase in the fixed-fee for a reduction in the penalty fee than are low types. As a result, the low-type contract can offer a discounted markup without distorting allocations or attracting high types.

To illustrate the preceding intuition, consider the following example:

**Example 1** *For low types,  $v_t \sim U[0, 10]$ , and for high types,  $v_t \sim U[0, 15]$ . Marginal cost is  $c = 5$ .*

*Contract L: Free first unit and a \$10 penalty:  $p_{1L} = p_{2L} = 0$  and  $p_{3L} = 10$ .*

*Contract H: Free first unit and a \$7.5 penalty:  $p_{1H} = p_{2H} = 0$  and  $p_{3H} = 7.5$ .*

Both contracts in Example 1 are efficient for their intended consumers. For low types who choose contract  $L$ , the optimal calling threshold is equal to marginal cost:  $v_L^* = 5$ . At this calling threshold, low types will purchase with probability  $1/2$  in each period so that, conditional on purchasing in the current period, the expected marginal price is  $1/2$  times the penalty fee or  $10/2 = 5$ . Similarly, for high types who choose contract  $H$ , the optimal calling threshold is also equal to marginal cost:  $v_H^* = 5$ . At this calling threshold, high types will purchase with probability  $2/3$  in each period. Thus the lower penalty fee is exactly offset by a higher purchase probability so that expected marginal price is the same:  $(2/3)7.5 = 5$ . Moreover, as contract  $H$  has a lower penalty fee than contract  $L$ , the firm can charge a higher fixed fee on contract  $H$ . In fact, the fixed fee on contract  $H$  can be up to \$1 higher while maintaining incentive compatibility. The \$1 difference in fixed fees corresponds to a higher markup on contract  $H$ :  $\mu_H = \mu_L + 1/6$ . Thus inattention and surprise penalty fees allow the firm to charge different markups without distorting allocations.

The preceding paragraphs focus on the case in which unconstrained optimal markups satisfy  $(\mu_H^* - \mu_L^*) \in [-X_L, X_H]$ , for which Proposition 6 shows that allocations are first best. If firms have sufficient market power, however, this situation does not arise and inattention and surprise penalty fees do not produce efficient outcomes. This is illustrated by Corollary 1.

**Corollary 1** *If a monopolist serves consumers with zero outside option (ZOOM), the downward ex ante incentive constraint binds and the low type's allocation is distorted below first best.*

Relative to a zero-outside-option monopoly, competition compresses unconstrained optimal markups. Firms with stiff competitors never want to charge high types too large a premium over low types because they would lose business to other firms. Thus, if consumers are inattentive, fairly competitive markets will satisfy  $(\mu_H^* - \mu_L^*) \in [-X_L, X_H]$  and yield efficient outcomes with surprise penalty fees. To state the result formally, let  $\tau_H = \tau H$  and  $\tau_L = \tau L$  for  $H > L > 0$  and  $\tau > 0$ , so that  $\tau$  parameterizes the degree of competition.

**Proposition 7** *Simple Duopoly Equilibrium (without bill-shock regulation):* If  $\tau > 0$  is sufficiently small then: Without bill-shock regulation, in the unique (up to penalty fees) symmetric equilibrium, all customers are served, allocations are first best, and mark-ups are  $\mu_s = \tau_s$ . Moreover, surprise penalty fees are charged but not disclosed at the point-of-sale and the set of equilibrium prices includes  $p_{1s}^i = p_{2s}^i = 0$  and  $p_{3s}^i = c / (1 - F_s(c))$ .

The intuition for Proposition 7 is as follows. In equilibrium, unconstrained optimal markups are closely related to transportation costs. Thus, in fairly-competitive markets when  $\tau$  is small, and hence the difference between  $\tau_H$  and  $\tau_L$  is also small, unconstrained optimal markups will satisfy  $(\mu_H^* - \mu_L^*) \in [-X_L, X_H]$ . (By their definition in Appendix A,  $X_L$  and  $X_H$  are independent of  $\tau$ .) As a result, Proposition 6 implies that, absent bill-shock regulation, firms will price discriminate without distorting allocations. Equilibrium without bill-shock regulation is efficient because competition ensures firms only want to charge high types a slightly higher markup and this can be achieved using only surprise penalty fees.<sup>19</sup>

#### 4.4 Consequences of bill-shock regulation

By comparing Propositions 4 and 7, Corollary 2 completes the first main result. The combination of surprise penalty fees and consumer inattention are socially valuable and bill-shock regulation is counterproductive whenever markets are fairly competitive.

**Corollary 2** *In a simple duopoly, if  $\tau > 0$  is sufficiently small then: bill-shock regulation would strictly decrease welfare and firm profits. Low types would lose while high types would win.*<sup>20</sup>

Corollary 2 follows by comparing Propositions 4 and 7. Absent bill-shock regulation, Proposition 7 shows equilibrium allocations are efficient for  $\tau > 0$  sufficiently small. In contrast, given bill-shock regulation, Proposition 4 implies equilibrium allocations are inefficient for all  $\tau > 0$ . Thus bill-shock regulation strictly lowers social welfare in fairly competitive markets. It does so because bill-shock regulation eliminates surprise penalty fees from firms' price-discrimination toolbox and forces firms to introduce quantity distortions on contract  $L$ .

To understand the distributional results, note that, by necessitating quantity distortions, bill-shock regulation makes price discrimination less profitable for firms. Thus firms also respond to

---

<sup>19</sup>Note that the efficiency result in Proposition 7 relies on several assumptions. If marginal costs were not constant, then a constant threshold strategy could not be efficient. If consumers were risk averse, penalty fees would have an inherent social cost. If some consumers were attentive, then those consumers would make inefficient consumption choices. Nevertheless, relaxing these assumptions slightly would only lead to a small deviation from efficiency because the model is continuous.

<sup>20</sup>Corollary 2 is also true for regulation banning penalty fees. See Online Appendix C.

bill-shock regulation by charging more similar markups: increasing the markup on contract  $L$  and reducing the markup on contract  $H$ . Low types are hurt both by the quantity distortion and the higher markup on contract  $L$  while high types benefit from the markup reduction on contract  $H$ . Firm market shares are unaffected in equilibrium, but profits are reduced because the loss from reducing markups on contract  $H$  exceed the gains from raising markups on contract  $L$  by a factor of  $H/L$ . This is because  $L$  types are more price-sensitive, so on the margin it is more expensive to raise markups on contract  $L$  in terms of market share.<sup>21</sup>

Finally, note that Corollary 2 does not extend beyond fairly-competitive markets. The impact of regulation becomes ambiguous when there is sufficient market power. Both surprise penalty fees and quantity distortions are useful tools for price discrimination. In some cases (including fairly-competitive markets) they are substitutes and regulation that eliminates surprise increases quantity distortions. In other cases they are complementary and the reverse is true:

**Corollary 3** *Let a monopolist serve consumers with zero outside option (ZOOM). Bill-shock regulation raises total welfare for some parameter values but lowers total welfare for others.*

The proof of Corollary 3 is by construction, applying Propositions 3 and 6 to the case in which: (1) high types' values are uniformly distributed between 0 and 1; (2) low types' values are uniformly distributed between 0 and  $1/2$  with probability  $3/4$  and are uniformly distributed between  $1/2$  and 1 with probability  $1/4$ . Then, for any  $c \in (0, 1)$ , the monopolist finds it strictly profitable to make penalty fees a surprise. Moreover, bill-shock regulation raises total welfare if  $c = 1/4$  but lowers total welfare if  $c = 1/2$ .

## 5 Biased Beliefs

Section 4 shows that consumer inattention makes surprise penalty fees an efficient tool for price discrimination. Corollary 2 concludes that, in fairly-competitive markets, bill-shock regulation reduces social welfare because it forces price-discriminating firms to substitute towards quantity distortions. This section explores an alternative role for surprise penalty fees: a tool for exploiting consumer bias. When consumers (who remain aware of their own inattention) underestimate their consumption of the add-on good or service, welfare implications for bill-shock regulation differ substantially. Bill-shock regulation may exacerbate or ameliorate allocative distortions created by

---

<sup>21</sup>Shifts in markups in each segment are already inversely weighted by shares of each segment  $\beta$  and  $(1 - \beta)$  since the shares reflect the cost of distorting that segment. Thus the difference in price sensitivity drives the difference in relative profit changes, rather than relative segment sizes.

biased beliefs depending on the size of marginal costs. However, the effect of first-order importance may be on surplus distribution rather than total welfare. Bill-shock regulation prevents *exploitation* of inattentive consumers:

**Definition 5** *A consumer is exploited if her average ex post utility is lower than her outside option.*<sup>22</sup>

## 5.1 Model

Return to the assumption in the benchmark model that consumers all have the same distribution of taste shocks  $F$ . Now, however, assume that consumers believe that the distribution is  $F^*$ , which is first-order stochastically dominated by  $F$  so that consumers underestimate their demand for the add-on service. Both  $F$  and  $F^*$  are continuous and strictly increasing on  $[0, 1]$  and the first-order stochastic dominance relationship is strict for some  $v \in (0, 1)$ .

A consumer's true expected gross utility from contracting with firm  $i$  at the contracting stage and making optimal consumption choices thereafter remains  $U^i = E[u(\mathbf{q}(\mathbf{v}; \mathbf{p}^i), \mathbf{v}) | F]$ . However, a consumer's perceived expected gross utility differs because expectations are taken with respect to consumer beliefs:  $U^{*i} = E[u(\mathbf{q}(\mathbf{v}; \mathbf{p}^i), \mathbf{v}) | F^*]$ . The fraction of consumers who buy from firm  $i$  depends on the perceived expected utility offered by firms rather than the true expected utilities:  $G(U^{*i}; U^{*-i})$ . Thus firm  $i$ 's expected profits are

$$\Pi^i = G(U^{*i}; U^{*-i}) E[P(\mathbf{q}(\mathbf{v}; \mathbf{p}^i), \mathbf{p}^i) - c(q_1(\mathbf{v}; \mathbf{p}^i) + q_2(\mathbf{v}; \mathbf{p}^i)) | F],$$

or rewritten in terms of true gross social surplus and consumers' true and perceived expected gross utilities:  $\Pi^i = G(U^{*i}; U^{*-i}) (S^i - U^i)$ .

To analyze the effect of bill-shock regulation, I begin by characterizing equilibrium pricing with bill-shock regulation, which is equivalent to all consumers being attentive. Afterwards, I continue by examining how pricing differs without bill-shock regulation.

## 5.2 Pricing with bill-shock regulation (Attentive Case)

If attentive consumers underestimate their demand for the service ex ante, firms have an incentive to set marginal charges above marginal cost, irrespective of competition (e.g. Grubb (2009)). This

---

<sup>22</sup>This is Eliaz and Spiegler's (2006) definition for an *exploitative contract*.

is reflected in Proposition 8, which characterizes pricing in the attentive case.<sup>23</sup>

**Proposition 8** *If attentive consumers underestimate demand, then the optimal contract is a two-part tariff ( $p_3 = 0, p_1 = p_2 = p$ ) with marginal price  $p = c + (F^*(p) - F(p)) / f(p)$  and profits*

$$\Pi = G(U^*) \left( v_0 - U^* + 2 \int_p^1 \left( v - c - \frac{F^*(v) - F(v)}{f(v)} \right) f(v) dv \right). \quad (13)$$

*No consumers are exploited and all transactions generate positive surplus. If  $F(c) < F^*(c)$  (bias is strict at  $p = c$ ) then  $p > c$  and allocations are inefficiently low.*

In the absence of inattention, bias distorts consumption downwards because a firm is limited in how much surplus it can extract ex ante through fixed fees by consumers' low estimate of their value for the service. The firm must wait until consumers draw high values and extract surplus through distortionary marginal charges. Nevertheless, there is no exploitation or surplus-reducing trade. Note that Proposition 8 implies that banning penalty fees is equivalent to bill-shock regulation because two-part tariffs are optimal when consumers are attentive.

### 5.3 Pricing without bill-shock regulation (Inattentive case)

The consumption threshold chosen by an inattentive consumer with biased beliefs satisfies

$$v^* = \frac{p_1 + p_2}{2} + p_3 (1 - F^*(v^*)), \quad (14)$$

which substitutes consumer beliefs in place of the true distribution of tastes in equation (6). As before, I focus on symmetric pricing  $p_1 = p_2 = \bar{p}$  and reframe the firm's problem in two ways. First, think of the firm choosing perceived expected-utility  $U^*$  so that the fixed fee  $p_0$  is given by

$$p_0 = -U^* + v_0 + 2 \int_{v^*}^1 v dF^*(v) - 2\bar{p}(1 - F^*(v^*)) - p_3 (1 - F^*(v^*))^2. \quad (15)$$

Second, think of the firm first choosing consumer threshold  $v^*$  and then choosing the best marginal prices  $\bar{p}$  and  $p_3$  which implement  $v^*$ . Using equations (14) and (15), firm profits can be written as a function of perceived expected-utility  $U^*$ , consumer threshold  $v^*$ , and penalty  $p_3$ :

$$\Pi = G(U^*) \left( v_0 - U^* + 2 \int_{v^*}^1 \left( v - c - \frac{F^*(v) - F(v)}{f(v)} \right) f(v) dv + p_3 (F^*(v^*) - F(v^*))^2 \right). \quad (16)$$

---

<sup>23</sup>Marginal pricing is the unit-demand analog of that characterized by Grubb (2009) for continuous demand and  $T = 1$ , repeated in each subperiod  $t \in \{1, 2\}$ .

Comparing equations (13) and (16) shows that the firm can make strictly higher profits by charging a surprise penalty fee to inattentive consumers than by selling to attentive consumers. Moreover, equation (16) shows that profits from inattentive consumers are increasing both in the surprise penalty fee and in the size of disagreement between firm and consumers about the probability consumers pay a penalty,  $(F^*(v^*) - F(v^*))^2$ . This is intuitive because the contract can be thought of as a bet to exploit the difference between firm and consumer beliefs. When consumers are inattentive, part of that bet is designed to exploit the difference in beliefs about the probability of paying a penalty. All else equal, profits from the bet increase in both the stakes of the bet, or the penalty fee, and the size of disagreement,  $(F^*(v^*) - F(v^*))^2$ .

For simplicity I impose a maximum penalty fee  $p^{\max} > 0$ , which implies that the optimal penalty fee is  $p^{\max}$ . (Many of the same economic forces identified in Remark 1 would also endogenously restrict penalty fees given biased beliefs.) It is worth pausing to point out why the bound on penalty fees is needed. When two risk-neutral parties have different prior beliefs about an publicly observable outcome, such as a coin flip, infinite bets are predicted. In this paper, when consumers are attentive, biased beliefs do not lead to infinite bets because the disagreement between parties is about the outcome of privately observed values  $v_t$ . Bets are limited in size by the fact that it must be incentive compatible for consumers to reveal values, including those they have bet against. When consumers are inattentive, however, this incentive constraint is relaxed. Consumers bet against receiving two high values and inattention means that they reveal their own losses (by purchasing twice) without realizing it. As discussed below, an important effect of bill-shock regulation is constraining the extent to which firms and consumers can bet over their difference in beliefs by reimposing the incentive constraint that limits the stakes of the bet.

Returning to the firm's pricing problem, it reduces to choosing  $U^*$  and  $v^*$  to maximize profits in equation (16) given  $p_3 = p^{\max}$ .<sup>24</sup> Moreover, the firm's objective only differs from that in the attentive case (equation (13)) by the additional term  $p_3 (F^*(v^*) - F(v^*))^2$ . Therefore, relative to the marginal price that is shown to be optimal for attentive consumers in Proposition 8, the firm's choice of  $v^*$  will be adjusted to increase the size of disagreement.

## 5.4 Consequences of bill-shock regulation

To understand the consequences of bill-shock regulation, begin by considering the firm's pricing problem for inattentive consumers from the point where the previous section finished. Having maximized the stakes of the penalty-fee bet by setting the maximum penalty fee, firms next have

---

<sup>24</sup>Given the positive penalty fee, equation (14) is sufficient for  $v^*$  to be incentive compatible.

an incentive to adjust consumers' threshold choice  $v^*$  to increase disagreement. This adjustment could increase or decrease quantity distortions relative to the attentive case, meaning that bill-shock regulation could be good or bad for welfare. In particular, for some intermediate marginal cost, maximizing disagreement ameliorates inefficiency and bill-shock regulation reduces welfare. However, if marginal costs are close to zero or one, then maximizing disagreement entails increased inefficiency and bill-shock regulation increases welfare. To state the result formally, I first parameterize consumers' degree of bias with the parameter  $\gamma$ : Let  $F$  and  $\hat{F}$  have full support with continuous densities on  $[0, 1]$  that cross finitely many times,  $F < \hat{F}$  for all  $v \in (0, 1)$ , and  $F^* = \gamma\hat{F} + (1 - \gamma)F$  for some  $\gamma \in (0, 1]$ . Consumers underestimate demand for any  $\gamma > 0$  but consumers' bias goes to zero as  $\gamma$  goes to zero.

**Proposition 9** *Consider either a monopolist satisfying ZOOM or a simple duopoly. If bias is sufficiently small ( $\gamma$  is sufficiently close to zero) and the maximum penalty fee  $p^{\max}$  is sufficiently large then: (1) When marginal cost  $c$  is close to zero or one, bill-shock regulation strictly improves welfare. (2) There exists an intermediate marginal cost  $c \in (0, 1)$  for which bill-shock regulation strictly decreases welfare. (3) Banning penalty fees has the same effect as bill-shock regulation.*

Note that Proposition 9 contrasts sharply with the benchmark result in Proposition 2, that regulation does not affect welfare when firms offer a single contract to inattentive but unbiased consumers. Thus the welfare effects identified in Proposition 9 arise entirely from the interaction between inattention and bias. In particular, only when consumers are inattentive do firms have an incentive to maximize disagreement over the probability of paying a penalty.

For intuition behind part (1) of Proposition 9, note that there is no disagreement between firms and consumers when  $v^*$  is zero or one. In these cases all parties either agree that consumers always purchase and pay a penalty ( $v^* = 0$ ) or agree that consumers never purchase and never pay a penalty ( $v^* = 1$ ). Thus, disagreement is increasing in  $v^*$  near zero but decreasing in  $v^*$  near one. For marginal costs near zero this means that  $v^*$  is distorted further above marginal cost than when consumers are attentive, thereby exacerbating underconsumption. For marginal cost near one this means that  $v^*$  is distorted downwards below marginal cost, replacing mild underconsumption when consumers are attentive with more severe overconsumption when consumers are inattentive. In both cases, bill-shock regulation reduces quantity distortions and increases welfare. Part (2) of Proposition 9 follows because when marginal cost is near the point of maximal disagreement, maximizing disagreement means moving  $v^*$  towards marginal cost and ameliorating underconsumption. In this case, bill-shock regulation exacerbates underconsumption and reduces welfare.

Although bill-shock regulation may increase or decrease total welfare depending on the level of marginal cost, clearer predictions can be made about its effect on the distribution of surplus between firms and consumers. This is because bill-shock regulation undermines the profitability of penalty fees so much that firms stop using them. Without bill-shock regulation, firms can use surprise penalty fees to charge a total marginal price for a second unit ( $p_2 + p_3$ ) that is far above a consumer's valuation because inattentive consumers buy when their value exceeds the expected marginal price. In contrast, such a high penalty fee would not be paid under bill-shock regulation because consumers would only buy when their value exceeded the realized marginal price of the second unit. Thus bill-shock regulation prevents firms from profitably exploiting different beliefs about the likelihood of paying a penalty.

Whether or not the loss of penalty fee revenue due to bill-shock regulation hurts firms and helps consumers depends on market structure. For a monopolist, bill-shock regulation directly lowers profits so that consumers can benefit even if total surplus falls. If duopolists compete on fixed fees, however, a reduction in penalty fee revenue can be offset by higher fixed fees in equilibrium, leaving firm profits unchanged:

**Proposition 10** (1) Monopoly: Assume the firm is a zero-outside-option monopolist. If the upper bound on penalty fees,  $p^{\max}$ , is sufficiently large then inattentive consumers are exploited. Bill-shock regulation eliminates exploitation by shifting surplus from the firm to consumers.

(2) Simple Duopoly: If the add-on is socially valuable ( $c < 1$ ) and the market is sufficiently competitive ( $\tau < (2/3)(v_0 + \int_c^1 (v - c) f^*(v) dv)$ ) then there is full-market-coverage and firm profits equal transport cost  $\tau$  independent of bill-shock regulation.

Proposition 10 suggests that bill-shock regulation could be beneficial for preventing consumer exploitation despite ambiguous effects on social welfare. However, Proposition 10 suggests that this is only true for a monopoly. Even if large revenues are earned on surprise penalty fees when consumers are inattentive, Proposition 10 shows that, under imperfect competition, these are rebated back to consumers through lower fixed fees so that firms earn identical markups to those charged under bill-shock regulation. Therefore consumers are residual claimants of social surplus with respect to bill-shock regulation.

The fact that competition can protect consumers from exploitation is intuitive. Nevertheless, part (2) of Proposition 10 relies on two assumptions that are not always realistic. The first is explicit, that the add-on is socially valuable ( $c < 1$ ). The second assumption is implicit, that firms can charge arbitrarily negative fixed fees and thereby offset penalty fee revenue, however large. If either assumption is relaxed, then bill-shock regulation can stop consumer exploitation and lower

firm profits even in fairly competitive markets. I discuss both extensions below.

First, suppose that the add-on service is not socially valuable and that  $c$  is above one. (The add-on has no social value if  $c > 1$  because  $v_t \leq 1$ .) In this case all product sales are inefficient. Yet because inattentive consumers underestimate their likely values for the product, sales can still take place. If  $p^{\max}$  is sufficiently large, then without bill-shock regulation firms will market the add-on solely to profit from surprise penalty fees, which consumers underestimate the likelihood of paying. This is true even in fairly competitive markets. In contrast, Proposition 8 implies that bill-shock regulation would lead to efficient shut down of the add-on market, eliminating both consumer exploitation and firm profits:

**Proposition 11** *Assume uniform Hotelling duopoly with an add-on that has no social value ( $c > 1$ ). For any fixed transport cost  $\tau > 0$ , if the maximum penalty fee  $p^{\max}$  is sufficiently large then: bill-shock regulation eliminates consumer exploitation and raises total welfare by shutting down the add-on market, which otherwise operates solely for firms to profit from surprise penalty fees.*

Finally, the assumption that firms can charge substantially negative fixed fees is often unrealistic. Thus Section 6 analyzes an extension that imposes a no-arbitrage condition that limits payments to be nonnegative.

## 6 Extension: Biased Beliefs and No-free-lunch

Proposition 10 suggests that, under monopoly, the effects of inattention and bill-shock regulation on total welfare may be less important than their effects on the distribution of surplus. However, Proposition 10 also shows that duopoly profits are invariant to bill-shock regulation: While bill-shock regulation limits penalty-fee revenue, in equilibrium these revenue losses are exactly off-set by increases in fixed fees. While this competitive result is intuitive, it is not robust because it relies on the unrealistic assumption that firms can charge substantially negative fixed fees.

To better understand the effect of bill-shock regulation on surplus distribution under imperfect competition with inattentive and biased consumers, I adapt the model of Section 5 and endogenously restrict penalty fees by imposing a no-arbitrage condition that I call the *no-free-lunch* (NFL) constraint. The NFL constraint arises endogenously if there exists a large pool of attentive potential-customers (or potential customers with a very low cost  $k$  of paying attention) with zero value for the service. Such consumers restrict the payment function to be nonnegative: If the payment function were negative at some allocation, such consumers would buy exactly the right quantities to earn the subsidy.

**Definition 6** *The no-free-lunch (NFL) constraint restricts consumer payments to be nonnegative at all allocations:  $p_0 \geq 0$ ,  $p_0 + p_1 \geq 0$ ,  $p_0 + p_2 \geq 0$ , and  $p_0 + p_1 + p_2 + p_3 \geq 0$ .*<sup>25</sup>

Finally, to focus on distributional issues, assume taste shocks have a Bernoulli distribution:  $v_t$  are independent and are equal to one with probability  $\alpha$  and zero otherwise.<sup>26</sup> Consumers underestimate their demand and believe that  $v_t$  equals one with probability  $\alpha' < \alpha$ . Also assume  $c \in [0, 1)$ . By introducing Bernoulli taste shocks, I ensure that firms induce first-best allocations with or without bill-shock regulation so that regulation only effects the distribution of surplus:

**Lemma 1** *Given Bernoulli taste shocks, attentive or inattentive consumers who underestimate demand ( $\alpha' < \alpha$ ),  $c \in (0, 1)$ , and the NFL constraint, firms set prices which induce the efficient allocation: consumers buy if and only if  $v_t = 1$ .*

As before, the first step to solve for equilibrium duopoly pricing is to write down and solve a monopolist's problem. This initial step, for both attentive and inattentive cases, is made in Online Appendix B where Propositions 14 and 15 characterize optimal monopoly pricing. In this section, I move directly to describing equilibrium duopoly pricing, beginning with the attentive case corresponding to the outcome with bill-shock regulation.

## 6.1 Pricing with bill-shock regulation Regulation (Attentive case)

Absent the NFL constraint, Hotelling duopolists would each offer a two-part tariff with marginal price of one to maximally exploit consumer bias. They would compete on fixed fees, which would be set to earn a markup of  $\tau$  in equilibrium. For  $\tau \geq 2\alpha(1 - c)$ , this is exactly the outcome with the NFL constraint. However, in more competitive markets with  $\tau < 2\alpha(1 - c)$ , this would require charging a negative fixed fee. When  $\tau$  falls below this threshold and the fixed fee has already been reduced to zero, firms must lower marginal fees if they wish to lower their markups. In more competitive markets competition shifts first to base marginal charges and finally to penalty fees. This progressively softens price competition and firms charge markups above  $\tau$ . Thus increasing competition is partially mitigated by reduced consumer price sensitivity.

Consider a firm's choice of marginal fees  $\{p_1, p_2, p_4\}$ , where  $p_4 = p_2 + p_3$  is the marginal price for a second-period purchase conditional on a first-period purchase. When lowering prices, a firm

---

<sup>25</sup>Similar results would follow from a negative lower-bound rather than zero lower-bound on payments. In fact, because fixed costs have been normalized to zero, the normalized lower bound should be equal to the negative of fixed costs. (The presence of high fixed-costs allows firms to subsidize consumers without making payments to consumers. Hardware discounts with cellular-phone-service contracts are an example.)

<sup>26</sup>Consumers may or may not want an extra unit but always have the same value when they do want one.

prefers to first cut those fees to which consumers are most price-sensitive. Consumers are less price-sensitive to  $p_1$  than to the fixed fee because they underestimate the chance of paying  $p_1$  by a factor  $\alpha'/\alpha$ . The reduced price-sensitivity is compounded for  $p_4$ , since consumers underestimate the chance of making two purchases by  $(\alpha'/\alpha)^2$ . However, the reduce price-sensitivity is mitigated for  $p_2$ . While consumers underestimate the chance of demanding a unit in the second period, they overestimate the chance that  $p_2$  is the relevant second-period price because they underestimate the likelihood of an initial purchase triggering a penalty fee. Thus, once fixed fees are reduce to zero, a firm reducing prices would like to first cut  $p_2$ , second cut  $p_1$ , and lastly cut  $p_4$ . (In fact a firm cutting  $p_2$  must simultaneously reduce  $p_1$  a proportion  $\alpha'$  as much to satisfy the incentive constraint  $v_1^* \leq 1$ .)

This program of price reduction in response to increasing competition leads to four qualitative pricing regions depicted in the top panel of Figure 1, which plots the equilibrium markup as a function of the transportation cost  $\tau$ .<sup>27</sup> Four dashed lines show the markups relevant for the four possible pricing regions. The solid bold line shows the equilibrium markup, which is increasing in  $\tau$  within pricing regions but constant between regions. Starting at the right side of the figure, and working leftward as  $\tau$  falls and competition increases, equilibrium begins in region 1 where firms compete on fixed fees and markups equal  $\tau$ . After fixed fees have reached zero, consumers become discontinuously less price-sensitive and markups are temporarily flat until equilibrium transitions to region 2 where firms begin competing on  $p_1$  and  $p_2$  and markups equal  $(\alpha/\alpha')(1 - \alpha + \alpha')\tau$ . As transportation costs fall, equilibrium continues to transition through the four competitive regions so that markups are weakly decreasing in absolute levels but weakly increasing as a proportion of transportation costs as competition shifts towards fees to which consumers are less and less price-sensitive. In region 3 firms compete on  $p_1$  and markups equal  $(\alpha/\alpha')\tau$  while in region 4 firms compete on the penalty fee and markups equal  $(\alpha/\alpha')^2\tau$ .

**Proposition 12** *Assume duopoly on a uniform Hotelling line, the NFL constraint, Bernoulli taste shocks, attentive consumers who underestimate demand ( $\alpha' < \alpha$ ) and  $c \in [0, 1)$ . Let base-good value  $v_0$  be sufficiently large for strict full-market-coverage.<sup>28</sup> There are four competitive regions*

---

<sup>27</sup>Note that the figure is shown for full market-coverage and  $c = 0$ . If marginal cost is strictly positive then some of the four competitive regions may not be relevant. For instance, if  $c > \alpha/2$  then region 4 is never reached. Under perfect competition ( $\tau = 0$ ) expected markups will always be zero, but for  $c > \alpha/2$  this means the penalty fee alone is insufficient for the firm to break even so other fees must remain positive.

over which markups are proportional to  $\tau$ . Markups are constant between regions.

<i>Region</i>	$\tau_{\min}$	$\tau_{\max}$	<i>markup <math>\mu</math></i>	<i>Competition</i>
1	$2\alpha(1-c)$		$\tau$	<i>fixed fees</i>
2	$\frac{2\alpha'(1-c)}{1-\alpha+\alpha'} - \alpha'$	$\frac{2\alpha'(1-c)}{1-\alpha+\alpha'}$	$(\alpha/\alpha')(1-\alpha+\alpha')\tau$	<i>base marginal charges</i>
3	$\alpha'(\alpha-2c)$	$\alpha'(\alpha-2c) + \alpha'(1-\alpha')$	$(\alpha/\alpha')\tau$	<i>base marginal charges</i>
4	0	$(\alpha'^2/\alpha)(\alpha-2c)$	$(\alpha/\alpha')^2\tau$	<i>penalty fees</i>

Duopoly profits equal the markup and consumers' true expected gross utility is  $U = S^{FB} - \mu \geq 0$ . No consumers are exploited.

An interesting feature is that although consumers are always made better off by increased competition, the cost of their bias is not decreasing monotonically with competition. Between pricing regions where markups are constant, increasing competition increases the gap between the markup and  $\tau$ . As the markup would be  $\tau$  if consumers were unbiased, this means the cost of their bias is locally increasing in competition. This contradicts a common intuition that increased competition reduces the importance of policy interventions to address consumer biases.

## 6.2 Pricing without bill-shock regulation (Inattentive case)

Proposition 12 shows that the NFL constraint softens competition when consumers are biased, leading to markups above  $\tau$  by forcing firms to compete on marginal fees rather than the fixed fee. Proposition 13 shows the same is true with inattentive consumers, but the magnitude of the effect is typically higher because inattention allows firms to raise penalty fees. As a result, Corollary 4 shows that bill-shock regulation typically helps consumers by intensifying competition.

To state the proposition, define  $Y \equiv \frac{(\alpha-\alpha')^2}{\alpha'(1-\alpha')}$ .

**Proposition 13** *Assume duopoly on a uniform Hotelling line, the NFL constraint, Bernoulli taste shocks, consumers who underestimate demand ( $\alpha' < \alpha$ ) and  $c \in [0, 1)$ . Let base-good value  $v_0$  be sufficiently large for strict full-market-coverage. Firms prefer to keep penalty fees a surprise, a strict preference for  $\tau > (\alpha')^2(\alpha - 2c)/\alpha$ . There are two competitive regions over which markups*

---

<sup>28</sup>See footnote 16.

are proportional to  $\tau$ . Markups are constant between regions.

<i>Region</i>	$\tau_{\min}$	$\tau_{\max}$	<i>markup <math>\mu</math></i>	<i>Competition</i>
1	$(2\alpha(1-c) + Y) / (1+Y) - \alpha'$		$(1+Y)\tau$	<i>all fees</i>
2	0	$(\alpha'/\alpha)(\alpha - 2c\alpha')$	$(\alpha/\alpha')^2\tau$	<i>penalty fees</i>

Duopoly profits equal the markup and consumers' true expected gross utility is  $U = S^{FB} - \mu$ .

The bottom panel of Figure 1 illustrates Proposition 13 by plotting the equilibrium markup as a function of the transportation cost  $\tau$ . Dashed lines show the markups relevant for the two possible pricing-regions as well as the markup  $\tau$  that would prevail in the absence of bias or the NFL constraint. The solid bold line shows the equilibrium markup, which is increasing in  $\tau$  within pricing regions but constant between regions. Starting at the right side of the figure, and working leftward as  $\tau$  falls and competition increases, equilibrium begins in region 1 where firms compete on a mixture of fees and markups equal  $(1+Y)\tau$ . In this region, penalty fees exceed  $1/\alpha'$ , which requires negative base marginal-charges to ensure expected marginal price does not exceed 1. Negative base marginal-charges in turn require positive fixed fees to satisfy NFL. Once penalty fees fall to  $1/\alpha'$ , all other fees are zero and firms compete on penalty fees alone. As a result, consumers become discontinuously less price-sensitive and markups are temporarily flat until equilibrium transitions to region 2 where markups equal  $(\alpha/\alpha')^2\tau$ .

### 6.3 Consequences of bill-shock regulation

A sufficient condition for the full-market-coverage assumption in Propositions 12 and 13 is  $\tau < 2v_0/3$ . Comparing Propositions 12 and 13 for  $\tau \in (0, 2v_0/3)$  uncovers the effects of bill-shock regulation under competition. Without bill-shock regulation, expected marginal-prices must be no-higher than one which allows penalty fees to be as high as  $1/\alpha'$  when base marginal-charges are zero and higher when base marginal-charges are negative. The important effect of bill-shock regulation is that it means implementing the efficient allocation (as is optimal) requires every marginal price be at most one, and hence penalty fees be no higher than one. Holding the level of bias fixed (and  $c < \alpha/2$ ), sufficient competition ( $\tau \leq (\alpha')^2(\alpha - 2c)/\alpha$ ) implies that this does not matter because firms choose to offer sufficiently high perceived-expected-utility levels that penalty fees must be less than one. As a result, firms offer the same contract and markup regardless of whether or not bill-shock regulation is implemented. For any higher level of market power ( $\tau > (\alpha')^2(\alpha - 2c)/\alpha$ ), however, bill-shock regulation does constrain firms' use of penalty fees. Typically this shifts competition towards fees to which consumers are more price-sensitive, thereby

intensifying competition and lowering firm markups. This is always the case for *severe bias*<sup>29</sup> ( $\alpha'/\alpha < \max\{1/2, (2\alpha - 1)/\alpha^2\}$ ) but the reverse can be true for intermediate values of  $\tau$  given *mild bias* ( $\alpha'/\alpha > \max\{1/2, (2\alpha - 1)/\alpha^2\}$ ) as made precise in parts 2 and 3 of Corollary 4.<sup>30</sup> The comparison is illustrated for severe and mild biases in top and bottom panels of Figure 2. Part 1 of Corollary 4 holds  $\tau > 0$  fixed and shows that sufficiently large bias leads to arbitrarily high markups and consumer exploitation. Thus while markups are unambiguously reduced for severe bias, for sufficiently high bias this reduction in markups also means an end to consumer exploitation.

**Corollary 4** *Assume duopoly on a uniform Hotelling line, the NFL constraint, Bernoulli taste shocks, inattentive consumers who underestimate demand ( $\alpha' < \alpha$ ), and  $c \in [0, 1)$ . Let  $\tau < (2/3)v_0$ . The market will be fully covered and allocations will be first best with or without bill-shock regulation.*

1. *For fixed  $\tau > 0$ , if bias is sufficiently large ( $\alpha'/\alpha$  is sufficiently small) then all consumers are exploited. Bill-shock regulation increases competition, strictly reduces markups, and eliminates consumer exploitation.*
2. *If bias is severe ( $\alpha'/\alpha < \max\{1/2, (2\alpha - 1)/\alpha^2\}$ ) then bill-shock regulation weakly reduces markups for all  $\tau \geq 0$ , and strictly reduces markups for all  $\tau > \max\{(\alpha')^2(\alpha - 2c)/\alpha, 0\}$  (which is for all  $\tau > 0$  if  $c \geq \alpha/2$ ).*
3. *If bias is mild ( $\alpha'/\alpha > \max\{1/2, (2\alpha - 1)/\alpha^2\}$ ) then bill-shock regulation effects markups as described for severe bias except for intermediate  $\tau \in [\tau_1, \tau_2]$ , where  $\tau_1 = (1 - \alpha + \alpha') / (\alpha - 2c\alpha')$  and  $\tau_2 = 2\alpha(1 - c) / (1 + Y)$ . For  $\tau \in (\tau_1, \tau_2)$ , bill-shock regulation strictly increases markups.*

Proposition 10 and Corollary 4 capture the second main result in the paper: combined with biased beliefs, inattention can cause consumer exploitation which is eliminated by bill-shock regulation. In the monopoly setting this is a direct result of the fact that bill-shock regulation constrains the size of penalty fees - precisely those fees which consumers most underestimate the chance of paying. In a competitive setting the result is more indirect. Absent additional constraints on prices, total markups would equal  $\tau$  independent of the fraction earned from penalty fees. However, the NFL constraint ensures that fairly-competitive firms compete on marginal charges rather than fixed fees. This leads to markups above  $\tau$  because consumers are less price-sensitive to marginal charges,

---

<sup>29</sup>The term *severe bias* is somewhat misleading. As  $\alpha$  approaches 1, a belief  $\alpha' < \alpha$  arbitrarily close to  $\alpha$  will satisfy the condition for severe bias.

<sup>30</sup>When bias is mild and  $\tau \in (\tau_1, \tau_2)$ , bill-shock regulation strictly increases equilibrium markups. In this case although keeping penalty fees a surprise is individually optimal for each firm, as an industry group firms would favor bill-shock regulation regulation.

which they underestimate the likelihood of paying. Bill-shock regulation limits the extent that this competition is over penalty fees, and typically forces firms to compete on more salient base marginal charges intensifying competition and protecting consumers. The online appendix shows that banning penalty fees has a similar effect. (Note that eliminating the underlying bias would be even better, making all fees equally salient and lowering markups to  $\tau$ . However, debiasing consumers completely is likely to be difficult and costly relative to implementing bill-shock regulation.)

## 7 Policy Applications

### 7.1 FCC's Proposed Bill-Shock Regulation

US cellular-phone customers are typically charged steep penalty fees for exceeding usage allowances, and the variation in usage allowances across calling plans is an essential instrument for encouraging consumers to self-select into different calling plans. On October 14th, 2010, the FCC proposed bill-shock regulation that would require carriers to notify customers, via voice or text alerts, when they are about to exceed plan limits and begin incurring overage charges (FCC 2010). The FCC's proposed bill-shock regulation had strong support from consumer groups but was opposed by major cellular carriers (Genachowski 2010, Deloney et al. 2011, Wyatt 2010, Stross 2011).<sup>31</sup> A year later, on October 17th, 2011, it was announced that cellular carriers will voluntarily begin providing such usage alerts by April 2013 and in return the FCC will drop bill-shock regulation (CTIA - The Wireless Association 2011).

The price-discrimination model in Section 4 provides an explanation both for carriers' use of surprise penalty fees on typical cellular contracts and for carriers' opposition to proposed bill-shock regulation. If one believes that cellular-phone customers are unbiased and the cellular market is sufficiently competitive, then Corollary 2 implies that the FCC's recent bill-shock agreement will be counterproductive and lower social welfare. Moreover, consumer groups' strong support for the agreement would be misplaced, since it would harm some consumers. (Their support is nevertheless understandable since the agreement would be unambiguously beneficial to all consumers but for the resulting endogenous price changes predicted by the model.)

There are two caveats to this criticism of the FCC's bill-shock agreement.<sup>32</sup> First, while Corollary 2 assumes a fairly-competitive market, the Department of Justice has argued that cellular

---

<sup>31</sup>See footnote 2.

<sup>32</sup>A third caveat is that the regulation would apply to fees beyond overage charges such as roaming fees which are typically the same across calling plans, and hence not used for price-discrimination purposes or relevant to this theoretical argument. Roaming charges were the target of recently adopted bill-shock regulation in the EU.

carriers do have substantial market power. (The Department of Justice made this argument in 2011 when it sued to block AT&T's proposed acquisition of T-Mobile (Justice 2011).) As illustrated by Corollary 3, sufficient market power means that bill-shock regulation could increase or decrease social surplus when consumers are unbiased. Second, Grubb (2009) and Grubb and Osborne (2012) show that cellular-phone customers have biased beliefs about their likely usage and in particular are overconfident. Overconfidence, like demand underestimation, causes consumers to underestimate the likelihood of paying penalties. Proposition 9 shows that such biases in beliefs provide a second reason (in addition to market power) that bill-shock regulation could raise rather than lower social surplus.

Note that US cellular phone companies do compete on contracts' fixed fees, which are typically \$40 per month or more. Moreover, equipment charges and marginal charges are typically nonnegative. Thus the no-free-lunch constraint is not binding in this market and hence, following Proposition 10, bill-shock regulation is unlikely to be substantially better at improving consumer surplus than it is at increasing social surplus.

In sum, the welfare impact of bill-shock regulation is ambiguous but there is a clear reason that it could be socially harmful. Thus, implementing the FCC's bill-shock agreement may lower not just firm profits but also lower total welfare and hurt some consumers. To resolve the theoretical ambiguity, complementary empirical work by Grubb and Osborne (2012) estimates a structural demand model of consumers' contract and calling choices using a panel of cellular billing data. Grubb and Osborne's (2012) counterfactual simulation predicts that bill-shock regulation will lower social welfare and consumer surplus by about \$2 per consumer annually while slightly reducing industry profits. In contrast, related work by Jiang (2011) predicts that bill-shock regulation will raise consumer surplus by about \$5 per household annually and increase industry profits by \$4 billion annually. (The difference between these empirical results may be due to a number of modeling differences. For instance, in contrast to Jiang (2011), Grubb and Osborne (2012) allow consumer beliefs to be biased and model consumers' endogenous change in calling behavior in response to information in bill-shock alerts.)

## 7.2 Overdraft Fees

Turning to a second application, consider overdraft fees: in 2009, US bank overdraft fee revenues from ATM and one-time debit-card transactions were \$20 billion (Martin 2010). Prior to the Fed's adoption of an opt-in rule, Bank of America and other banks charged high (often \$35) overdraft fees on debit and ATM transactions without notifying customers at the point of sale. When the

Fed proposed opt-in regulation, banks opposed it.<sup>33</sup> Nevertheless, effective August 15, 2010 (July 1, 2010 for new accounts) new Federal Reserve Board rules “prohibit financial institutions from charging consumers fees for paying overdrafts on automated teller machine (ATM) and one-time debit-card transactions, unless a consumer consents, or opts-in, to the overdraft service for those types of transactions” (Federal Reserve Board 2009a).

In response to opt-in regulation, Bank of America chose to stop offering overdraft protection on debit-card transactions, despite the fact that Bank of America is estimated to have earned \$2.2 Billion from ATM and debit-card-transaction overdraft fees in 2009 (Sidel and Fitzpatrick 2010). Other major banks have been accused of responding with deceptive marketing campaigns to induce opt-in. For instance, customers filed a federal class-action lawsuit against JPMorgan Chase for such bad behavior in August 2010 (Dinzeo 2010, McCune, Wright, Arevalo and Kim 2010). More broadly, the Consumer Financial Protection Bureau (CFPB) announced in February 2012 that it will be investigating reports of such misleading marketing (Wyatt 2012).

The model used throughout the paper is stylized and fits the overdraft application imperfectly. In particular, while quantity is one-dimensional in the model, overdraft fees depend on both dollars spent and the number of transactions. Overdraft fees are only triggered once dollars spent exceed the account balance, but are then charged on a per-transaction basis. Moreover, marginal costs are likely increasing rather than constant. Nevertheless, the model captures important features of the setting.

For instance, consider the model of biased beliefs in Section 6 in which values  $v_t$  are one with probability  $\alpha$  and zero otherwise but consumers underestimate their demand, believing  $v_t$  is high with probability  $\alpha' < \alpha$ . We can interpret  $\alpha$  as the probability a consumer wishes to make a purchase with her debit card. Then  $v = 1$  is her value for making the purchase with her debit card, rather than an alternative such as a credit card. Consumers underestimate the likelihood of purchases to be  $\alpha'$  because they are partially-naive beta-delta discounters who not only undersave and overspend due to time inconsistency (Laibson 1997) but also underestimate how much they spend due to partial naivete (O'Donoghue and Rabin 2001).<sup>34</sup>

Do the results in Section 4 imply that the combination of consumer inattention and overdraft fees could be socially valuable by making price discrimination by banks less distortionary? In fact they do not apply. While banks offer different types of checking accounts, prior to the regulation

---

<sup>33</sup>Prior to regulating overdraft fees, the Federal Reserve solicited public comment. Industry commenters sought to undermine the regulation in every possible way. For instance “industry commenters...urged the Board to permit institutions to vary the account terms...for consumers who do not opt in [to overdraft protection]” (Federal Reserve Board 2009b). Clearly banks wanted to be able to make declining overdraft protection an expensive account feature.

<sup>34</sup>This bias is consistent with Ausubel (1991) and Ausubel and Shui's (2005) findings for credit card spending.

banks typically charged the same overdraft fees on all accounts (e.g. Bank of America (2010)). Thus heterogeneity in expectations of overdraft usage is typically not an important dimension of self-selection across checking accounts.

Neither the benchmark model nor Section 4’s model of price discrimination explain banks’ widespread use of overdraft fees, failure to notify consumers at the point-of-sale, or aversion to opt-in regulation. A more compelling explanation for these facts is that consumers underestimate the incidence of overdraft fees. Given such bias, Proposition 9 shows that bill-shock regulation could raise or lower social surplus depending on the level of marginal costs. However, the more important effect of regulation may be to shift surplus from banks to consumers.

Interestingly, Proposition 11 shows that for products with no social value ( $c > 1$ ) bill-shock regulation would end consumer exploitation and increase social surplus by shutting down the market for the add-on. This possibility is an intriguing explanation for Bank of America’s choice to stop a service which had been earning an estimated \$2.2 billion per year and JPMorgan Chase’s choice to resort to deceptive marketing tactics to encourage opt-in for overdraft coverage of ATM and debit-card transactions.<sup>35</sup>

Even if overdraft coverage for ATM and debit-card transactions is socially valuable, there is still reason to believe that bill-shock regulation would help consumers. The recent rise in the use of overdraft fees has coincided with the rise of *free* checking accounts (Burhouse, Cashman, Cordeiro, Critchfield, Lee, Pawelski and Samolyk 2008, Stango and Zinman 2012). As the Deputy Director of the CFPB put it, “With these free checking accounts, much of the costs to the consumer were buried in overdraft fees” (Date 2011). This suggests that a nonnegative pricing constraint has been binding in the industry and that results that assume the constraint (Corollary 4) will be a better guide to consumer protection than those which do not (Proposition 10), Jamie Dimon’s claims to the contrary notwithstanding (see footnote 6).

Thus Corollary 4 implies that bill-shock regulation could limit overdraft revenues, forcing banks to raise and compete on more salient fees, thereby intensifying competition, lowering markups, and ending consumer exploitation. Corollary 4 does not actually predict that bill-shock regulation would increase fixed fees.<sup>36</sup> However, it is worth noting that this would be the prediction under severe bias and intermediate market power if firms were further restricted to charging nonnegative

---

<sup>35</sup>An alternative explanation for Bank of America’s choice is regulatory threat.

<sup>36</sup>This is because the model predicts that, absent regulation, firms may raise fixed fees in order to charge negative base marginal charges without violating the no-free-lunch constraint. In turn, negative base marginal charges allow firms to raise surprise penalty fees while keeping expected marginal price low. Thus, given intermediate market power and the no-free-lunch constraint, bill-shock regulation will eliminate negative base marginal charges while stiffening competition and lowering both fixed fees and markups.

marginal prices as well as nonnegative total prices. (Moreover, checking account pricing typically satisfies this stricter constraint.)

Consistent with these results, between 2009 and 2011 the fraction of banks with at least \$50 billion in assets offering free checking fell from 96% to 35% (Aspan 2011). This fall in free checking coincides with the introduction of the overdraft opt-in requirement in 2010 and a debit card interchange fee cap in 2011. While the reduction in free checking is (somewhat counter-intuitively) a promising sign that recent regulation may be stiffening competition, the change in overdraft revenues is not so promising. Analysts initially predicted that opt-in regulation would dramatically reduce overdraft-fee revenue (Campbell 2009). In fact, opt-in rates have been high (75%) and overdraft-fee revenue has been relatively stable (Benoit 2010).

High opt-in rates for overdraft fees may be due to deceptive marketing practices currently under investigation (Wyatt 2012). An alternative explanation, however, is that the opt-in regulation adopted by the Fed is weaker than bill-shock regulation. If (as this paper assumes) consumers are aware of their own inattention, then opt-in regulation should have similar effects to bill-shock regulation: It should limit overdraft fees and consumer exploitation because consumers will opt-out if overdraft fees are too high. However, consumers may be naive about their own inattention, believing themselves to be attentive. A consumer who believes himself to be attentive should always opt-in.

Whether due to deceptive marketing or naivete about inattention, recent experience with opt-in regulation shows that it is a poor substitute for bill-shock regulation. The CFPB apparently believes that the current opt-in regulation is inadequate and has proposed that a “penalty fee box” should appear on checking account statements detailing overdraft fees charged during the month (Wyatt 2012). While this helps alert consumers when they have been charged overdraft fees, it still falls short of bill-shock regulation which would help them avoid such fees (and associated \$40 cups of coffee) at the point of sale.

## 8 Conclusion

Bill-shock regulation can help inattentive consumers avoid surprise penalty fees. While this is good for consumers holding prices fixed (and hence may attract the support of consumer groups), it is essential for policy evaluation to incorporate firms’ pricing response.

If unbiased consumers with heterogeneous forecasts of their future demand are inattentive, surprise penalty fees become a useful tool for price discrimination. The combination of inattention and surprise penalty fees are socially valuable when firms view them to be a substitute for inefficient

quantity distortions. This is always the case in fairly-competitive markets, but may not be true when firms have sufficient market power. Thus, in fairly-competitive markets, bill-shock regulation will be socially harmful because firms will continue to price discriminate but they will be forced to impose greater allocative inefficiencies to do so.

When consumers underestimate their future demand, the combination of consumer inattention and surprise penalty fees can be highly profitable for firms. The bias in beliefs make high penalty fees an attractive way for firms and consumers to take opposing sides of a bet. Inattention relaxes incentive constraints that would otherwise limit the size of penalty fees and the resulting bet. When firms compete on fixed fees to sell a socially valuable product, high revenues from surprise penalty fees are off-set in equilibrium by lower fixed fees. In this context, bill-shock regulation is superfluous to competition for protecting consumers from exploitation.

However, inattention and surprise penalty fees can enable firms to profit from selling a product with no social value. In these cases, consumers are exploited in the sense that doing business with the firm makes them worse off. Moreover, in some markets firms do not compete on fixed fees because they are zero and can't be lowered. In such cases, regulation can protect consumers from exploitation by shifting competition away from penalty fees towards fixed fees or base marginal charges to which consumers are more price-sensitive, thereby intensifying competition.

The results suggest that regulators should require bill-shock alerts for services such as overdraft protection that are not differentially priced to sort consumers across contracts. This is particularly true because of the prevalence of free checking accounts in consumer banking, which implies that competition is not sufficient to protect consumers from exploitation. However, regulators should be more cautious about the FCC's proposed bill-shock regulation and other applications for which penalty fees do help sort consumers across contracts and fixed fees are positive. The limited impact of the Fed's opt-in regulation on overdraft-fee revenues suggests it is a poor substitute for bill-shock regulation, perhaps because consumers are unaware of their own inattention. Such naivete presents an interesting avenue for future work.

## 9 Figures

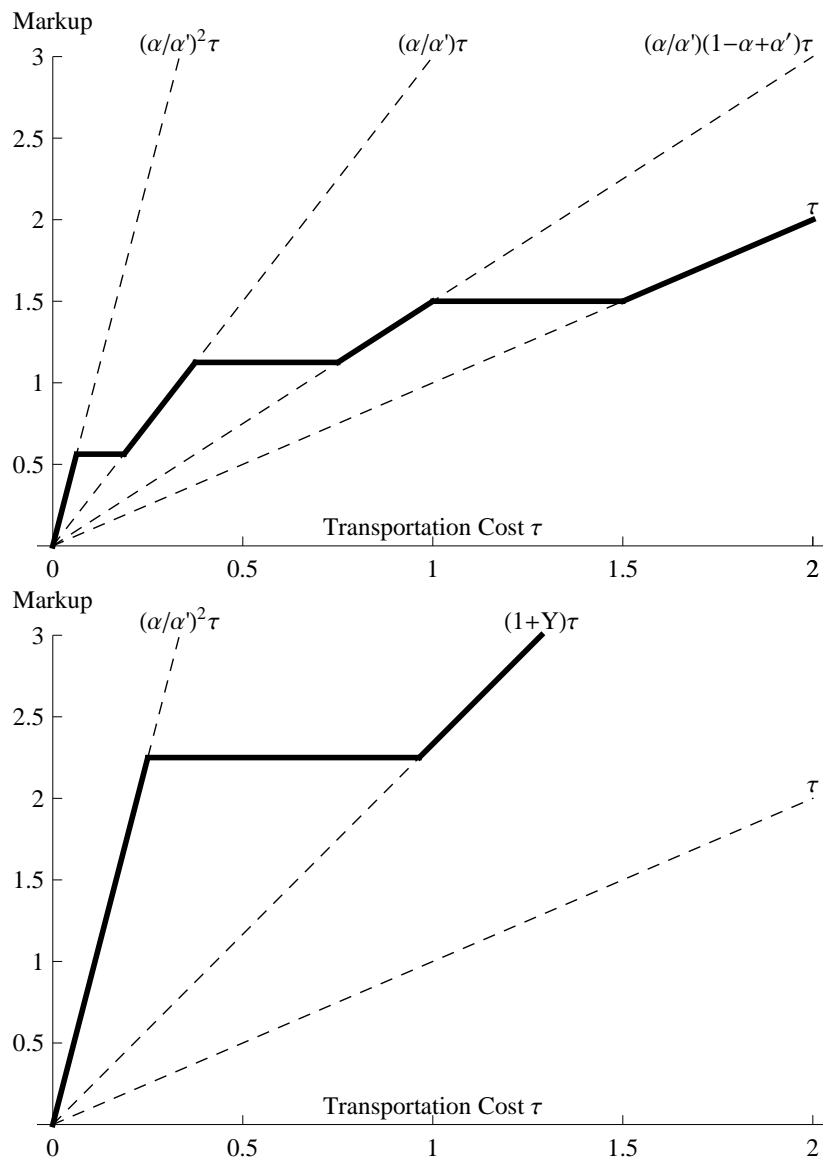


Figure 1: Firm markup as a function of transportation cost  $\tau$  in a Hotelling duopoly with the no-free-lunch constraint and consumers who receive Bernoulli taste shocks and underestimate demand ( $\alpha' < \alpha$ ). Top panel: attentive consumers. Bottom panel: inattentive consumers. The figure is plotted for  $c = 0$ ,  $\alpha = 3/4$ ,  $\alpha' = 1/4$ , and  $v_0$  sufficiently high for strict full-market-coverage.

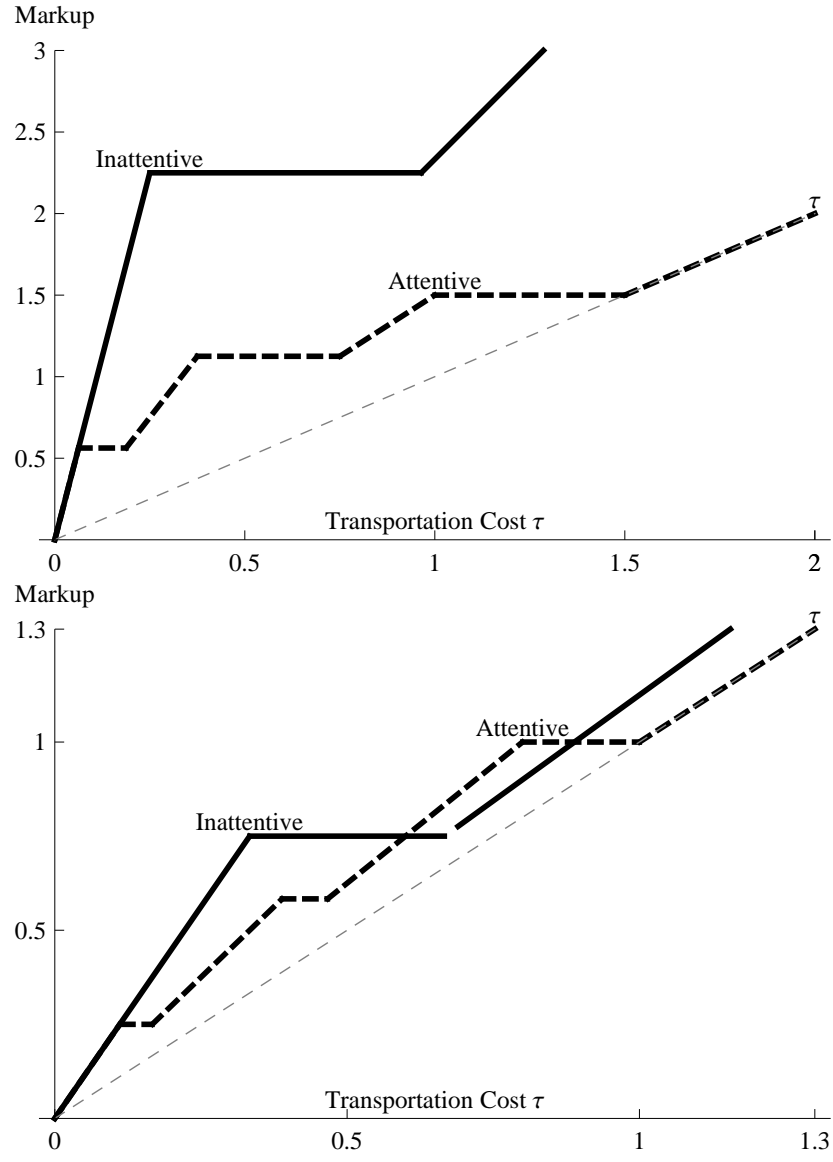


Figure 2: Firm markup as a function of transportation cost  $\tau$  in a Hotelling duopoly with the no-free-lunch constraint and consumers who receive Bernoulli taste shocks and underestimate demand ( $\alpha' < \alpha$ ). Solid line: attentive consumers. Dashed line: inattentive consumers. Top panel depicts severe bias:  $\alpha = 3/4$  and  $\alpha' = 1/4$ . Bottom panel depicts mild bias:  $\alpha = 1/2$  and  $\alpha' = 1/3$ . In both cases,  $c = 0$  and  $v_0$  is sufficiently high for strict full-market-coverage.

## A Additional Details

### A.1 Pricing first-order conditions referenced in Proposition 3

Given case (2) of Proposition 3 ( $\mu_H^* > \mu_L^*$ ), first-order conditions for marginal prices on contract  $L$  are:

$$v_L = c + \int_{p_{2L}}^{p_{2L}+p_{3L}} (v - c) f_L(v) dv + \frac{-\partial\Pi}{\partial U_H} \frac{F_L(v_L) - F_H(v_{HL})}{(1 - \beta) G_L(U_L) f_L(v_L)}, \quad (17)$$

$$p_{2L} = c + \frac{F_H(v_{HL}) - \partial\Pi}{F_L(v_L)} \frac{F_L(p_{2L}) - F_H(p_{2L})}{\partial U_H (1 - \beta) G_L(U_L) f_L(p_{2L})}, \quad (18)$$

$$p_{2L} + p_{3L} = c + \frac{(1 - F_H(v_{HL})) - \partial\Pi}{(1 - F_L(v_L))} \frac{F_L(p_{2L} + p_{3L}) - F_H(p_{2L} + p_{3L})}{\partial U_H (1 - \beta) G_L(U_L) f_L(p_{2L} + p_{3L})}, \quad (19)$$

where  $v_{HL} = v_L + \int_{p_{2L}}^{p_{2L}+p_{3L}} (F_L(v) - F_H(v)) dv$ .

Given case (3) of Proposition 3 ( $\mu_H^* < \mu_L^*$ ), first-order conditions for marginal prices on contract  $H$  are:

$$v_H = c + \int_{p_{2H}}^{p_{2H}+p_{3H}} (v - c) f_H(v) dv - \frac{-\partial\Pi}{\partial U_L} \frac{F_L(v_{LH}) - F_H(v_H)}{\beta G_H(U_H) f_H(v_H)}, \quad (20)$$

$$p_{2H} = c - \frac{F_L(v_{LH}) - \partial\Pi}{F_H(v_H)} \frac{F_L(p_{2H}) - F_H(p_{2H})}{\partial U_L \beta G_H(U_H) f_H(p_{2H})}, \quad (21)$$

$$p_{2H} + p_{3H} = c - \frac{(1 - F_L(v_{LH})) - \partial\Pi}{(1 - F_H(v_H))} \frac{F_L(p_{2H} + p_{3H}) - F_H(p_{2H} + p_{3H})}{\partial U_L \beta G_H(U_H) f_H(p_{2H} + p_{3H})}, \quad (22)$$

where  $v_{LH} = v_L - \int_{p_{2L}}^{p_{2L}+p_{3L}} (F_L(v) - F_H(v)) dv$ .

Consider case (2)  $\mu_H^* > \mu_L^*$ . To understand the marginal prices characterized by equations (17)-(19) it is helpful to begin by understanding optimal pricing in a simplified model with only one purchase opportunity rather than two. Given  $\mu_H^* > \mu_L^*$ , the downward incentive constraint binds and the optimal marginal price on the low contract is distorted above marginal cost:<sup>37</sup>

$$p_L = c + \frac{-\partial\Pi}{\partial U_H} \frac{F_L(p_L) - F_H(p_L)}{(1 - \beta) G_L(U_L) f_L(p_L)}. \quad (23)$$

The distortion away from marginal cost follows from standard price discrimination logic. High types find increases in marginal prices more costly than do low types because high types make more purchases. Thus raising marginal prices on the low contract relaxes the downward incentive constraint at the cost of distorting low-types' allocations downwards. The optimal distortion in marginal price, given by equation (23), follows from a first-order condition which equates the marginal cost of distortions,  $(p_L - c)(1 - \beta) G_L(U_L) f_L(p_L)$ , to the marginal benefit of relaxing

---

<sup>37</sup>Given ZOOM, for which  $-\partial\Pi/\partial U_H = \beta$  and  $G_L(U_L) = 1$ , this matches Courty and Li (2000).

the constraint,  $-\partial\Pi/\partial U_H (F_L(p_L) - F_H(p_L))$ . Unpacking the cost of distortion,  $(p_L - c)$  is the lost surplus from the marginal foregone purchase and  $(1 - \beta) G_L(U_L) f_L(p_L)$  is the likelihood a consumer is a low-type on the margin. Unpacking the benefit,  $(F_L(p_L) - F_H(p_L))$  is the amount by which raising  $p_L$  relaxes the downward incentive constraint, and  $-\partial\Pi/\partial U_H$  is the shadow value of relaxing the constraint.

Optimal marginal pricing for the case  $\mu_H^* > \mu_L^*$  can now be understood by comparing equations (17)-(19) to equation (23). First-period marginal cost in equation (17) is adjusted by  $\int_{p_{2L}}^{p_{2L}+p_{3L}} (v - c) f_L(v) dv$ , which is the second-period surplus lost when a first-period purchase triggers the penalty fee in period two. Second-period distortions away from marginal cost in equations (18)-(19) are adjusted by the additional terms  $F_H(v_{HL})/F_L(v_L) < 1$  and  $(1 - F_H(v_{HL})) / (1 - F_L(v_L)) > 1$  respectively. This implies that penalty fee  $p_{3L}$  is positive. Marginal prices are distorted upwards to discourage the high type from choosing the low contract. A positive penalty fee makes the second-period distortion larger after an initial purchase. This is optimal because a deviating high type is more likely to purchase in the first period than a low type.

## A.2 Pricing first-order conditions and definitions referenced in Proposition 6

Given case (2) of Proposition 6, the threshold  $v_L$  satisfies the first-order condition:

$$v_L = c + \frac{\beta}{1 - \beta} \frac{F_L(v_L) - F_H(v_{HL})}{f_L(v_L)} \frac{-\partial\Pi/\partial U_H}{\beta G_L(U_L)} \left( (1 + p_{3L} f_L(v_L)) + \frac{1}{2} (F_L(v_L) - F_H(v_{HL})) h'_L(v_L) \right), \quad (24)$$

where  $v_{HL} = v_L + p_{3L} (F_L(v_L) - F_H(v_{HL}))$ .

Given case (3) of Proposition 6, the threshold  $v_H$  satisfies the first-order condition:

$$v_H = c - \frac{1 - \beta}{\beta} \frac{F_L(v_{LH}) - F_H(v_H)}{f_H(v_H)} \frac{-\partial\Pi/\partial U_L}{(1 - \beta) G_H(U_H)} \left( (1 + p_{3H} f_H(v_H)) - \frac{1}{2} (F_L(v_{LH}) - F_H(v_H)) h'_H(v_H) \right), \quad (25)$$

where  $v_{LH} = v_H - p_{3H} (F_L(v_{LH}) - F_H(v_H))$ .

Constants  $X_H$  and  $X_L$ , which delineate cases in Proposition 6, follow. First,  $X_H$  is defined as

$$X_H \equiv 2 \int_c^{v_{HL}} (v - c) dF_H(v) + (v_{HL} - c) (F_L(c) - F_H(v_{HL})), \quad (26)$$

where  $v_{HL}$  uniquely satisfies  $v_{HL} = c + h_L(c) (F_L(c) - F_H(v_{HL}))$ . Second,  $X_L$  is defined as

$$X_L \equiv 2 \int_{v_{LH}}^c (c - v) dF_H(v) - (c - v_{LH}) (F_L(v_{LH}) - F_H(c)), \quad (27)$$

where  $v_{LH}$  uniquely satisfies  $v_{LH} = c - h_H(c) (F_L(v_{LH}) - F_H(c))$ . Note that  $X_L > 0$  and  $X_H > 0$ .

## References

- Agarwal, Sumit, John C. Driscoll, Xavier Gabaix, and David I. Laibson**, “Learning in the Credit Card Market,” Working Paper April 23 2011. <http://ssrn.com/abstract=1091623>.
- Altschul, Michael F., Christopher Guttman-McCabe, and Brian M. Josef**, “Comments of CTIA - The Wireless Association,” January 10th 2011. <http://fjallfoss.fcc.gov/ecfs/document/view?id=7021025497>.
- Armstrong, Mark and John Vickers**, “Competitive Price Discrimination,” *RAND Journal of Economics*, 2001, 32 (4), 579–605.
- and —, “Consumer protection and contingent charges,” MPRA Paper No. 37239 March 2012. <http://mpa.ub.uni-muenchen.de/37239/>.
- Aron-Dine, Aviva, Liran Einav, Amy Finkelstein, and Mark Cullen**, “Moral Hazard in Health Insurance: How important is forward looking behavior?,” Mimeo 2011.
- Aspan, Maria**, “Free Checking Thrives at Smaller Banks, Durbin Notwithstanding,” *American Banker*, August 29 2011. [http://www.americanbanker.com/issues/176\\_168/free-checking-durbin-debit-interchange-1041641-1.html](http://www.americanbanker.com/issues/176_168/free-checking-durbin-debit-interchange-1041641-1.html).
- Ausubel, Lawrence M.**, “The Failure of Competition in the Credit Card Market,” *The American Economic Review*, 1991, 81 (1), 50–81.
- and **Haiyan Shui**, “Time inconsistency in the credit card market,” Working Paper 2005.
- Bank of America**, “Personal Schedule of Fees,” Website March 1 2010. <https://www3.bankofamerica.com/efulfillment/documents/91-11-3000ED.20100201.htm>.
- Benoit, David**, “Customers Opt in for Overdraft Protection,” *The Wall Street Journal*, November, 26 2010, *Business*.
- Borenstein, Severin**, “To What Electricity Price Do Consumers Respond? Residential Demand Elasticity Under Increasing-Block Pricing,” Preliminary Draft April 30 2009.
- Bubb, Ryan and Alex Kaufman**, “Consumer Biases and Firm Ownership,” NYU Law and Economics Research Paper No. 11-35. 2011. <http://ssrn.com/abstract=1945852>.
- Burhouse, Susan, Patricia Cashman, Robert Cordeiro, Timothy Critchfield, Yan Lee, Victoria Pawelski, and Katherine Samolyk**, “FDIC Study of Bank Overdraft Programs,” 2008. [http://www.fdic.gov/bank/analytical/overdraft/FDIC138\\_Report\\_Final\\_v508.pdf](http://www.fdic.gov/bank/analytical/overdraft/FDIC138_Report_Final_v508.pdf).
- Campbell, Dakin**, “U.S. Overdraft-Fee Limits May Cost Banks \$15 Billion (Update1),” *Bloomberg*, November 24 2009. <http://www.bloomberg.com/apps/news?pid=newsarchive&sid=a0Dvn7hZ2cVc&pos=6>.
- Courty, Pascal and Hao Li**, “Sequential Screening,” *The Review of Economic Studies*, 2000, 67 (4), 697–717.

- CTIA - The Wireless Association**, “CTIA-The Wireless Association, Federal Communications Commission and Consumers Union Announce Free Alerts to Help Consumers Avoid Unexpected Overage Charges,” October 17 2011. <http://www.ctia.org/media/press/body.cfm/prid/2137>.
- Dash, Eric and Nelson D. Schwartz**, “Banks Seek to Keep Profits as New Oversight Rules Loom,” *The New York Times*, July 15 2010, p. B1.
- Date, Raj**, “Lessons Learned from the Financial Crisis: The Need for the CFPB,” 2011. <http://www.consumerfinance.gov/speeches/lessons-learned-from-the-financial-crisis-the-need-for-the-cfpb/>.
- DellaVigna, Stefano**, “Psychology and economics: Evidence from the field,” *Journal of Economic Literature*, 2009, 47 (2), 315–372.
- and **Ulrike Malmendier**, “Contract Design and Self-Control: Theory and Evidence,” *The Quarterly Journal of Economics*, 2004, 119 (2), 353–402.
- Deloney, Amalia, Linda Sherry, Susan Grant, Parul P. Desai, Chris M. Riley, Matthew F. Wood, John D. Breyault, Jessica J. Gonzalez, and Benjamin Lennett**, “Comments of the Center for Media Justice, Consumer Action, Consumer Federation of America, Consumers Union, Free Press, Media Access Project, National Consumers League, National Hispanic Media Coalition and New America Foundation Open Technology Initiative in response to notice of proposed rulemaking,” January 10th 2011. <http://fjallfoss.fcc.gov/ecfs/document/view?id=7021025418>.
- Diamond, Peter A.**, “A model of price adjustment,” *Journal of Economic Theory*, 1971, 3 (2), 156–168.
- Dinzeo, Maria**, “Customers Say Bank Sticks it to Them,” *Courthouse News Service*, August 11 2010. <http://www.courthousenews.com/2010/08/11/29486.htm>.
- Edlin, Aaron S. and Chris Shannon**, “Strict monotonicity in comparative statics,” *Journal of Economic Theory*, 1998, 81 (1), 201–219.
- Eliaz, Kfir and Ran Spiegler**, “Contracting with Diversely Naive Agents,” *The Review of Economic Studies*, 2006, 73 (3), 689–714.
- and — , “Consumer Optimism and Price Discrimination,” *Theoretical Economics*, 2008, 3 (4), 459–497.
- Ellison, Glenn**, “A Model of Add-on Pricing,” *The Quarterly Journal of Economics*, 2005, 120 (2), 585–637.
- FCC**, “FCC Proposes Rules to Help Mobile Phone Users Avoid ‘Bill Shock’,” Press Release October 14th 2010. [http://hraunfoss.fcc.gov/edocs\\_public/attachmatch/DOC-302192A1.pdf](http://hraunfoss.fcc.gov/edocs_public/attachmatch/DOC-302192A1.pdf).
- Federal Reserve Board**, “Federal Reserve announces final rules prohibiting institutions from charging fees for overdrafts on ATM and one-time debit card transactions,” Press Release November 12 2009.
- , “Regulation E final rule,” *Federal Register*, 2009, 74 (220), 59033–59056.

- Gabaix, Xavier and David Laibson**, “Shrouded Attributes, Consumer Myopia, and Information Suppression in Competitive Markets,” *Quarterly Journal of Economics*, 2006, 121 (2), 505–540.
- Genachowski, Julius**, “Prepared Remarks of Chairman Julius Genachowski,” 2010.
- Grubb, Michael D.**, “Selling to Overconfident Consumers,” *American Economic Review*, 2009, 99 (5), 1770–1807.
- , “Consumer Inattention and Bill-Shock Regulation,” Working Paper 2012.
- and **Matthew Osborne**, “Cellular Service Demand: Biased Beliefs, Learning, and Bill Shock,” Working Paper 2012.
- Hadar, Josef and William R. Russell**, “Rules for Ordering Uncertain Prospects,” *American Economic Review*, 1969, 59 (1), 25.
- Heidhues, Paul, Botond Köszegi, and Takeshi Murooka**, “The Market for Deceptive Products,” Working Paper 2012.
- Ito, Koichiro**, “Do Consumers Respond to Marginal or Average Price? Evidence from Nonlinear Electricity Pricing,” Working Paper November 20 2010.
- Jiang, Lai**, “The Welfare Effects of ‘Bill Shock’ Regulation in Mobile Telecommunication Markets,” Working Paper July 2011.
- Laibson, David**, “Golden Eggs and Hyperbolic Discounting,” *The Quarterly Journal of Economics*, 1997, 112 (2), 443–477.
- Liebman, Jeffrey B. and Richard Zeckhauser**, “Schmeduling,” Working Paper October 2004.
- Martin, Andrew**, “Bank of America to End Debit Overdraft Fees,” *The New York Times*, March 10 2010.
- McCune, Richard D., David C. Wright, Kristy M. Arevalo, and Jae K. Kim**, “Class Action Complaint,” US District Court, Northern District of California, Case No. CV 10 3463 August 6 2010.
- Miao, Chun-Hui**, “Consumer myopia, standardization and aftermarket monopolization,” *European Economic Review*, 2010, 54 (7), 931–946.
- O’Donoghue, Ted and Matthew Rabin**, “Choice and Procrastination,” *Quarterly Journal of Economics*, 2001, 116 (1), 121–160.
- of Justice, Department**, “Justice Department Files Antitrust Lawsuit to Block AT&Ts Acquisition of T-Mobile,” Office of Public Affairs Press Release August 31 2011. <http://www.justice.gov/opa/pr/2011/August/11-at-1118.html>.
- Pavan, Alessandro, Ilya Segal, and Juuso Toikka**, “Dynamic Mechanism Design: Incentive Compatibility, Profit Maximization and Information Disclosure,” Working Paper 2011.
- Piccione, Michele and Ariel Rubinstein**, “On the Interpretation of Decision Problems with Imperfect Recall,” *Games and Economic Behavior*, 1997, 20 (1), 3–24.

- Rochet, Jean-Charles and Lars A. Stole**, “Nonlinear Pricing with Random Participation,” *The Review of Economic Studies*, 2002, 69 (1), 277–311.
- and **Lars Stole**, “The Economics of Multidimensional Screening,” in M. Dewatripont, Lars Peter Hansen, and Stephen J. Turnovsky, eds., *Advances in economics and econometrics: theory and applications - eighth world Congress*, Vol. 36 of *Econometric Society Monographs*, New York: Cambridge University Press, 2003.
- Saez, Emmanuel**, “Do Taxpayers Bunch at Kink Points?,” Working Paper June 2002.
- Sidel, Robin and Dan Fitzpatrick**, “End Is Seen to Free Checking,” *The Wall Street Journal*, June 16 2010.
- Skiba, Paige Marta and Jeremy Bruce Tobacman**, “Payday Loans, Uncertainty and Discounting: Explaining Patterns of Borrowing, Repayment, and Default,” Vanderbilt Law and Economics Research Paper No. 08-33 2008.
- Spiegler, Ran**, *Bounded Rationality and Industrial Organization*, Oxford University Press, 2011.
- Stango, Victor and Jonathan Zinman**, “What do Consumers Really Pay on Their Checking and Credit Card Accounts? Explicit, Implicit, and Avoidable Costs,” *American Economic Review Papers and Proceedings*, 2009, 99 (2).
- and — , “Limited and Varying Consumer Attention: Evidence from Shocks to the Salience of Overdraft Fees,” Working Paper March 2012.
- Stole, Lars A.**, “Price Discrimination and Competition,” in Mark Armstrong and Robert K. Porter, eds., *Handbook of Industrial Organization*, Vol. Volume 3, Elsevier, 2007, chapter 34, pp. 2221–2299.
- Stross, Randall**, “Please, just tell me when I’m nearing my limit,” *The New York Times*, January 22 2011, *Business*. [http://www.nytimes.com/2011/01/23/business/23digi.html?\\_r=2&hpw](http://www.nytimes.com/2011/01/23/business/23digi.html?_r=2&hpw).
- Wyatt, Edward**, “F.C.C. Wants to Stop Cellphone ‘Bill Shock’,” *The New York Times*, October 13 2010, *Business Day*, B3.
- , “Consumer Inquiry Focuses on Bank Overdraft Fees,” *The New York Times*, February 22 2012, *Business Day*. <http://www.nytimes.com/2012/02/22/business/bank-overdraft-fees-to-be-scrutinized-by-consumer-bureau.html>.
- Yao, Song, Carl F. Mela, Jeongwen Chiang, and Yuxin Chen**, “Determining Consumers’ Discount Rates with Field Studies,” Working Paper June 15 2011.