Penalty Pricing:
Optimal Price-Posting Regulation with Inattentive Consumers

Michael D. Grubb
MIT Sloan School of Management
Cambridge, MA 02142
mgrubb@mit.edu
www.mit.edu/˜mgrubb

January 21, 2011

Abstract

For many goods and services, such as cellular-phone service, electricity, health-care, and debit or credit-card transactions, the marginal price of the next unit of service depends on past usage. As a result, consumers who are inattentive to their past usage may be aware of contract terms and yet still uncertain about the marginal price of the next unit. I develop a model of inattentive consumption, derive optimal pricing-policies when consumers are inattentive, and evaluate price-posting regulation requiring firms to publish marginal price at the time of each transaction. When consumers are homogeneous and have unbiased beliefs, inattention has no substantive effect on market outcomes. Otherwise, inattention leads firms to charge surprise penalty-fees for high usage. When consumers are heterogeneous ex ante and have unbiased beliefs, inattention and penalty fees increase welfare in sufficiently competitive markets, and price-posting regulation is counterproductive. Under these conditions, cellular-phone usage-alerts under consideration by the FCC could reduce welfare and harm consumers. If consumers are homogeneous ex ante but underestimate their demand, then price-posting regulation has an ambiguous impact on total welfare but may have large distributional benefits by increasing price competition and protecting consumers from exploitation. Hence the Federal Reserve's new opt-in rule for debit-card overdraft-protection could substantially benefit consumers.

1 Introduction

In many important situations, consumers may be fully aware of the full schedule of marginal charges when making an ex ante decision to sign a contract, but nevertheless, ex post are uncertain about the marginal price of any given transaction. This occurs whenever marginal prices vary with the level of consumption (as they do when firms levy penalty fees for excessive usage) and, due to inattention, consumers are unaware of their past consumption when making additional consumption choices. Note that marginal prices vary with usage for a wide variety of products and services including electricity, cellular-phone service, health insurance, and debit and credit-card transactions. In each case, inattention would create uncertainty about marginal price at the point of sale.

For example, a cellular phone customer may be fully aware that the first 500 minutes are billed at zero cents a minute and later minutes at a penalty (or "overage") rate of 45 cents a minute. However, he may be uncertain whether the next call will be billed at zero cents or 45 cents per minute, because he does not know how much he has already used the phone. Similarly, a new checking account enrollee may be fully aware that overdraft penalty fees are $35 per transaction, but be unaware whether her next debit transaction will be free or incur a $35 penalty because she is uncertain about her checking balance. (Stango and Zinman (2010) find in survey data "that 60% of overdrafters reported overdrafting because they 'thought there was enough money in my account'".) Notice that the same price uncertainty 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.

In each example, firms have the ability to disclose to consumers whether or not a penalty fee is applicable at the point of sale. A mobile phone screen could flash "overage rate applies" before making any call once included minutes are used up. A debit-card processing terminal could ask "Overdraft fee applies. Continue - Yes/No?" before processing transactions on an overdrawn account. Firms' choices not to make this information so readily available suggest that firms benefit from consumer uncertainty about marginal price. Recent regulation of overdraft fees by the Federal Reserve Board and consideration of "bill shock" regulation by the FCC suggest that regulators believe that the lack of transparency is bad for consumers and bad for welfare. 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 clearly agree (Genachowski 2010).

In this paper, I develop a model to answer the following three questions: First, if consumers are inattentive to their own past consumption, do firms profit by charging surprise penalty-fees for excessive usage? Second, if so, does price-posting regulation requiring firms to disclose marginal price at the point of sale benefit consumers more than it harms firms and thus increase welfare? Third, how do the conclusions depend on the level of competition between firms? In the process I also answer a fourth question: how do the conclusions depend on consumer heterogeneity and consumer biases?

I begin by modeling the consumption behavior of inattentive consumers. I assume that once an inattentive consumer signs a cellular-phone contract or opens a bank account, consumption opportunities arise sequentially and each decision to make an additional phone call or debit-card transaction is made without any recollection of prior usage. Moreover, I assume that consumers are aware of their own inattention when making plans. In Section 3, I show that for any price schedule, an inattentive consumer's optimal strategy is to use a threshold rule and consume only those units valued above the endogenous expected marginal price. 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 (2010). (These papers use the threshold rules in demand or labor supply estimation, while I explore the supply-side ramifications of such behavior.)

In Section 3, I develop a benchmark model which assumes that at the time of contracting consumers are homogeneous (so there is no scope for price discrimination) and consumers have correct beliefs (so there are no biases to exploit). For simplicity, I assume throughout the paper that there are only two consumption opportunities. As a result, the effect of price-posting regulation is to make inattentive consumers attentive. (With more consumption opportunities, greater disclosure would be needed to make inattentive consumers attentive.) To analyze the effect of price-posting regulation, I therefore solve for equilibrium prices under two conditions: first with attentive consumers and second with inattentive consumers.

Under the benchmark-model assumptions, the primary result is equivalence. Regardless of the level of market competition, neither consumer inattention nor price-posting regulation affect substantive market outcomes including allocations, firm profits, and consumer surplus. The only effect of price-posting regulation is to restrict the set of feasible equilibrium prices. Firms would find it optimal to set marginal price equal to marginal cost and not charge any penalty fees for excessive usage, regardless of whether consumers are attentive or inattentive. However these prices are uniquely optimal only with attentive consumers or price-posting regulation. Thus price-posting regulation could induce firms to eliminate penalty fees but compensate with other charges. This captures the argument of some critics of price-posting regulation - that it would only cause firms to recoup lost penalty fees through fixed fees and other charges (Federal Reserve Board 2009a).[1] However, the result relies heavily on the assumption of homogeneity. Moreover, it clearly does not explain the widespread use of penalty fees, choices by firms not to disclose marginal price or alert consumers to penalty fees at the point of sale, or firms' expressed aversion to regulation which would help consumers avoid purchases that trigger penalties (Federal Reserve Board 2009a).

One reason that penalty fees are used in practice may be that they are useful for discriminating between consumers with heterogeneous expectations about their own future demand for the product or service. For instance, cellular phone overage fees are not only designed to generate revenue directly (Grubb (2009) finds 22 percent of revenues were from overage charges), but also to encourage consumers who anticipate high demand to self select into larger calling plans. Section 4 enriches the benchmark model by incorporating two ex ante types, with low and high expectations of future demand. Given such heterogeneity, I find that if consumers are inattentive, surprise penalty-fees and the resulting price uncertainty can strictly increase not only firm profits but also welfare. The intuition is that price uncertainty relaxes incentive constraints which otherwise limit a firm's ability to price discriminate. This allows firms with market power to extract more information rents from consumers and increase profits - which can explain firm aversion to price-posting regulation. Perhaps more surprising is the fact that inattention may help some consumers and increase overall welfare. It can allow firms to price discriminate effectively while imposing smaller allocative distortions than they would otherwise. This is not always the case (sometimes inattention can increase firm profits but also cause them to increase distortions and reduce welfare), but it is always true when markets are fairly competitive. Thus, the first of two main-results is that in fairly-competitive markets with heterogeneous inattentive-consumers who have correct beliefs, penalty fees are socially valuable and price-posting regulation is counter productive.

The paper's first main result could suggest caution in adopting bill-shock regulation under consideration by the FCC, which would require carriers to alert customers of rapidly accumulating fees by text message (FCC 2010). A fundamental part of cellular-phone-service pricing is separating consumers with different expectations of usage among different contracts with different allowances of included minutes. If one believes that cellular phone customers have correct beliefs and the cellular market is sufficiently competitive, then inattention is good for welfare - and price-posting regulation would be counter-productive. Moreover, while price-posting regulation is unambiguously good for consumers holding prices fixed, prices will change such that some consumers are made worse off. Thus consumer groups' advocacy for the policy may be misplaced. But note that these

[1] 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).

results depend on assumptions about beliefs and competition that may not be valid. In fact, evidence shows that cellular customers have biased beliefs (Grubb 2009, Grubb and Osborne 2010) and it is not obvious that the industry is highly competitive. As a result the welfare impact of price-posting regulation is ambiguous.[2]

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). Effective July 1, 2010 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 2009b). Does Section 4's model of heterogeneous consumers with correct beliefs suggest this regulation is welfare reducing? In fact it does not apply. Prior to the regulation, banks typically did not differentiate checking accounts by varying overdraft fees. For instance, before ending overdraft protection on ATM and debit-card transactions, Bank of America offered a variety of checking accounts, but offered the same overdraft fee schedule on all of them (Bank of America 2010). Thus heterogeneity in expectations of overdraft usage is typically not an important dimension of self-selection across checking accounts.

Since neither the benchmark model nor Section 4's model of price discrimination explain banks' widespread use of overdraft fees, I explore a more compelling alternative: that consumers underestimated the incidence of overdraft fees. There is substantial evidence that consumers often have biased beliefs at the time of contracting (Ausubel and Shui 2005, DellaVigna and Malmendier 2006, Grubb 2009). Moreover, demand underestimation would arise in the context of overdraft fees if consumers were 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). Section 5 enriches the benchmark model by assuming that consumers underestimate their own future demand. Firms can profit from this bias by raising marginal prices that consumers underestimate the likelihood of paying. However, attentive consumers who underestimate their own value for a service cannot be exploited in the sense that they can never be induced to pay more than their average value for a product or service. In contrast, the paper's second main result is that if consumers are both inattentive and underestimate their own values for a service, they can be grossly exploited and firms can extract profits orders of magnitude higher than the total surplus using penalty fees. This is true even

[2]Moreover, 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.

in fairly competitive markets, as the combination of penalty fees and consumer inattention can significantly soften price competition.

The total welfare effects of price-posting regulation are ambiguous, but may be second order relative to the distributional effects. Regulation requiring the posting of the marginal price at the point of each transaction would mitigate the consumer welfare losses due to biased beliefs and ensure that consumers are not exploited. The redistribution of surplus from firms to consumers involved in ending exploitation could be orders of magnitude larger than the total surplus generated by the market. In fact, it is possible for a service with zero or negative social value to be sold at high profit (and high consumer loss) prior to regulation, but be efficiently shut down by price-posting regulation. This may explain the fact that Bank of America, the bank with the largest overdraft fee revenue in 2009 (estimated to be $2.2 billion per year by a Sandler O'Neill + Partners report (Sidel and Fitzpatrick 2010)), responded to the Fed's new "opt-in" requirement by ending overdraft protection for one-time debit-card transactions (Martin 2010).[3]

This paper considers settings where consumers are inattentive to their own past consumption and shows that firms optimally charge penalty fees for excessive usage to take advantage of such inattention. In such settings, the results suggest that regulators should require price-posting for products such as overdraft protection that are not differentially priced to sort consumers into different contracts. However, regulators should be more cautious for products such as cellular-phone calls that are an important dimension of consumers' self selection across contracts. In particular, it predicts that the Federal Reserve Board's opt-in rule for overdraft fees on debit transactions could strongly benefit consumers, but that the bill shock regulation under consideration by the FCC has the potential to be counter productive.

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 equivalence result. Section 4 analyzes the model enriched with ex ante heterogeneity, which explores the role of inattention, penalty fees, and price-posting regulation in price discrimination. Section 5 makes the alternative extension to biased consumer beliefs, for which inattention can increase the scope for exploitation. Finally Section 6 concludes. All proofs not included in the text are provided in the appendix.

[3]Bank of America still offers an alternative overdraft protection service that transfers money from a linked savings or credit card account for a $10 fee. This opt-in service has always been available, but was typically unused by customers who failed to opt-in and ended up paying the higher $35 overdraft fees that are now subject to regulation.

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 assumption is made in both empirical work (Cardon and Hendel 2001, Reiss and White 2005, Gaynor, Shi, Telang and Vogt 2005, Lambrecht, Seim and Skiera 2007, Huang 2008) and throughout the theoretical literatures on nonlinear pricing (Wilson 1993) and two-period sequential screening (Baron and Besanko 1984, Riordan and Sappington 1987, Miravete 1996, Courty and Li 2000, Miravete 2005, Grubb 2009). When applied to settings in which consumers make many separate consumption decisions within in a billing period, the implicit assumption is that consumers have perfect foresight to predict all these individual choices at the start of the billing period. This is usually implausible and is empirically rejected by the lack of bunching at tariff kink points in electricity (Borenstein 2009) and cellular-phone-service (Grubb and Osborne 2010) consumption.

Relaxing the perfect foresight assumption, if firms charge penalty fees for excessive consumption, attentive consumers must solve a dynamic programming problem similar to the airline revenue management problem surveyed by McAfee and te Velde (2007). A key feature of the solution is that attentive consumers reduce consumption after penalty fees are triggered (equation (1)). Using detailed call-level data, Grubb and Osborne (2010) find no evidence of this behavior among cellular phone subscribers, suggesting that they are in fact inattentive to their own past usage within the billing cycle. In the context of checking-account overdraft-fees, Stango and Zinman (2009) find even more direct evidence of inattention: the median consumer could avoid more than 60% of overdraft charges by using alternative cards (checking or credit) with available liquidity. Using a different data set, Stango and Zinman (2010) 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'".[4]

Formally, the inattentive consumer's decision problem analyzed in Section 3 exhibits Piccione and Rubinstein's (1997b) *absentmindedness*. Subject to the information constraint imposed by absentmindedness, consumers behave optimally. Psychology experiments demonstrate that attention is a limited resource (Broadbent 1958). DellaVigna (2009) surveys recent work in economics which examines inattention to shipping costs, nontransparent taxes, financial news, and other information. I show that inattentive consumers purchase all units valued above the endogenous expected

[4] Stango and Zinman (2010) 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 (2008) finding that accruing one credit card late penalty fee reduces the likelihood of incurring one in the following month.

marginal price.

Liebman and Zeckhauser (2004) analyze optimal pricing given alternative deviations from unbounded rationality by consumers faced with multi-part tariffs. Liebman and Zeckhauser's (2004) deviations, which they dub "ironing" and "spotlighting", are based on decision errors rather than an information limitation. Liebman and Zeckhauser's (2004) first model (ironing) is static. It assumes that consumers make a single quantity choice and confuse the average price with the marginal price. Liebman and Zeckhauser's (2004) second model (spotlighting) is dynamic. It assumes consumers make consumption decisions one unit at a time and myopically base their consumption choices on the marginal price of the current unit.

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 marginal 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 (Diamond 1971), purchase a base product such as a printer (Ellison 2005), select a hotel (Gabaix and Laibson 2006), or open a checking account (Bubb and Kaufman 2009). As a result, marginal fees for add-on products or services are set at monopoly levels in spite of competition or the use of two-part tariffs, either of which would normally lead to marginal cost pricing.

Section 4's model of price discrimination is related to the literature on sequential screening (Baron and Besanko 1984, Riordan and Sappington 1987, Miravete 1996, Courty and Li 2000, Miravete 2005, Grubb 2009, Pavan, Segal and Toikka 2009), in which consumers first choose from a menu of contracts and then make quantity choices after the arrival of more information. Both Courty and Li (2000) and Pavan et al. (2009) model monopoly pricing when consumers have zero outside options. Under this market condition, my model with attentive consumers would coincide with Courty and Li (2000) if consumers made only a single purchase decision. Pavan et al.'s (2009) results show that with a continuum of ex ante types at the contracting stage (and my assumption that values are distributed independently conditional on ex ante type) the optimal contract with multiple purchase opportunities is a repetition of the Courty and Li (2000) solution. However, I assume two ex ante types at the contracting stage rather than a continuum and Pavan et al.'s (2009) approach and result do not apply: the optimal contract is not a repetition of the Courty and Li (2000) solution. In addition, I solve my attentive model under more general market conditions: monopoly with heterogeneous outside options and duopoly.

Although I am unaware of other work on competitive sequential-screening, there is related work on competitive static-nonlinear-pricing, for which Stole (2007) provides an excellent survey. In particular, I incorporate competition following a similar approach to that taken by Armstrong and Vickers (2001) and Rochet and Stole (2002). Armstrong and Vickers (2001) and Rochet and Stole (2002) both contain versions of the same result: that sufficient competition in nonlinear price-schedules leads to two-part-tariff pricing at marginal cost and first-best allocations. This is a knife-edge result, which depends on the assumption that the optimal markup (ignoring incentive constraints) is exactly the same for all customer segments. I find an analogous result in my attentive model with competitive sequential screening. The first-best-allocation result (although not the two-part-tariff-pricing result) also extends to competitive sequential-screening with inattentive consumers, but in this case is more general as it holds even if optimal markups differ across customer segments.

The model explored in Section 5 assumes that at the time of contracting consumers underestimate their demand for the good or service for sale. Such consumers exhibit similar behavior to naive quasi-hyperbolic-discounters (DellaVigna and Malmendier 2004, Eliaz and Spiegler 2006) or myopic consumers (Gabaix and Laibson 2006, **?**). There is a small related literature on optimal pricing when attentive consumers have biased beliefs (Sandroni and Squintani 2007, Eliaz and Spiegler 2008, Grubb 2009, Bubb and Kaufman 2009). A common finding is that demand underestimation, due either to biased beliefs (Eliaz and Spiegler 2008, Grubb 2009), myopia (Gabaix and Laibson 2006, **?**), or naive quasi-hyperbolic-discounting (DellaVigna and Malmendier 2004, Eliaz and Spiegler 2006), leads to high marginal prices above marginal cost.

In competitive markets, economic models typically predict that firms offset high marginal fees with lower fixed fees (**?**, Gabaix and Laibson 2006, Grubb 2009). **?** shows that profits from aftermarket sales are not necessarily competed away in primary market competition because firms cannot set negative prices for primary goods. I also consider a constraint that (total) prices be non-negative, motivated by a related no-arbitrage condition that I call the *no-free-lunch* constraint. I find that biased beliefs can soften price competition given such a constraint, by forcing firms to compete on add-on fees rather than fixed fees. More importantly, I show that inattention exacerbates this softening of competition due to biased beliefs and makes consumers even worse off. 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.

Gabaix and Laibson (2006) and Bubb and Kaufman (2009) focus on the cross-subsidization of unbiased consumers by biased consumers. Despite cross-subsidization, biased consumers who are attentive can never be exploited in the sense that they always achieve at least their outside options.[5] In contrast, I show that inattention allows consumers to be exploited and can drastically exacerbate the cost of biased beliefs to consumers, even in fairly competitive markets.

3 Benchmark Model

This section develops the underlying model structure used throughout the paper. The benchmark assumptions that are relaxed later are that consumers have correct beliefs and are homogeneous 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 price-posting regulation formally, which in the context of the model is equivalent to making inattentive consumers attentive. Comparing equilibrium pricing with inattentive consumers to that with attentive consumers thus illuminates the effect of price-posting regulation. The primary result in this section is an equivalence result: neither inattention nor price-posting regulation affect substantive market outcomes.

3.1 Model

Game players are mass 1 of consumers and $N \geq 1$ firms. Consumers privately learn a vector of N firm-specific (brand) taste shocks $\mathbf{x}$ that is mean zero (and could for instance capture location on a Hotelling line). At the contracting stage ($t = 0$), firms simultaneously offer contracts, and each consumer either signs a contract or receives their 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 $P^i(q_1, q_2)$ to firm i, as a function of past quantity choices. Firm i's offered contract can be any deterministic price schedule:[6]

$$P^i(q_1, q_2) = p_0^i + p_1^i q_1 + p_2^i q_2 + p_3^i q_1 q_2,$$

[5] In Bubb and Kaufman's (2009) model, biased consumers correctly predict their value for the bundle of the base good and the add-on, but overestimate their value of the base good without the add-on. Since they are over-estimating the value of the base good, they can be induced to over-pay and be exploited.

[6] See Rochet and Stole (2002) for an insightful discussion of this assumption.

characterized by the vector of prices $\mathbf{p}^i = \left(p_0^i, p_1^i, p_2^i, p_3^i\right)$.

A consumer's base payoff 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:

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

Conditional on signing a contract with prices $\mathbf{p}$, a consumer's optimal consumption strategy can be described by a function mapping valuations to quantity choices: $\mathbf{q}\left(\mathbf{v}; \mathbf{p}\right)$. A consumer's base expected payoff from contracting with firm i at the contracting stage and making optimal consumption choices thereafter is $U^i = E\left[u\left(\mathbf{q}\left(\mathbf{v}; \mathbf{p}^i\right), \mathbf{v}\right)\right]$. Similarly, let $S^i = v_0 + E\left[\sum_{t=1}^{2}\left(v_t - c\right) q_t\left(\mathbf{v}; \mathbf{p}^i\right)\right]$ be the expected surplus generated by a consumer contracting with firm i and making optimal consumption choices at $t \in \{1, 2\}$.

A consumer's total expected payoff, $U^i + x^i$, includes brand taste x^i. Thus, fraction $G\left(U^i; U^{-i}\right)$ of consumers of type s buy from firm i if firm i offers base expected utility of U^i, while competitors offer U^{-i}:

$$G\left(U^i; U^{-i}\right) = \Pr(U^i + x^i \geq \max_{j \neq i}\{U^j + x^j\}).$$

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

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

which can always be rewritten in terms of total surplus and consumer utility,

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

3.2 Consumer Strategies

The first step in analyzing the game is to solve the consumers problem. As I do so below, I suppress the firm i superscript from my notation.

The optimal decision rule for an attentive consume 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 $v^*\left(q^{t-1}, t\right)$ which is a function of the date t and the vector of past usage choices q^{t-1}. Let the period one and two thresholds be v_1^* and $v_2^*\left(q_1\right)$ respectively. Then

$$v_2^*\left(q_1\right) = p_2 + p_3 q_1, \tag{1}$$

and v_1^* depends on the distribution of taste shocks:

$$v_1^* = p_1 + (1 - F(p_2 + p_3))\,p_3 + \int_{p_2}^{p_2+p_3} (v - p_2)\,f(v)\,dv \tag{2}$$

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 (2) can be simplified to

$$v_1^* = p_1 + \int_{p_2}^{p_2+p_3} (1 - F(v))\,dv \tag{3}$$

Note that the period two threshold is a function of past usage if and only if the penalty fee is nonzero ($p_3 \neq 0$). Formally, the attentive consumers optimal consumption rule can be written as:

$$\mathbf{q}^A(\mathbf{v};\mathbf{p}) = \left(1_{v_1 \geq v_1^*},\, 1_{v_1 \geq v_1^*} 1_{v_2 \geq (p_2+p_3)} + 1_{v_1 < v_1^*} 1_{v_2 \geq p_2}\right). \tag{4}$$

An inattentive consumer is one who cannot condition her strategy on the date t or on past usage q^{t-1} because she does not keep track of these variables. She exhibits imperfect recall. (Note that while everyone knows the calendar date, it takes more effort to track the date within ones billing cycle for any particular service.) Otherwise, I assume that inattentive consumers are entirely rational and, in particular, are aware of their own inattention and plan accordingly.[7] The optimal strategy of an inattentive consumer is also to consume if and only if value v_t is above a threshold v^*, but an inattentive consumer's threshold is simply a constant, since it cannot be conditioned on t or q^{t-1}.

Formally, the consumer's decision problem exhibits Piccione and Rubinstein's (1997b) *absentmindedness*. Piccione and Rubinstein's (1997b) paradoxical absent minded driver example shows that analysis of such decision problems can be problematic, and there are different views on how to handle them (Piccione and Rubinstein 1997b, Piccione and Rubinstein 1997a, Gilboa 1997, Battigalli 1997, Grove and Halpern 1997, Halpern 1997, Lipman 1997, Aumann, Hart and Perry 1997a, Aumann, Hart and Perry 1997b). In particular, optimal strategies need not be time consistent. In this case, however, there is no problem.[8] Consumers' optimal thresholds from an

[7] 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^*. Assuming that consumers do not attend to the date makes the model more tractable but does not qualitatively change the primary welfare results. (See also footnote 10.)

[8] In Piccione and Rubinstein's (1997b) paradoxical absent minded driver example, time inconsistency arises because past decisions to exit or stay on the free-way determine whether or not the decision is faced again in the future. Thus

ex ante planning view point are time consistent and also optimal during execution. Hence the standard Bayesian Nash Equilibrium is an appropriate solution concept. Note that I assume that consumers plan ahead and choose a consumption strategy at the time they sign a contract. This rules out suboptimal equilibria that exist in the game modeled between multiple selves.[9]

A feasible inattentive strategy is a function $b(v_t)$ which describes a purchase probability for each valuation v_t to be implemented at all $t > 0$ independently of date or past usage. 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^*: $\mathbf{q}^I(\mathbf{v};\mathbf{p}) = (1_{v_1 \geq v^*}, 1_{v_2 \geq v^*})$. The optimal consumption threshold v^* is equal to the expected marginal price conditional on purchasing in the current period and satisfies the first order condition:*

$$v^* = \frac{p_1 + p_2}{2} + (1 - F(v^*))\, p_3. \tag{5}$$

Equation (5) 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 (5) has a unique solution and is sufficient as 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.

Proof. Assume that at the contracting stage a consumer plans to take strategy b^* but later considers a one time deviation to strategy b. At the planning stage, the consumer chooses b^* to maximize $U(b^*, b^*)$:

$$U(b^*, b^*) = v_0 - p_0 + 2\int_0^1 \left(v - \frac{p_1 + p_2}{2}\right) b^*(v)\, dF(v) - p_3 \left(\int_0^1 b^*(v)\, dF(v)\right)^2.$$

The plan is time consistent if, when considering a one time deviation to strategy b at the implementation stage, the resulting payoff $U(b^*, b)$ is maximized at $b = b^*$.

$$\begin{aligned} U(b^*, b) &= v_0 - p_0 + \int_0^1 \left(v - \frac{p_1 + p_2}{2}\right) b^*(v)\, dF(v) \\ &\quad + \int_0^1 \left(v - \frac{p_1 + p_2}{2}\right) b(v)\, dF(v) - p_3 \left(\int_0^1 b^*(v)\, dF(v)\right)\left(\int_0^1 b(v)\, dF(v)\right). \end{aligned}$$

simply arriving at a free-way exit is informative about which exit it is. In this paper, a consumer always has exactly two consumption choices to make, and hence being presented with a choice is entirely uninformative about past purchasing.

[9] 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 better equilibrium, which seems reasonable for family members who choose to setup joint accounts.

Inspection of the first order conditions for point-wise maximization at the planning and implementation stages,

$$\frac{dU(b^*, b)}{db(v)} = \frac{1}{2}\frac{dU(b^*, b^*)}{db^*(v)} = f(v)\left(v - \frac{p_1 + p_2}{2} - p_3 \int_0^1 b^*(v)\, dF(v)\right),$$

shows that the optimal strategy at the planning stagey is a threshold strategy satisfying equation (5) and that it is time consistent. A non-negative penalty fee is sufficient for $\frac{d}{dv^*}\left(\frac{1}{f(v^*)}\frac{dU(v^*)}{dv^*}\right) = -2(1 + f(v^*)\, p_3)$ to be strictly negative, which in turn is a sufficient second order condition for the consumer's maximization problem. ■

Note that given fixed prices and a positive penalty fee, equation (5) 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 Price Posting Regulation

Suppose that a firm faced some inattentive consumers and had the option either to disclose nothing or to make inattentive consumers attentive by disclosing the pair $\{t, q^{t-1}\}$ at the point of sale. I refer to the joint disclosure of $\{t, q^{t-1}\}$ as price-posting, since in this model it is equivalent to disclosing the date and the marginal price of the current unit.[10]

Definition 1 *Price-Posting Regulation (PPR) is the requirement that firms disclose* $\{t, q^{t-1}\}$ *at the point of sale.*

Note that in a richer model with more than two purchase opportunities, reporting the full purchase history q^{t-1} would require more than posting transaction prices at point of sale. However, firms commonly set prices only as a function of total purchases $\sum_{t=1}^{T} q_t$ rather than the full vector q^T.[11] In this case, disclosing total purchases to date rather than q^{t-1} is sufficient to make inattentive consumers attentive.

[10] I do not consider the possibility that firms might disclose t but conceal q^{t-1} or vice versa. This is purely for simplicity. Disclosing q^{t-1} without t leads consumers make inferences about t from q^{t-1}. A model in which consumers know t and are only inattentive to q^{t-1} has the following feature: When penalty fees are sufficiently high, a consumer who knew t but not q^{t-1} would choose different thresholds $v_1^* \neq v_2^*$ in each period. This would endogenously limit the size of penalty fees but would not qualitatively affect the primary pricing or welfare predictions.

[11] For instance, while cellular-phone service-providers typically differentiate between peak and off-peak calling, the total bill depends only on the total peak and total off-peak calling. It does not matter 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

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

Definition 2 *Constant-Marginal-Price Regulation (CMPR) 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 first model of biased beliefs in Section 5.1) 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 5.2 are shown only for price-posting regulation, the two regulations would have qualitatively similar effects in both the price discrimination and biased-beliefs models.

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 p_0 (balancing the trade-off between mark-up and volume in the standard way). As a result, neither inattention nor price-posting regulation have any substantive effect on market outcomes.

Proposition 2 *If consumers have homogeneous unbiased beliefs,* $v_t \sim F(v_t)$*, then there is a unique equilibrium outcome in which equilibrium allocations are efficient. If at least some consumers are attentive, then equilibrium contracts must offer marginal cost pricing (*$p_1 = p_2 = c$ *and* $p_3 = 0$*). If all consumers are inattentive, 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* $p \in [0, c]$*. Price-posting and constant-marginal-price regulations would both restrict equilibrium prices but have no effect on allocations, firm profits, or consumer surplus.*

The equivalence result in Proposition 2 captures the argument of some critics of price-posting regulation - that it would only cause firms to recoup lost penalty fees through fixed fees and other charges (Federal Reserve Board 2009a). However, the result relies heavily on the joint assumptions of homogeneity and correct beliefs. Further, Proposition 2's prediction that firms are indifferent to the use of penalty fees and disclosing marginal price at the point of sale appears inconsistent with firm behavior. In particular, Proposition 2 does not explain banks' choices not to voluntarily post transaction prices, by alerting consumers at the point of sale whether a given transaction

---

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 7 or Corollaries 2 and 3.

will result in an overdraft charge, nor their expressed aversion to regulatory intervention (Federal Reserve Board 2009a).[12] Similarly, Proposition 2 does not explain why cellular phone companies do not actively alert consumers to accruing overage charges, as the FCC is now considering making a requirement.

4 Price Discrimination Model

In this section, I relax the assumption of ex ante homogeneity imposed in the benchmark model and show that heterogeneity 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 price-posting regulation does affect substantive market-outcomes. In particular, price-posting regulation will be counter-productive in fairly competitive markets.

4.1 Model

Game players are mass 1 of consumers who have unbiased beliefs, but are heterogeneous ex ante, and $N \geq 1$ firms. At the contracting stage ($t = 0$), each consumer privately receives one of two private signals $s \in \{L, H\}$, where $\Pr(s = H) = \beta$. In addition, consumers privately learn a vector of N firm-specific taste shocks $\mathbf{x}$ that is mean zero conditional on s. Each firm i simultaneously offers a menu with a choice of two contracts, $s \in \{L, H\}$. 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, $t \in \{1, 2\}$, consumers privately learn a taste shock v_t, which measures a consumer's value for a unit of add-on service. Conditional on receiving signal s, a consumer's consumption taste shocks 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 (FOSD): $F_L(v) \geq F_H(v)$. After learning their taste shocks v_t, 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 $P^i(q_1, q_2; \hat{s})$ to firm i, as a function of past quantity choices and the chosen contract $\hat{s}$:

$$P^i(q_1, q_2; \hat{s}) = p^i_{0\hat{s}} + p^i_{1\hat{s}} q_1 + p^i_{2\hat{s}} q_2 + p^i_{3\hat{s}} q_1 q_2,$$

[12] 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 2009a). Clearly banks wanted to be able to make declining overdraft protection an expensive account feature.

characterized by the vector of prices $\mathbf{p}^i_{\hat{s}} = \left(p^i_{0\hat{s}}, p^i_{1\hat{s}}, p^i_{2\hat{s}}, p^i_{3\hat{s}}\right)$.

A consumer's base payoff 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 chosen contract $\hat{s}$:

$$u\left(\mathbf{q}, \mathbf{v}, \hat{s}\right) = v_0 + q_1 v_1 + q_2 v_2 - P^i(q_1, q_2; \hat{s}).$$

Conditional on signing a contract with prices $\mathbf{p}$, a consumer's optimal consumption strategy can be described by a function mapping valuations to quantity choices: $\mathbf{q}\left(\mathbf{v}; \mathbf{p}\right)$. The expected base 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 $U^i_{s\hat{s}} = E\left[u\left(\mathbf{q}\left(\mathbf{v}; \mathbf{p}^i_{\hat{s}}\right), \mathbf{v}, \hat{s}\right) \mid s\right]$. Define $U^i_s \equiv U^i_{ss}$ to be the expected base utility of a consumer who chooses the intended contract from firm i. Similarly, let $S_s = v_0 + E\left[\sum_{t=1}^{2} \left(v_t - c\right) q_t\left(\mathbf{v}; \mathbf{p}^i_s\right) \mid s\right]$ be the expected surplus from a consumer of type s who chooses contract s and makes optimal consumption choices at $t \in \{1, 2\}$.

A consumer's total expected payoff, $U^i_s + x^i$, includes brand taste x^i. Fraction $G_s\left(U^i_s; U^{-i}_s\right)$ of consumers of type s buy from firm i if firm i offers contract s with base expected utility of U^i_s, while competitors offer U^{-i}_s:

$$G_s\left(U^i_s; U^{-i}_s\right) = \Pr(U^i_s + x^i \geq \max_{j \neq i}\{U^j_s + x^j\}).$$

Firm profits per consumer are equal to payments less fixed costs (normalized to zero w.l.o.g.) and marginal cost $c \in [0, 1)$ per unit served. Thus, suppressing competitors offers U^{-i}_s and firm i superscripts from the notation, the firm's expected profit maximization problem is:

$$\begin{aligned} &\max_{\mathbf{p}_L, \mathbf{p}_H} \left(\left(1 - \beta\right) G_L\left(U_L\right)\left(S_L - U_L\right) + \beta G_H\left(U_H\right)\left(S_H - U_H\right)\right) \\ &\text{s.t. } U_s \geq U_{s\hat{s}} \;\; \forall s, \hat{s} \in \{L, H\}. \end{aligned}$$

This initial statement of the firm's problem encompasses both attentive and inattentive cases. They vary only by the consumers' optimal consumption rule $\mathbf{q}\left(\mathbf{v}; \mathbf{p}\right)$, which is given as a function of prices by equation (4) in the attentive case but by Proposition 1 in the inattentive case.

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 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 $\mu_s \equiv \left(S_s - U_s\right)$ are the firm's markup on each contract and the profit per customer served. Absent ex ante incentive constraints, the choice of markup would be a standard monopoly pricing problem.

I make one of two assumptions: (1) Zero outside option monopoly (ZOOM): $G_L(x) = G_H(x) = 1_{x\geq 0}$, which captures a monopolist serving customers with zero outside option. (2) Heterogeneous outside options (HOO): $G_s(x)$ 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. Define μ_s^* to be the optimal markup given first-best allocations and ignoring ex ante incentive constraints: $\mu_s^* = S_s^{FB} - \hat{U}_s$ where $\hat{U}_s \equiv \arg\max_U G_s(U)\left(S_s^{FB} - U\right)$. In the first case (ZOOM), $\hat{U}_s = 0$ and $\mu_s^* = S_s^{FB}$. In the latter case (HOO), $\mu_s^* = \frac{G_s(\hat{U}_s)}{g_s(\hat{U}_s)}$ where $\hat{U}_s$ uniquely satisfies $S_s^{FB} = \hat{U}_s + \frac{G_s(\hat{U}_s)}{g_s(\hat{U}_s)}$. Under ZOOM, $\mu_H^* > \mu_L^*$, and under HOO I will often focus on the case in which $\mu_H^* \geq \mu_L^*$. This is a natural assumption if high-average-value customers are high-income customers who have a lower marginal-value of money.

4.2 Attentive Case

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 et al. (2009). However, because I assume period-zero types are discrete rather than continuous, Pavan et al.'s (2009) results do not apply. Importantly 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.

Let $v_{1s\hat{s}}^*$ be the optimal first-period consumption-threshold of an attentive consumer of type s who chooses contract $\hat{s}$ and let $v_{1s}^* = v_{1ss}^*$. The expression for $v_{1s\hat{s}}^*$ is an extension of equation (3):

$$v_{1s\hat{s}}^* = p_{1\hat{s}} + \int_{p_{2\hat{s}}}^{p_{2\hat{s}}+p_{3\hat{s}}} (1 - F_s(v))\, dv. \tag{6}$$

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

$$\begin{aligned} U_{s\hat{s}} \quad = \quad & v_0 - p_{0\hat{s}} + \int_{v_{1s\hat{s}}^*}^{1} (v - p_{1\hat{s}})\, dF_s(v) \\ & + F_s(v_{1s\hat{s}}^*) \int_{p_{2\hat{s}}}^{1} (v - p_{2\hat{s}})\, dF_s(v) + (1 - F_s(v_{1s\hat{s}}^*)) \int_{p_{2\hat{s}}+p_{3\hat{s}}}^{1} (v - p_{2\hat{s}} - p_{3\hat{s}})\, dF_s(v), \end{aligned} \tag{7}$$

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

$$S_s = v_0 + \int_{v_{1s}^*}^{1} (v - c)\, dF_s(v) + \int_{p_{2s}+p_{3s}}^{1} (v - c)\, dF_s(v) + F_s(v_{1s}^*) \int_{p_{2s}}^{p_{2s}+p_{3s}} (v - c)\, dF_s(v). \tag{8}$$

It is useful to reframe the firm's problem in two ways. First, it is convenient to think of the firm choosing offered utility levels U_s rather than setting fixed fees p_{0s}. In this case the base fee p_{0s} is given by equation (7) evaluated at $\hat{s} = s$ as function of U_s. Second, it is convenient to think of the firm first choosing consumer's first period threshold v^*_{1s} rather than marginal price p_{1s}. Given a choice of v^*_{1s}, it is necessary for p_{1s} to satisfy equation (6) evaluated at $\hat{s} = s$. The firm's problem can be written as:

$$\max_{\substack{U_L, v^*_{1L}, p_{2L}, p_{3L} \\ U_H, v^*_{1H}, p_{2H}, p_{3H}}} \left((1-\beta) G_L(U_L)(S_L(v^*_{1L}, p_{2L}, p_{3L}) - U_L) + \beta G_H(U_H)(S_H(v^*_{1H}, p_{2H}, p_{3H}) - U_H) \right)$$
$$\text{s.t. } U_s \geq U_{s\hat{s}} \ \forall s, \hat{s} \in \{L, H\},$$

where $U_{s\hat{s}}$ and S_s are given by equations (7) and (8) and p_{1s} and p_{0s} are given by equations (6) and (7) evaluated at $\hat{s} = s$.

Proposition 3 characterizes the solution to a single firm's problem, and Proposition 4 applies the result to a Hotelling duopoly.

Proposition 3 *Given $U_s + \frac{G_s(U_s)}{g_s(U_s)}$ increasing and $c > 0$, there are three cases:*

(1) If $\mu^*_L = \mu^*_H$, then a single marginal cost contract, $P(q_1, q_2) = p_0 + c(q_1 + q_2)$, is offered and both types receive the first-best allocation.

(2) If $\mu^*_H > \mu^*_L$, then the high type receives the first-best allocation via marginal-cost pricing ($p_{3H} = 0$, $p_{1H} = p_{2H} = c$), while the low type's allocation is distorted downwards below first best. The penalty fee p_{3L} is strictly positive and allocations are strictly below first best: $v_{1L} > c$ and $p_{2L} > c$. The downward incentive constraint is binding, $U_H = U_{HL}$, while the upward incentive constraint can be ignored. The triple $\{v_{1L}, p_{2L}, p_{3L}\}$ must satisfy the following first order conditions:

$$v_{1L} = c + \int_{p_{2L}}^{p_{4L}} (v - c) f_L(v)\, dv + \frac{\beta}{1-\beta} \frac{-\partial\Pi/\partial U_H}{\beta G_L(U_L)} \frac{F_L(v_{1L}) - F_H(v_{1HL})}{f_L(v_{1L})}, \tag{9}$$

$$p_{2L} = c + \frac{\beta}{1-\beta} \frac{-\partial\Pi/\partial U_H}{\beta G_L(U_L)} \frac{F_H(v_{1HL})}{F_L(v_{1L})} \frac{F_L(p_{2L}) - F_H(p_{2L})}{f_L(p_{2L})}, \tag{10}$$

$$p_{2L} + p_{3L} = c + \frac{\beta}{1-\beta} \frac{-\partial\Pi/\partial U_H}{\beta G_L(U_L)} \frac{(1 - F_H(v_{1HL}))}{(1 - F_L(v_{1L}))} \frac{F_L(p_{2L} + p_{3L}) - F_H(p_{2L} + p_{3L})}{f_L(p_{2L} + p_{3L})}, \tag{11}$$

where v_{1HL} is given by equation (6).

(3) If $\mu^*_H < \mu^*_L$, then the low type receives the first-best allocation via marginal-cost pricing, while the high type's allocation is distorted upwards above first best. The fee p_{3H} is strictly negative, corresponding to a volume discount rather than a penalty fee, and allocations are strictly above first best: $v_{1H} < c$ and $p_{2H} < c$. The upward incentive constraint is binding, $U_L = U_{LH}$, while the downward incentive constraint can be ignored. The triple $\{v_{1H}, p_{2H}, p_{3H}\}$ must satisfy the following

first order conditions:

$$v_{1H} = c + \int_{p_{2H}}^{p_{4H}} (v - c) f_H(v)\, dv - \frac{1-\beta}{\beta} \frac{-\partial\Pi/\partial U_L}{(1-\beta) G_H(U_H)} \frac{F_L(v_{1LH}) - F_H(v_{1H})}{f_H(v_{1H})},$$

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

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

where v_{1LH} is given by equation (6).

To understand optimal pricing with attentive consumers characterized by case (2) of Proposition 3 for $\mu_H^* > \mu_L^*$ it is helpful to compare it to the optimal marginal price for ZOOM and $T = 1$ characterized by Courty and Li (2000):

$$p^{CL} = c + \frac{\beta}{1-\beta} \frac{F_L(p^{CL}) - F_H(p^{CL})}{f_L(p^{CL})}. \tag{12}$$

The first-order conditions for optimal second-period marginal prices characterized by equations (10)-(11) differ from equation (12) in two respects. First, there is an additional term,

$$\frac{-\partial\Pi/\partial U_H}{\beta G_L(U_L)} > 0,$$

which results from allowing for heterogeneous outside-options. This term is equal to 1 in the case of ZOOM. Second, equations (10)-(11) incorporate the additional terms $F_H(v_{1HL})/F_L(v_{1L}) < 1$ and $(1 - F_H(v_{1HL}))/(1 - F_L(v_{1L})) > 1$ respectively. These imply that $p_{3L} > 0$ and under ZOOM that $p_{2L} + p_{3L} > p^{CL} > p_{2L}$. Marginal prices are distorted upwards to discourage the high type from choosing the low contract. In the second period this distortion is more worthwhile after an initial purchase since a deviating high type is more likely to purchase in the first period than a low type. (In fact the additional terms simply adjust the likelihood ratio $\beta/(1-\beta)$ to condition on first period purchase information.) Hence a positive penalty fee is optimal.

The first-order condition for the optimal first-period threshold in equation (9) differs from equation (12) because of the heterogeneous-outside-options term and because of the positive penalty-fee potentially charged in period two. The cost of selling a unit in the first period is no longer simply the production cost c but also includes the lost surplus $\int_{p_{2L}}^{p_{4L}} (v - c) f_L(v)\, dv$ that results from making the penalty fee applicable in period 2.[13]

[13] A final difference is that $v_{1HL} \neq v_{1L}$. This is the reason that Pavan et al.'s (2009) approach does not work in

There are two features of the solution worth highlighting. First, Proposition 3 shows that optimal pricing for attentive consumers requires a positive penalty-fee or a volume discount and in general is not only a function of total usage since p_{1s} will differ from p_{2s} for one of the two contracts. Note that while the result can explain penalty fees it does not address the role of surprise penalty-fees.

The second feature of the attentive solution worth highlighting is that allocations are first best only when unconstrained optimal-markups are identical for low and high types. As Proposition 4 shows, this implies that allocations are only efficient in a Hotelling duopoly when both market segments have identical transportation costs.

Proposition 4 *Let duopolists with strictly positive marginal costs $c > 0$ compete on a uniform Hotelling line with transport costs τ_H and $\tau_L > 0$ for high and low types respectively. (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 tariff with fixed-fee markup of τ. (2) If $\tau_H \neq \tau_L$, then all equilibria are inefficient. (3) If $\tau_H > \tau_L$, then in all symmetric equilibria, high types receive first-best allocations, while low types' allocation is distorted downwards below first best. 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. In contrast, the striking result in the next section is that firms can charge different markups to different segments without distorting allocations if consumers are inattentive.

4.3 Inattentive case

I first solve the firm's problem assuming that the firm does not disclose $\left\{t, q^{t-1}\right\}$ to consumers. Later I consider whether nondisclosure is optimal.

---

this setting. Pavan et al. (2009) impose only incentive constraints that restrict one-step deviations and then check that the solution to this relaxed problem is also incentive compatible against multi-step deviations. However, the positive penalty-fee ensures that the multi-step deviation to $\{L, v_{1HL}\}$ is more tempting for the high type than the single step deviation to $\{L, v_{1L}\}$.

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 (5):

$$v^*_{s\hat{s}} = \frac{p_{1\hat{s}} + p_{2\hat{s}}}{2} + p_{3\hat{s}} \left(1 - F_s\left(v^*_{s\hat{s}}\right)\right). \tag{13}$$

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

$$U_{s\hat{s}} = v_0 - p_{0\hat{s}} + 2\int_{v^*_{s\hat{s}}}^1 v dF_s(v) - (p_{1\hat{s}} + p_{2\hat{s}})\left(1 - F_s\left(v^*_{s\hat{s}}\right)\right) - p_{3\hat{s}}\left(1 - F_s\left(v^*_{s\hat{s}}\right)\right)^2, \tag{14}$$

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

$$S_s = \int_{v^*_s}^1 (v - c)\, dF_s(v). \tag{15}$$

Define $\bar{p}_s = (p_{1s} + p_{2s})/2$. When consumers are inattentive, any pair $\{p_{1s}, p_{2s}\}$ which have the same average $\bar{p}_s$ are equivalent, both in terms of allocations and surplus division. I focus on symmetric pricing $p_{1s} = p_{2s} = \bar{p}_s$, for which the firm's problem reduces to the choice of p_{0s}, $\bar{p}_s$, and p_{3s} for $s \in \{L, H\}$. It is useful to reframe the firm's problem in two ways. First, it is convenient to think of the firm choosing offered utility levels U_s rather than setting fixed fees p_{0s}. In this case the base fee p_{0s} is given by equation (16):

$$p_{0s} = -U_s + v_0 + 2\int_{v^*_s}^1 v dF_s(v) - 2\bar{p}_s\left(1 - F_s\left(v^*_s\right)\right) - p_{3s}\left(1 - F_s\left(v^*_s\right)\right)^2. \tag{16}$$

Second, it is convenient to think of the firm first choosing consumer threshold v^*_s and then choosing the best marginal prices $\bar{p}_s$ and p_{3s} which implement v^*_s. Given any fixed choice of offered utility U_s and consumer threshold v^*_s, by Proposition 1 it is necessary[14] for $\bar{p}_s$ to satisfy the first order condition:

$$\bar{p}_s = v^*_s - p_{3s}\left(1 - F_s\left(v^*_s\right)\right). \tag{17}$$

The firm's problem can be written as:

$$\begin{aligned}
&\max_{\substack{U_L, v^*_L, p_{3L} \\ U_H, v^*_H, p_{3H}}} \left((1-\beta) G_L(U_L)\left(S_L\left(v^*_L\right) - U_L\right) + \beta G_H(U_H)\left(S_H\left(v^*_H\right) - U_H\right)\right) \\
&\text{s.t. } U_s \geq U_{s\hat{s}} \;\forall s, \hat{s} \in \{L, H\}, \\
&v^*_s \in \arg\max_x \left\{2\int_x^1 v f_s(v)\, dv - 2\bar{p}_s\left(1 - F_s(x)\right) - p_{3s}\left(1 - F_s(x)\right)^2\right\},
\end{aligned}$$

[14] Up to the fact that all thresholds above one are equivalent, and all thresholds below zero are equivalent.

where $U_{s\hat{s}}$, S_s, p_{0s}, and $\bar{p}_s$ are given by equations (14) through (17).

Notice that only offered utilities U_s and consumer thresholds v_s^* enter the objective function directly. Penalty fee p_{3s} only affects profits via the incentive constraints. The first order condition in equation (17) is sufficient for v_s^* to be incentive compatible for all $p_{3s} \geq 0$. Moreover, for any $v_s^* > 0$, increasing p_{3s} weakly relaxes both ex ante incentive incentive constraints, from which it follows that it is weakly optimal to set p_{3s} as large as possible.

Proposition 5 *Increasing p_{3s} weakly relaxes both ex ante incentive constraints. It is weakly optimal to choose non-negative penalties p_{3s} as large as possible.*

Proof. Substituting equations (16-17) into equation (14) yields

$$U_{s\hat{s}} = U_{\hat{s}} + 2\int_{v_{s\hat{s}}}^{1} (v - v_{\hat{s}})\, dF_s(v) - 2\int_{v_{\hat{s}}}^{1} (v - v_{\hat{s}})\, dF_{\hat{s}}(v) - p_{3\hat{s}} \left(F_{\hat{s}}(v_{\hat{s}}) - F_s(v_{s\hat{s}})\right)^2. \tag{18}$$

By the envelope condition:

$$\frac{d}{dp_{3\hat{s}}} U_{s\hat{s}} = \frac{\partial}{\partial p_{3\hat{s}}} U_{s\hat{s}} = -\left(F_{\hat{s}}(v_{\hat{s}}) - F_s(v_{s\hat{s}})\right)^2 \leq 0. \tag{19}$$

■

Proposition 5 suggests that the solution to the firm's problem could involve unreasonably high penalty fees. There are many forces which could endogenously limit penalty fees, some of which I discuss in Section 4.4. For simplicity, I exogenously impose one of two restrictions. Either I impose a cap on the penalty fees, or I require marginal prices to be non-negative. Both restrictions can be expressed as upper bounds on penalty fees: $p_{3s} \leq h_s(v_s)$. A cap on penalty fees corresponds to $h_s(v_s) = p^{\max} > 0$, while non-negative marginal prices correspond to $h_s(v_s) = v_s / (1 - F_s(v_s))$. Notice that all prior results and statements remain true with this addition to the problem.[15]

I solve the firm's problem separately for three cases. In each case I relax one or both ex ante incentive compatibility constraints and then confirm that the relaxed solution satisfies the ignored constraints and therefore solves the original problem. In the attentive problem, both ex ante incentive constraints can be relaxed and 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. This can be achieved because strictly-positive penalty-fees relax the ex ante incentive-constraints when consumers are inattentive.

[15] In particular, the constraint is symmetric such that any pair $\{p_{1s}, p_{2s}\}$ which have the same average $\bar{p}_s$ are still equivalent.

To state the proposition, first define $X_H \equiv 2\int_c^{v_{HL}} (v-c)\, dF_H(v) + (v_{HL}-c)(F_L(c) - F_H(v_{HL}))$ where v_{HL} uniquely satisfies $v_{HL} = c + h_L(c)(F_L(c) - F_H(v_{HL}))$ and $X_L \equiv 2\int_{v_{LH}}^{c} (c-v)\, dF_H(v) - (c - v_{LH})(F_L(v_{LH}) - F_H(c))$ where v_{LH} uniquely satisfies $v_{LH} = c - h_H(c)(F_L(v_{LH}) - F_H(c))$. Note that both X_L and X_H are strictly positive.

Proposition 6 *Assume (1) $U_s + \frac{G_s(U_s)}{g_s(U_s)}$ increasing, (2) the firm chooses not to disclose $\{t, q^{t-1}\}$, and (3) either penalty fees p_{3s} are exogenously restricted to be less than $p^{\max}$ ($h_s(v_s) = p^{\max}$), or marginal prices are exogenously restricted to be non-negative ($h_s(v_s) = v_s/(1-F_s(v_s))$). If $\mu_H^* \neq \mu_L^*$, then the firm offers a menu of two distinct contracts and sets at least one penalty fee strictly positive. If $\mu_H^* > \mu_L^*$, then any weakly positive penalty fee $p_{3H} \geq 0$ is optimal on the high contract, but the low contract must charge a strictly positive penalty fee $p_{3L} > 0$. The reverse is true for $\mu_H^* < \mu_L^*$. There are three cases: (1) If*

$$-X_L \leq \mu_H^* - \mu_L^* \leq X_H, \tag{20}$$

then both types receive the first-best allocation, $v_L^* = v_H^* = c$, and contract mark-ups are μ_L^* and μ_H^* respectively. (2) If $\mu_H^* - \mu_L^* > X_H$, then the high type receives the first-best allocation, $v_H^* = c$, and any weakly positive penalty fee $p_{3H} \geq 0$ is optimal on the high contract. However the downward incentive-constraint (IC-H) binds and the low-type's allocation is distorted downwards below first best: $v_L^* > c$. Moreover, the low type pays a strictly positive penalty fee $p_{3L} = h_L(v_L^*) > 0$ and v_L^* must satisfy the first order condition:

$$v_L = c + \frac{\beta}{1-\beta}\frac{F_L(v_L) - F_H(v_{HL})}{f_L(v_L)}\frac{-\partial\Pi/\partial U_H}{\beta G_L(U_L)}\left((1 + p_{3L} f_L(v_L)) + \frac{1}{2}(F_L(v_L) - F_H(v_{HL}))\, h_L'(v_L)\right), \tag{21}$$

where $v_{HL} = v_L + p_{3L}(F_L(v_L) - F_H(v_{HL}))$. (3) If $\mu_H^* - \mu_L^* < -X_L$, then the low type receives the first-best allocation $v_L^* = c$ and any weakly positive penalty fee $p_{3L} \geq 0$ is optimal on the low contract. However the upward incentive-constraint (IC-L) binds and the high type's allocation is distorted upwards above first best: $v_H^* < c$. Moreover, the high type pays a strictly positive penalty fee $p_{3H} = h_H(v_H^*) > 0$ and v_H^* must satisfy the first order condition:

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

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

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 equation (20) then nondisclosure is strictly optimal. Comparing Propositions 3 and 6 shows that 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 across different consumer segments are different, but not too different.

Suppose that $\mu_H^* > \mu_L^*$. If a firm were unconstrained by ex ante incentive-constraints (as in the case of third-degree price-discrimination between low and high types) it would always be optimal to offer both groups first-best allocations via marginal-cost pricing. The firm would then like to offer a discounted fixed-fee to low types. This is not feasible if consumers are attentive, as high types would always choose the discounted contract intended for low types. To satisfy the ex ante incentive-constraint and give low types a discount, the firm combines a discounted fixed-fee with high marginal prices that distort the low type's allocation downwards.

The striking result for inattentive consumers is that this is no longer the case for small discounts. By combining high penalty-fees with discounted fixed-fees, the firm can offer the low type a discounted markup and maintain incentive compatibility. By making the penalty fees surprise penalty-fees so that inattentive consumers respond to the expected marginal-price, the firm can avoid allocative distortions. This is possible because high types expect to pay penalty fees more often, and hence the firm can choose a high penalty-fee on the low contract and a low penalty-fee on the high contract such that both low and high types perceive expected marginal-price to equal marginal cost on their respective contracts.

Proposition 7 states the first of two main results in the paper. The combination of penalty fees and consumer inattention are socially valuable and price-posting regulation is counter-productive whenever markets are fairly competitive.

Proposition 7 *Let duopolists compete on a uniform Hotelling line, high types have transportation costs $\tau_H = \tau H$ strictly higher than low types $\tau_L = \tau L$, and marginal cost c be strictly positive. If $\tau > 0$ is sufficiently small, then: (1) In the unique (up to penalty fees) symmetric-pure-strategy 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 $p_{3s}^i = c/\left(1 - F_s\left(c\right)\right)$. (2) Price-posting regulation would strictly decrease welfare and firm profits. Low types would be losers while high types would be winners.*

Note that part (2) of Proposition 7 is also true for regulation banning penalty fees. See Appendix [to be completed].

The intuition behind the result in Proposition 7 that PPR is socially detrimental is as follows. Consider starting at the inattentive equilibrium and introducing PPR. At existing prices, PPR would cause the downward incentive-constraint to be violated, and firms could no longer charge markups that were so different. To restore incentive compatibility, firms would reduce markups on contract H, increase markups on contract L, and distort allocations on contract L downwards to reduce the need to adjust markups even further. The changes in markups drive the consumer surplus results, while the allocative distortion causes the reduction in social welfare. 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 expensive to raise markups on contract L in terms of market share.[16]

The FCC's proposed bill-shock regulation has strong support from consumer groups but is opposed by major cellular carriers (Genachowski 2010, Wyatt 2010). Proposition 7 easily explains industry opposition as it predicts lower industry profits. Support by consumer groups is less obvious since some consumers are made worse off by the policy. Note, however, that were prices held fixed, PPR would unambiguously help consumers. The fact that low types are worse off is due to the endogenous change in prices with regulation. The strong advocacy for the FCC's proposed bill-shock regulation by consumer groups (Genachowski 2010) is therefore not surprising (although misguided) if it is based on reasoning that fails to take into account future changes in prices.

In contrast with fairly-competitive markets, sufficient market power implies that penalty fees and inattentive consumers do not produce efficient outcomes. Corollary 1 illustrates this for the zero outside option monopoly.

Corollary 1 *Let the firm be a monopolist serving consumers with zero outside option and* $F_H < F_L$ *for all* $v \in (0,1)$ *(a strong form of strict FOSD). The upward ex ante incentive-constraint binds and the low-type's allocation is distorted below first best.*

Proof. By assumption, $G_s(U_s) = 1_{U_s \geq 0}$ and β is sufficiently small that it makes sense to serve the low types. (If not $v_L^* = 1 > c$ and the result is true as well). Hence, at the optimum, $G_L(U_L) = G_H(U_H) = 1$ and $\frac{-\partial\Pi/\partial U_H}{\beta G_L(U_L)} = 1$. When neither IC-L nor IC-H bind, $U_L = U_H = 0$. However, the high type can always mimic the low type by choosing contract L and a threshold v_{HL} such that $F_H(v_{HL}) = F_L(v_L)$. In this case, the high type makes the same expected payments and the same number of purchases, but at FOSD higher valuations. Thus $U_{HL} > U_L = U_H = 0$, which violates IC-H. ■

[16] Shifts in markups in each segment are already inversely weighted by shares of each segment β 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.

When there is sufficient market power the impact of regulation becomes ambiguous. Let the firm be a monopolist serving consumers with zero outside option. Without a binding revenue rasing requirement, a regulator with sufficient information and authority would optimally set marginal price equal to marginal cost to achieve efficient allocations. In this case inattention and price-posting regulation have no effect on outcomes. If a revenue raising requirement was binding, then a regulator setting optimal Ramsey prices would keep marginal prices hidden from inattentive consumers for the same reason an unregulated firm would: inattention allows revenues to be more efficiently extracted from high types. If a regulator is unable to directly regulate prices, but could require marginal prices to be posted at the time of transaction, such regulation may or may not be beneficial.

4.4 Constraints on penalty fees

Proposition 7 shows that, given sufficient competition, case (1) of Proposition 6 applies, ex ante incentive constraints are slack, and finite penalty fees are optimal. Thus with sufficient competition, restrictions on penalty fees do not bind, and the precise form of restriction does not matter. Hence Proposition 7 and the result it highlights – that in fairly-competitive markets the combination of surprise penalty-fees and consumer inattention can be socially valuable – are robust to a variety of restrictions on penalty fees.

When equation (20) isn't satisfied in equilibrium, then it is strictly optimal to set at least one penalty fee as high as possible. Without restriction this leads to the unreasonable prediction of negative infinity base marginal prices and positive infinity penalty fees. For simplicity and tractability, in the preceding analysis I imposed one of two exogenous constraints on penalty fees: either (a) that penalty fees must be below some exogenous upper bound $p^{\max}$, or (b) that marginal prices be non-negative. However, there are many natural economic forces absent from the model that would endogenously restrict penalty fees. This is particularly true because profits are bounded (strictly) below first-best surplus. Thus as penalty fees grow large, the remaining profit increase from increasing them all the way to infinity becomes arbitrarily small. Hence any arbitrarily small cost of raising penalty fees would be sufficient to endogenously limit penalty fees to finite levels.

Economic forces that would endogenously restrict penalty fees include: (1) Limited liability; (2) Mild consumer risk aversion; (3) A small risk of regulatory intervention that increases in the size of penalty fees; (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 their while; (6) Consumers who attend to the date and could condition v^* on the date.

(1) Limited liability restricts total price to always be below a consumer's wealth: $p_{0s} \leq W$,

$p_{0s} + p_{1s} \leq W$, $p_{0s} + p_{2s} \leq W$, and $p_{0s} + p_{1s} + p_{2s} + p_{3s} \leq W$. Combining equations (16) and (17), the last constraint imposes an upper bound on the penalty fee p_{3s} for any fixed v_s^* and U_s:

$$p_{3s} \leq h\left(v_s, U_s\right) = \frac{W + U_s - v_0 - 2\int_{v_s^*}^{1} v dF_s\left(v\right) - 2v_s^* F_s\left(v_s^*\right)}{2F_s^2\left(v_s^*\right)}.$$

(2) Low base-marginal-fees combined with high penalty-fees ensure that an inattentive consumer's bill is a lottery, the size of which will be limited by even mild risk-aversion. (3) Regulatory threat is self explanatory. (4) Any consumers who are attentive can ensure they purchase one and only one unit of the add-on service. If penalty fees are too high this will be costly to the firm, since (combining equations (16) and (17)) the firm would end up paying them a subsidy of at least[17]

$$-\left(p_{0s} + \bar{p}_s\right) = U_s - v_0 - 2\int_{v_s}^{1} \left(v - v_s\right) dF_s\left(v\right) - v_s + p_{3s}\left(1 - F_s\left(v_s\right)\right) F_s\left(v_s\right), \tag{23}$$

which for any fixed U_s and v_s is increasing linearly in the penalty p_{3s}. (5) Rational inattention limits penalty fees because increasing penalty fees increases consumers' returns to attention and thus the number of consumers who endogenously choose to be attentive. (6) Consumers who attend to the date would restrict penalty fees because if penalty fees were sufficiently high, a consumer who attended to the date would never buy in the first period, but always buy in the second (or vice-versa), thereby always avoiding the penalty fee, but receiving at least the subsidy in equation (23).

As already noted, the pricing predictions would be qualitatively robust under any of these modifications given strong competition, since for strong competition small penalty fees are sufficient. Clearly introducing additional players (attentive consumers), or costs (risk aversion) would affect welfare predictions, but only slightly if the additions are small. With sufficient market power, modifications (4) or (5) could qualitatively change pricing predictions by making asymmetric prices ($p_{1s} \neq p_{2s}$) optimal. This would be in response to the information asymmetry between periods in the attentive consumers' dynamic-programming problem. There would be no qualitative change in pricing predictions from modifications (2), (3), or (6). The limited liability constraint on penalty fees is relaxed when utility offers are increased, and hence would have an additional affect on markups (beyond the indirect affect via limiting penalty fees), but otherwise would not qualitatively affect pricing predictions.

An additional endogenous restriction on penalty fees would come from the existence of a large pool of attentive potential customers (or potential customers with a very low cost k of paying

[17]The subsidy would be higher if the firm chose asymmetric prices $p_{1s} \neq p_{2s}$.

attention) with zero value for the service. The existence of such potential customers imposes a no-arbitrage condition that I call the *no-free-lunch* (NFL) constraint. This restricts consumer payments to be non-negative at all allocations: $p_{0s} \geq 0$, $p_{0s} + p_{1s} \geq 0$, $p_{0s} + p_{2s} \geq 0$, and $p_{0s} + p_{1s} + p_{2s} + p_{3s} \geq 0$.[18] Otherwise, the large pool of attentive consumers with zero value for the product would purchase exactly the right quantity to get paid by the firm. This limits penalty fees, since holding v_s and U_s fixed, increasing p_{3s} towards infinity sends $p_{0s} + \bar{p}_s$ towards negative infinity (equation (23)). In fact, the NFL constraints $p_{0s} + p_{1s} \geq 0$ and $p_{0s} + p_{2s} \geq 0$ are equivalent to:

$$p_{3s} \leq h(v_s, U_s) = \frac{v_0 - U_s + 2\int_{v_s}^{1} (v - v_s)\, dF_s(v) + v_s}{(1 - F_s(v_s))\, F_s(v_s)}.$$

I explore the NFL constraint further (under the assumption that consumer beliefs are biased) in Section 5.2.

5 Biased Beliefs Model

The previous section showed that when consumers are inattentive, penalty fees may be used to more efficiently price discriminate between customer segments with stochastically low and high demand for an add-on good or service. This provides an explanation for two choices by cellular-phone companies: first to offer tariffs with steep penalty charges for high usage and second to avoid actively notifying consumers of the accruing charges until the end of the month. The analysis also suggested caution with respect to the bill-shock regulation under consideration by the FCC, since if the cellular market is sufficiently competitive (and consumers are unbiased) then actively notifying customers about accruing charges could undermine the social benefits of consumer inattention.

Unfortunately, the analysis in the previous section sheds no light on why, prior to the Federal Reserve Board's adoption of an opt-in rule, Bank of America and other banks charged high ($35) overdraft fees on debit and ATM transactions without notifying customers at the point of sale. Banks like Bank of America do price discriminate by offering multiple types of checking accounts with different terms into which different customer segments self select. However, Bank of America and others typically did not use overdraft charges as a tool to encourage self selection. On the contrary, the terms of overdraft charges were typically the same across different types of accounts. (For example, Figure 1 shows Bank of America's March 1st, 2010 menu of 4 types of checking

[18]Similar results would result 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.)

accounts and Figure 2 describes overdraft fees which were the same for all 4 types of checking accounts.)

This section explores an explanation for firms' valuation of penalty fees and consumer inattention that does apply to the case of overdraft fees: that consumers have biased beliefs and underestimate their consumption of the add-on good or service. As discussed in the introduction, this would arise in the context of overdraft fees if consumers are partially naive beta-delta discounters who consistently underestimate their spending. Consumer inattention may exacerbate or ameliorate allocative distortions created by biased beliefs. When marginal costs are extreme relative to the distribution of consumer valuations, inattention creates allocative distortions that are worse than those with biased beliefs alone, thereby lowering total welfare. When marginal costs are high, the allocative distortion is overconsumption and there are surplus reducing trades. For some intermediate marginal costs, inattention ameliorates distortions created by biased beliefs and price-posting regulation would reduce total welfare. However, the effect of first-order importance may be on surplus distribution rather than total surplus. Inattention means that consumers can be exploited and receive payoffs far below their outside options. Price-posting regulation ensures that consumers receive at least their outside option.

5.1 Continuous taste shocks and welfare

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 like the true distribution F is continuous and strictly increasing on $[0,1]$. Moreover, assume that F first-order-stochastically-dominates F^* so that consumers underestimate their demand for the add-on services.[19]

A consumer's true base expected payoff from contracting with firm i at the contracting stage and making optimal consumption choices thereafter remains

$$U^i = E\left[u\left(\mathbf{q}\left(\mathbf{v};\mathbf{p}^i\right),\mathbf{v}\right) \mid F\right] = \int_0^1 \int_0^1 u\left(\mathbf{q}\left(\mathbf{v};\mathbf{p}^i\right),\mathbf{v}\right) dF\left(v_1\right) dF\left(v_2\right).$$

However, a consumer's perceived expected payoff differs because expectations are taken with respect to consumer beliefs:

$$U^{*i} = E\left[u\left(\mathbf{q}\left(\mathbf{v};\mathbf{p}^i\right),\mathbf{v}\right) \mid F^*\right] = \int_0^1 \int_0^1 u\left(\mathbf{q}\left(\mathbf{v};\mathbf{p}^i\right),\mathbf{v}\right) dF^*\left(v_1\right) dF^*\left(v_2\right).$$

[19] To capture overconfidence with only two subperiods, consumers would need to underestimate the correlation in v_t across periods.

The fraction $G\left(U^{*i};U^{*-i}\right)$ of consumers of type s who buy from firm i depends on the perceived base-expected-utility offers of firms rather than the true expected-utilities:

$$G\left(U^{*i};U^{*-i}\right)=\Pr(U^{*i}+x^{i}\geq\max_{j\neq i}\{U^{*j}+x^{j}\}).$$

Thus firm i's expected profits are

$$\Pi^{i}=G\left(U^{*i};U^{*-i}\right)E\left[P^{i}\left(\mathbf{q}\left(\mathbf{v};\mathbf{p}\right)\right)-c\left(q_{1}\left(\mathbf{v};\mathbf{p}\right)+q_{2}\left(\mathbf{v};\mathbf{p}\right)\right)\mid F\right],$$

which can be rewritten in terms of total surplus and consumers' true and perceived expected-utilities:

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

5.1.1 Attentive Case

If attentive consumers underestimate their demand for the service ex ante, then we know that firms have an incentive to set marginal charges above marginal cost, irrespective of competition (e.g. Grubb (2009)). This is reflected in Proposition 8, which characterizes optimal pricing in the attentive case.[20]

Proposition 8 *If all consumers are attentive and homogeneously underestimate demand, then the optimal contract is a two-part tariff ($p_3=0$, $p_1=p_2=p$) with marginal price*

$$p=c+\frac{F^{*}\left(p\right)-F\left(p\right)}{f\left(p\right)},$$

and profits

$$\Pi=G\left(U^{*}\right)\left(-U^{*}+2\int_{p}^{\infty}\left(v-c-\frac{F^{*}\left(v\right)-F\left(v\right)}{f\left(v\right)}\right)f\left(v\right)dv\right). \tag{24}$$

All consumers are weakly better off than choosing their outside options, and all transactions generate positive surplus. If $F\left(c\right)<F^{*}\left(c\right)$ (demand underestimation is strict at $p=c$) then marginal price is above marginal cost and allocations are inefficiently low.

Proposition 8 shows the potential for biased beliefs to reduce welfare in the absence of inattention by distorting consumption downwards. The intuition is that the firm is limited in how much surplus it can extract ex ante through a fixed fee since consumers underestimate their value for

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

the service. The firm must wait until consumers draw high value-realizations and extract surplus through distortionary marginal-charges. Proposition 8 also points out that when attentive consumers underestimate their value for a good or service they cannot be exploited (they must receive at least their outside option) and there are no surplus-reducing trades.

5.1.2 Inattentive case

Now consider the inattentive case. The consumption threshold chosen by an inattentive consumer with biased beliefs satisfies the first order condition,

$$v^* = \frac{p_1 + p_2}{2} + p_3 \left(1 - F^* \left(v^*\right)\right) \tag{25}$$

which substitutes consumer beliefs in place of the true distribution of tastes in equation (5). As before, I focus on symmetric pricing $p_1 = p_2 = \bar{p}$ and it is useful to reframe the firm's problem in two ways. First, it is convenient to think of the firm choosing perceived expected-utility U^* rather than setting fixed fee p_0. In this case the fixed fee p_0 is given by equation (26):

$$p_0 = -U^* + v_0 + 2 \int_{v^*}^{1} v dF^* \left(v\right) - 2\bar{p} \left(1 - F^* \left(v^*\right)\right) - p_3 \left(1 - F^* \left(v^*\right)\right)^2 . \tag{26}$$

Second, it is convenient to think of the firm first choosing consumer threshold v^* and then choosing the best marginal prices $\bar{p}$ and p_3 which implement v^*. Given any fixed choice of perceived expected-utility U^* and consumer threshold v^*, by Proposition 1, it is necessary for $\bar{p}$ to satisfy the first order condition:

$$\bar{p} = v^* - p_3 \left(1 - F^* \left(v^*\right)\right) . \tag{27}$$

Using equations (26) and (27), firm profits can be written as a function of perceived expected-utility U^*, penalty p_3, and consumer threshold v^*:

$$\Pi = G\left(U^*\right) \left(-U^* + 2 \int_{v^*}^{\infty} \left(v - c - \frac{F^*\left(v\right) - F\left(v\right)}{f\left(v\right)} \right) f\left(v\right) dv + p_3 \left(F^*\left(v^*\right) - F\left(v^*\right)\right)^2 \right) . \tag{28}$$

Comparing equations (28) and (24) shows that the firm can make strictly higher profits charging a surprise penalty-fee to inattentive consumers than by selling to attentive consumers. Since profits increase linearly in the penalty fee p_3, the optimal penalty fee will be positive and the local incentive constraint of equation (27) is sufficient for v^* to be globally optimal. Moreover, without any additional constraints, firms optimally choose $p_3 = \infty$ and $v^* \in (0, 1)$. This contract transfers infinite wealth from consumers to the firm. Infinite penalty fees are implausible because many forces will restrict the size of penalty fees in practice, as discussed in Section 4.4. An important difference with biased beliefs is that the returns to increasing penalty-fees are constant rather than decreasing. Thus a fraction of consumers who are attentive would still endogenously restrict penalty fees, but only if the fraction were sufficiently large. For simplicity I impose a maximum penalty fee $p^{\max}$. Since the optimal penalty fee is $p^{\max}$, the firm's problem can then be written as:

$$\max_{U^*,v^*} G(U^*)\left(-U^* + 2\int_{v^*}^{\infty}\left(v - c - \frac{F^*(v) - F(v)}{f(v)}\right) f(v)\,dv + p^{\max}\left(F^*(v^*) - F(v^*)\right)^2\right)$$

Equation (28) shows that for any fixed finite-penalty-fee $p^{\max}$, profits are increasing in the size of the disagreement between consumer and firm about the consumer's per period purchase probability $|F^*(v^*) - F(v^*)|$. Given a restriction on penalty fees, firms have an incentive to adjust consumers' threshold choice v^* to increase this disagreement. In general, the incentive to maximize disagreement could increase or decrease distortions relative to the attentive case. For some intermediate marginal-costs, maximizing disagreement ameliorates inefficiency and price-posting regulation reduces welfare. However, if marginal costs are extreme relative to consumer valuations, then maximizing disagreement entails increased inefficiency and price-posting regulation increases welfare (Proposition 9). For example, if marginal cost is zero, then $F^*(c) = F(c) = 0$ and at marginal-cost pricing both firm and consumers agree that the consumer always purchases. Hence attentive pricing is at marginal cost, but with inattentive consumers firms raise the expected marginal price v^* above zero to create exploitable disagreement.

To state the next result, I parameterize consumers' degree of bias. Let F and $\hat{F}$ have full support with continuous densities on $[0,1]$ that cross finitely many times, $F < \hat{F}$ for all $v \in (0,1)$ (a strong form of strict FOSD), 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 γ goes to zero.

Proposition 9 *Assume penalty fees are exogenously restricted by an upper bound $p^{\max} > 0$. If the bias is sufficiently small (γ is sufficiently close to zero holding F and $\hat{F}$ fixed) then: (1) When marginal cost $c \geq 0$ is close to zero, inattention exacerbates underconsumption ($v^* > v^A > c$). (2) There exists an intermediate marginal cost $c \in (0,1)$ for which inattention ameliorates overconsumption ($v^A > v^* \geq c$). In addition, if the maximum penalty fee $p^{\max}$ is sufficiently large then: (3) When marginal cost is close to 1, inattention creates overconsumption worse than the attentive underconsumption ($v^* < c \leq v^A$). In cases (1) and (3) price-posting regulation strictly improves welfare, while in case (2) it strictly reduces welfare. Constant-marginal-price regulation would have an identical effect.*

Note that Proposition 9 points out that when consumers are inattentive, consumers who underestimate their demand for a product or service may be induced to overconsume. For instance when c is slightly above 1, all product sales are inefficient. Yet because inattentive consumers underestimate their likely values for the product, sales take place. Price posting regulation would increase consumer surplus and total welfare by ending sales of these products. This is potentially why Bank of America chose to stop offering overdraft protection on debit-card transactions following the Federal Reserve's new 'opt-in' requirement, 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).

5.2 Bernoulli taste shocks and surplus distribution

The effects of inattention and price-posting regulation on total welfare may in fact be second order relative to their effects on the distribution of surplus. In some situations, the welfare effects are likely to be small, for instance because costs and values are similar or small, or because valuations have a concentrated distribution. But more importantly, since inattentive consumers who underestimate their demand can be exploited, surplus distribution effects of price-posting regulation are not limited by first-best surplus but can be orders of magnitude higher.

To focus on distributional issues, I make an alternative assumption about the distribution of taste shocks. For the rest of the paper, assume taste shocks have a Bernoulli distribution: v_t are drawn independently and are equal to 1 with probability α and zero otherwise. Consumers underestimate their demand and believe that v_t equals 1 with probability $\alpha' < \alpha$.[21] Also assume $c \in (0, 1)$. Finally, rather than exogenously imposing an upper bound on penalty fees or imposing that marginal prices be non-negative, I will endogenously restrict penalty fees by imposing the no-free-lunch constraint.

5.2.1 Attentive Case

To solve the firm's problem in both attentive and inattentive cases I proceed in two steps. First, I fix a perceived utility U^* to be offered and solve for the optimal price vector and allocation which implements U^* subject to the NFL constraint. The price vector and allocation determine expected surplus S and true expected-utility U. Hence the first step derives an optimal markup, $\mu(U^*)$, to be charged as a function of U^*. The second step is to choose the perceived expected-utility U^* which maximizes profits, $\Pi = G(U^*)\,\mu(U^*)$, subject to feasibility under NFL.

[21] This might arise in the context of checking-account overdraft-fees if: (1) purchases occur with probability α, (2) consumers value making purchases with a checking-account debit-card a fixed amount $v = 1$ over an alternative form of payment, and (3) consumers underestimate the likelihood of purchases because they are naive beta-delta discounters.

When consumers are attentive, the firm finds it optimal to induce the efficient allocation and charge a nonnegative penalty fee.

Lemma 1 *Given Bernoulli taste shocks, attentive consumers who underestimate demand ($\alpha' < \alpha$), $c \in (0,1)$, and the no-free-lunch constraint, firms set prices which induce the efficient allocation: consumers buy if and only if $v_t = 1$. Moreover firms charge a nonnegative penalty fee $p_3 \geq 0$.*

Following Lemma 1 and conditional on a level of perceived utility U^* to be offered, the firm's problem reduces to choosing the vector of prices which maximize the expected markup μ subject to a set of constraints. For the attentive consumers problem, it is useful to omit the penalty fee p_3 and work with the price vector $\{p_0, p_1, p_2, p_4\}$ where $p_4 = p_2 + p_3$ is the marginal price for a second period purchase conditional on a first period purchase. The constraints are the NFL constraints, the constraint that the efficient allocation be incentive compatible, and the offered utility constraint that U^* is in fact as specified. Note that the two NFL constraints $p_0 + p_2 \geq 0$ and $p_0 + p_1 + p_4 \geq 0$ are redundant given the IC constraints $p_2, p_4 \geq 0$. Thus the firm's problem is:

$$\max_{p_0,p_1,p_2,p_3} \left(S^{FB} - U\right)$$

1. NFL: $p_0 \geq 0,\ p_0 + p_1 \geq 0,$
2. Offered Utility: $U^* = v_0 - p_0 + 2\alpha' - \alpha' \left(p_1 + p_2 + \alpha' \left(p_4 - p_2\right)\right),$
3. IC: $p_2, p_4, v_1^* \in [0,1]\,,\ v_1^* = p_1 + \alpha' \left(p_4 - p_2\right),$

where the true and perceived expected utilities are:

$$U = v_0 - p_0 + 2\alpha - \alpha \left(p_1 + p_2 + \alpha \left(p_4 - p_2\right)\right), \tag{29}$$

$$U^* = v_0 - p_0 + 2\alpha' - \alpha' \left(p_1 + p_2 + \alpha' \left(p_4 - p_2\right)\right). \tag{30}$$

Since the firm's objective and constraints are linear in prices, it is possible to solve the firm's problem as follows. Begin by setting the fixed fee sufficiently high that perceived expected utility is too low and setting marginal charges as high as possible under incentive constraints: $p_1 = p_2 = p_4 = 1$. This contract is a two-part tariff with no penalty fees ($p_3 = 0$) which satisfies IC and NFL constraints but offers too little perceived expected utility. Thus the firm must reduce prices. For $U^* = v_0 + 2\alpha'$, there is only one feasible option, which is to set all prices to zero: $p_0 = p_1 = p_2 = p_4 = 0$. (Higher perceived utility offers are not feasible under NFL.) However, for lower perceived utility offers the firm has discretion over which prices to reduce and by how much.

The firm would like to raise the perceived expected utility as cheaply as possible. That is, the firm would like to begin by reducing prices which will reduce the expected markup least for a given increase in perceived expected utility. If consumers were unbiased, it would not matter which prices were reduced. In all cases a price reduction that raised perceived expected utility by a dollar would lower the expected markup by a dollar as well. However, because consumers underestimate their demand this is no longer true. Instead consumers are more sensitive to some prices than others.

I calculate the *bang-for-the-buck* of independently decreasing price p_n as $\gamma_n \equiv -\frac{dU}{dp_n}/\frac{dU^*}{dp_n}$, which measures the decrease in markup for a one dollar increase in U^* due to decreasing p_n. Computing derivatives from equations (29)-(30), these bang-for-the-buck coefficients are: $\gamma_0 = -1$, $\gamma_1 = -\alpha/\alpha'$, $\gamma_2 = -\alpha\left(1-\alpha\right)/\left(\alpha'\left(1-\alpha'\right)\right)$, and $\gamma_4 = -\alpha^2/\alpha'^2$. These can be ranked as follows:

$$\gamma_0, \gamma_2 > \gamma_1 > \gamma_4.$$

Note that the ranking of γ_0 and γ_2 depends on whether $(\alpha + \alpha')$ is less than or greater than one. Reducing the price with a higher bang for the buck (e.g. p_1 has a higher bang-for-the-buck than p_4) will result in a smaller cut in markup for the same increase in perceived expected-utility U^*. Thus, absent a binding constraint, the firm will always reduce the price with the highest bang-for-the-buck first.

The ranking $\gamma_0 > \gamma_1 > \gamma_4$ is relatively straight forward. There is no confusion about the fixed fee - consumers and the firm both agree that a dollar reduction in p_0 is a transfer of a dollar from firm to consumers. However, consumers underestimate their likelihood of purchase in the first period by a factor α'/α. Thus they undervalue reductions in p_1 by the same factor, and the expected markup falls at rate α/α' faster than perceived utility rises. The problem is compounded for p_4, since consumers underestimate the chance of making two purchases by $\left(\alpha'/\alpha\right)^2$. On the other hand, the problem 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. Hence $\gamma_2 > \gamma_1 > \gamma_4$.

An important point is that p_2 has a higher bang-for-the-buck than p_1 and p_4 (and also than p_0 if $\alpha + \alpha' > 1$). However it cannot be independently reduced before p_1 and p_4 without violating the incentive constraint $v_1^* \leq 1$. Either p_1 or p_4 must be reduced at the same time to maintain incentive compatibility. It is shown in the proof of Proposition 10 that after reducing fixed fees to zero, the highest bang-for-the-buck is achieved by simultaneously reducing p_1 and p_2 such that the incentive constraint $v_1^* \leq 1$ is binding. This leads to the following conclusion. It is optimal for the firm to reduce prices in the following order (stopping as soon as the prescribed U^* is achieved): (1) First reduce the fixed fee p_0 until U^* is achieved or $p_0 = 0$. (2) Second, reduce p_2 and p_1 simultaneously such that $p_1 = 1 - \alpha' (1 - p_2)$ until U^* is achieved or $p_2 = 0$. (3) Third, reduce p_1 until U^* is achieved or $p_1 = 0$ (at which point the NFL constraint $p_0 + p_1 \geq 0$ binds because $p_0 = 0$). (4) Finally the firm should reduce p_4 until U^* is achieved. This procedure stops at the optimal contract conditional on U^*.

This program of price reduction leads to four qualitative pricing regions described in Proposition 10. The cost of raising the offered perceived-expected-utility ($-d\mu/dU^*$) is increasing across the four pricing regions as the firm begins cutting prices that consumers are less and less sensitive to.

Proposition 10 *Given Bernoulli taste shocks, attentive consumers who underestimate demand ($\alpha' < \alpha$), $c \in (0, 1)$, and the no-free-lunch constraint: Firms offer the first-best allocation and charge nonnegative penalty-fees. Consumers are not exploited: $U \geq 0$. Moreover, there are four qualitative pricing regions as a function of the offered perceived expected-utility U^*: (1) $U^* \in [0, v_0]$, (2) $U^* \in [v_0, v_0 + \alpha']$, (3) $U^* \in [v_0 + \alpha', v_0 + \alpha' (2 - \alpha')]$, (4) $U^* \in [v_0 + \alpha' (2 - \alpha'), v_0 + 2\alpha']$. ($U^* > v_0 + 2\alpha'$ is not feasible given the no-free-lunch constraint.) Prices and in each region are summarized in the following table:*

Region	p_0	p_1	p_2	p_3
1	$v_0 - U^*$	1	1	0
2	0	$1 - \alpha' (1 - p_2)$	$1 - (U^* - v_0) / \alpha'$	$1 - p_2$
3	0	$2 - \alpha' - (U^* - v_0) / \alpha'$	0	1
4	0	0	0	$2/\alpha' - (U^* - v_0) / \alpha'^2$

Corresponding markups for each pricing region are given in the appendix. Importantly, consumer price sensitivity is declining across the four regions, as measured by the increasing cost of raising perceived-expected utility $-d\mu/dU^*$:

Region	*1*	*2*	*3*	*4*
$-d\mu/dU^*$	$1 <$	$(\alpha/\alpha') (1 - \alpha + \alpha') <$	$(\alpha/\alpha') <$	$(\alpha/\alpha')^2$

Without the NFL constraint, the firm would always offer a two-part tariff with marginal price of 1 as in pricing region (1) described in Proposition 10. In duopoly competition on a uniform Hotelling line with full market coverage this would result in firms competing on fixed fees and charging markups equal to the transportation cost τ. This will still be the result with the NFL constraint under weak competition when firms offer low perceived-expected-utilities and the equilibrium is in region 1. However in more competitive markets after firms drop fixed fees to zero the NFL constraint forces them to compete first on base marginal charges (regions 2 and 3) and finally on penalty fees (region 4). This progressively softens price competition and firms charge markups above τ. Thus increasing competition is partially mitigated by reduced consumer price sensitivity.

Proposition 11 *Assume duopoly competition on a uniform Hotelling line, the no-free-lunch constraint, Bernoulli taste shocks, consumers who underestimate demand ($\alpha' < \alpha$) and $c \in [0, 1)$. Let transportation cost τ be sufficiently small or v_0 be sufficiently large that the market is fully covered.*[22] *There are four competitive regions corresponding to the four pricing regions in Proposition 10 over which markups are proportional to τ at rate $-d\mu/dU^*$. Markups are constant in τ over the gaps between the four regions.*

Region	$\tau_{\min}$	$\tau_{\max}$	*markup* μ	*Competition*
1	$2\alpha(1-c)$		τ	*Fixed fees*
2	$\frac{2\alpha'(1-c)}{1-\alpha+\alpha'} - \alpha'$	$\frac{2\alpha'(1-c)}{1-\alpha+\alpha'}$	$(\alpha/\alpha')(1-\alpha+\alpha')\tau$	*Base marginal charges*
3	$\alpha'(\alpha-2c)$	$\alpha'(\alpha-2c)+\alpha'(1-\alpha')$	$(\alpha/\alpha')\tau$	*Base marginal charges*
4	0	$\left(\alpha'^2/\alpha\right)(\alpha-2c)$	$(\alpha/\alpha')^2\tau$	*Penalty fees*

Duopoly profits equal the markup and consumers' true expected utility is $U = S^{FB} - \mu \geq 0$.

The top panel of Figure 3 illustrates Proposition 11 by plotting the equilibrium markup as a function of the transportation cost τ. The 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 τ within pricing regions and then constant between regions. Starting at the right hand side of the figure, and working leftward as τ falls and competition increases, equilibrium begins in pricing region 1 where firms compete on fixed fees. After fixed fees have reached zero, consumers become discontinuously less price sensitive and markups are temporarily flat until equilibrium transitions to pricing region 2 where firms begin competing on base marginal charges. 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.[23]

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. In particular,

---

[22] *To illustrate how the nature of competition changes over a wide range of transportation costs, assume that for any transportation cost τ under consideration that v_0 is sufficiently large for there to be full market coverage. This means that the value of the base good is sufficiently high that in equilbirium a firm's marginal consumer always strictly prefers the competition to the outside good.*

[23] Note that the figure is shown for $c = 0$. 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.

between pricing regions where markups are constant, increasing competition increases the gap between the markup and τ. Since the markup would always be τ if consumers were unbiased this means the cost of their bias is increasing in competition in these regions. This contradicts a common intuition that increased competition reduces the importance of policy interventions to correct consumer biases.

5.2.2 Inattentive case

If consumers are inattentive, firms charge positive penalty fees but still induce efficient consumption. Thus total surplus is first best irrespective of inattention, price-posting regulation, or consumers' biased beliefs. The distribution of surplus, however, varies significantly with inattention and price-posting regulation.

Given Bernoulli taste shocks, an inattentive consumer's strategy is described by the pair $\{b_0, b_1\}$. These are the probabilities of purchase conditional on realizing $v_t = 0$ or $v_t = 1$ respectively: $b_0 = \Pr(q_t(v_t = 0) = 1)$ and $b_1 = \Pr(q_t(v_t = 1) = 1)$. A consumer's perceived expected-utility U^* is given by equation (31) as a function of prices and the strategy $\{b_0, b_1\}$:

$$U^*(b_0, b_1) = -p_0 + v_0 + 2\left(1-\alpha'\right) b_0 (-\bar{p}) + 2\alpha' b_1 (1-\bar{p}) - \left(\left(1-\alpha'\right) b_0 + \alpha' b_1\right)^2 p_3. \tag{31}$$

Firm profits, as a function of prices, perceived expected-utility U^*, and the allocation $\{b_0, b_1\}$ are given by equation (32):

$$\Pi = G(U^*)\Big(p_0 + 2\left((1-\alpha) b_0 + \alpha b_1\right)(\bar{p} - c) + \left((1-\alpha) b_0 + \alpha b_1\right)^2 p_3\Big). \tag{32}$$

The first result is that it will be optimal for firms to set prices which induce the efficient allocation $\{b_0, b_1\} = \{0, 1\}$.

Lemma 2 *Given Bernoulli taste shocks, inattentive consumers who underestimate demand ($\alpha' < \alpha$), $c \in (0,1)$, and the no-free-lunch constraint, firms set prices which induce the efficient allocation: consumers buy if and only if $v_t = 1$.*

To induce the efficient allocation, the firm must set expected marginal price conditional on a purchase to be between zero and one: $0 \leq \bar{p} + \alpha' p_3 \leq 1$. Applying Lemma 2, the firm's problem can thus be reduced to the following:[24]

$$\begin{aligned}
\max_{U^*,\bar{p},p_3} \Pi &= G(U^*)\left(p_0 + 2\alpha(\bar{p} - c) + \alpha^2 p_3\right) \\
\text{such that} &: \\
\text{IC:} & \quad 0 \leq \bar{p} + \alpha' p_3 \leq 1 \\
\text{NFL:} & \quad p_0 \geq 0,\ p_0 + \bar{p} \geq 0,\ p_0 + 2\bar{p} + p_3 \geq 0 \\
\text{Fixed Fee} &: \quad p_0 = -U^* + v_0 + 2\alpha'(1 - \bar{p}) - \alpha'^2 p_3
\end{aligned}$$

Proposition 12 characterizes optimal prices given a fixed perceived-expected-utility U^*. For low utility offers, the NFL constraint $p_0 + \bar{p} \geq 0$ and the IC constraint $\bar{p} + \alpha' p_3 \leq 1$ both bind. For medium utility offers, the two NFL constraints $p_0 \geq 0$ and $p_0 + \bar{p} \geq 0$ both bind. Higher utility offers above $v_0 + 2\alpha'$ are not feasible given the NFL constraint. To state the proposition, define the notation

$$Y \equiv \frac{(\alpha - \alpha')^2}{\alpha'(1 - \alpha')}. \tag{33}$$

Proposition 12 *Given Bernoulli taste shocks, inattentive consumers who underestimate demand ($\alpha' < \alpha$), $c \in (0,1)$, and the NFL constraint: Firms offer the first-best allocation and charge nonnegative surprise-penalty-fees, preferring not to disclose them at the point of sale (a strict preference for offers U^* below $v_0 + \alpha'(2 - \alpha')$). Moreover, there are two qualitative pricing regions as a function of the offered perceived expected-utlity U^*: (1) Conditional on offering $U^* \in [0, v_0 + \alpha']$, optimal prices are*

$$p_1 = p_2 = -p_0 = -\frac{v_0 + \alpha' - U^*}{1 - \alpha'},\ p_3 = \frac{v_0 + 1 - U^*}{(1 - \alpha')\alpha'}. \tag{34}$$

(2) Conditional on offering $U^* \in [v_0 + \alpha', v_0 + 2\alpha']$, optimal prices are:

$$p_1 = p_2 = p_0 = 0,\ p_3 = \left(2\alpha' + v_0 - U^*\right)/\alpha'^2. \tag{35}$$

(Offering $U^* > v_0 + 2\alpha'$ is not feasible under NFL..) Corresponding markups for each pricing region are given in the appendix. Importantly, consumer price sensitivity declines across the two regions, as measured by the increasing cost of raising perceived-expected utility $-d\mu/dU^*$:

Region	*1*	*2*
$-d\mu/dU^*$	$-(1+Y) <$	$(\alpha/\alpha')^2$

[24] It is strictly optimal to set prices symmetrically, $p_1 = p_2 = \bar{p}$, since keeping $\bar{p}$ constant but setting $p_1 < p_2$ would tighten the NFL constraint $p_0 + p_1 \geq 0$ without otherwise effecting consumer incentives or firm profits. Similarly, setting $p_2 < p_1$ would tighten the NFL constraint $p_0 + p_2 \geq 0$.

Together, Propositions 10 and 12 shows that to offer $U^* \in [v_0 + \alpha'(2-\alpha'), v_0 + 2\alpha']$, optimal contracts are the same whether consumers are attentive or inattentive and involve only one positive fee - a penalty fee $p_3 \in [0,1]$ On the other hand, to make a strictly lower offer $U^* \in [0, v_0 + \alpha'(2-\alpha'))$, firms would charge inattentive consumers a penalty fee $p_3 > 1$ and earn a higher markup than were consumers attentive. The difference arises because attentive consumers would never pay a penalty fee above 1. For high U^* offers this doesn't matter because the penalty fee must be below 1 to achieve such high U^*. For lower U^* offers the constraint binds when consumers are attentive reducing margins and profitability.

Propositions 10 and 12 characterize optimal prices and markup $\mu(U^*)$ as a function of perceived expected-utility U^*. Corollary 2 applies Propositions 10 and 12 to a zero-outside-option monopoly for which the optimal utility offer is $U^* = 0$. The result compares attentive and inattentive cases and evaluates the effect of price-posting regulation:

Corollary 2 *Assume a zero-outside-option monopoly, the no-free-lunch constraint, Bernoulli taste-shocks, consumers who underestimate demand ($\alpha' < \alpha$), and $c \in (0,1)$. If consumers are attentive, the monopolist charges $p_0 = v_0$, $p_1 = p_2 = 1$, and $p_3 = 0$, induces efficient consumption, and captures the full surplus ($\Pi = S^{FB}$, $CS = 0$). Let $Y \equiv (\alpha - \alpha')^2 / (\alpha'(1-\alpha'))$. If consumers are inattentive, the monopolist charges $p_1 = p_2 = -p_0 = -(v_0 + \alpha')/(1-\alpha')$ and $p_3 = (v_0+1)/(\alpha'(1-\alpha'))$. While still inducing efficient consumption, the monopolist now captures more than the entire first-best surplus ($\Pi = S^{FB} + (1+v_0)Y$) and consumers are exploited, receiving less than their outside option ($CS = -(1+v_0)Y < 0$). Price posting regulation does not affect total welfare, but redistributes $(1+v_0)Y$ from firm to consumers and eliminates consumer exploitation.*

Proof. A direct application of Propositions 10 and 12 given that the optimal utility offer is $U^* = 0$ given ZOOM. ■

Note that my choice of the no-free-lunch constraint, rather than an alternative restriction on penalty fees, does not qualitatively effect the results in Corollary 2, only the magnitude of the shift in surplus $(1+v_0)Y$ would vary with alternative constraints. The assumption has a more substantive role in competitive markets however. For instance, with a simple upper bound of $p^{\max}$ imposed on penalty fees, the redistributive effects of price-posting regulation would vanish with Hotelling competition, because additional profits extracted from inattentive consumers through penalty fees would be rebated through fixed fees due to competition. Firms would always earn markup τ regardless of price-posting regulation. Analysis of the attentive case, however, shows that the no-free-lunch constraint softens competition leading to markups above τ 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, price-posting regulation can have a substantive effect on market outcomes even in the competitive case.

Proposition 13 *Assume duopoly competition on a uniform Hotelling line, the no-free-lunch constraint, Bernoulli taste shocks, consumers who underestimate demand ($\alpha' < \alpha$) and $c \in [0, 1)$. Let transportation cost τ be sufficiently small or v_0 be sufficiently large that the market is fully covered. Firms charge surprise-penalty-fees preferring not to disclose $\{q^{t-1}, t\}$ (a strict preference for $\tau > (\alpha')^2 (\alpha - 2c)/\alpha$). There are two competitive regions corresponding to the two pricing regions in Proposition 12 over which markups are proportional to τ at rate $-d\mu/dU^*$. Markups are constant in τ over the gap between the two regions.*

Region	$\tau_{\min}$	$\tau_{\max}$	*markup* μ	*Competition*
1	$(2\alpha(1-c) + Y)/(1+Y) - \alpha'$		$(1+Y)\tau$	*All fees*
2	0	$(\alpha'/\alpha)(\alpha - 2c\alpha')$	$(\alpha/\alpha')^2 \tau$	*Penalty fees*

Duopoly profits equal the markup and consumers' true expected utility is $U = S^{FB} - \mu \geq 0$.

The bottom panel of Figure 3 illustrates Proposition 13 by plotting the equilibrium markup as a function of the transportation cost τ. Dashed lines show the markups relevant for the two possible pricing regions as well as the markup τ that would prevail in the absence of bias or the no-free-lunch constraint The solid bold line shows the equilibrium markup, which is increasing in τ within pricing regions and then constant between regions. Starting at the right hand side of the figure, and working leftward as τ falls and competition increases, equilibrium begins in pricing region 1 where firms compete on a mixture of fees. After all fees excluding the penalty fee have reached zero, consumers become discontinuously less price sensitive and markups are temporarily flat until equilibrium transitions to pricing region 2 where firms begin competing on penalty fees alone.

A sufficient condition for full market coverage assumed in Propositions 11 and 13 is $\tau < 2v_0/3$. Comparing Propositions 11 and 13 for $\tau \in (0, 2v_0/3)$ uncovers the effects of price-posting regulation under competition. Without price-posting 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). The important effect of price-posting 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 price-posting regulation is implemented. For any higher level of market power ($\tau > (\alpha')^2 (\alpha - 2c) / \alpha$), however, price-posting 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*[25] ($\alpha'/\alpha < \max\{1/2, (2\alpha - 1)/\alpha^2\}$) but the reverse can be true for intermediate values of τ given *mild bias* ($\alpha'/\alpha > \max\{1/2, (2\alpha - 1)/\alpha^2\}$) as made precise in parts 2 and 3 of Corollary 3.[26] The comparison is illustrated for severe and mild biases in top and bottom panels of Figure 4. Part 1 of Corollary 3 holds $\tau > 0$ fixed and shows that sufficiently large bias leads to arbitrarily high markups and consumers 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 3 *Assume duopoly competition on a uniform Hotelling line, the no-free-lunch 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 PPR.*

1. *For fixed $\tau > 0$, if bias is sufficiently large (α'/α is sufficiently small) then all consumers are exploited. Price-posting 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 PPR 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 PPR effects markups as described for severe bias except for intermediate transportation costs within the interval $\tau \in [\tau_1, \tau_2]$, where $\tau_1 = (1 - \alpha + \alpha') / (\alpha - 2c\alpha')$ and $\tau_2 = 2\alpha (1 - c) / (1 + Y)$. For $\tau \in (\tau_1, \tau_2)$, PPR strictly increases markups.*

[25] The term "severe bias" is somewhat misleading. As α approaches 1, a belief $\alpha' < \alpha$ arbitrarily close to α will satisfy the condition for "severe bias".

[26] When bias is mild and $\tau \in (\tau_1, \tau_2)$, PPR 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 PPR regulation.

Corollaries 2 and 3 capture the second main result in the paper – that, combined with biased beliefs, inattention can cause consumers to receive payoffs far below their outside option and that price-posting regulation will eliminate this exploitation. In the monopoly setting this is a direct result of the fact that price-posting regulation constrains the size of penalty fees - precisely those fees which consumers most misestimate the chance of paying. In a competitive setting the result is more indirect. Absent additional constraints on prices, total markups would equal τ independent of the fraction earned from penalty fees. However, the no-free-lunch constraint ensures that sufficiently competitive firms compete on marginal charges rather than fixed fees. This leads to markups above τ because consumers are less price sensitive to marginal charges, which they underestimate the likelihood of paying. Price-posting regulation limits the extent that this competition is over penalty fees, and forces firms to compete on more salient base marginal charges intensifying competition and protecting consumers.

6 Conclusion

If consumers have unbiased beliefs, but have heterogeneous forecasts of their future demand for an add-on good or service, the combination of consumer inattention and penalty fees can help firms price discriminate between customer segments with stochastically low and high demand forecasts. Price-posting regulation, by providing inattentive consumers with the same information recalled by attentive consumers, can help consumers avoid penalty fees. While this is good for consumers holding prices fixed, it undermines the value of penalty fees and will cause firms to change their prices. When firms have substantial market power, it is ambiguous whether this will increase or decrease total welfare. In fairly competitive markets, however, price-posting 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.

The model provides an explanation for two facts about cellular-phone service in the US. First, customers are 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. Second, firms do not actively alert customers to accruing charges prior to the end of the month. If one believes that cellular phone customers have correct beliefs and the cellular market is sufficiently competitive, then the FCC's considered bill-shock regulation, which requires carriers to alert consumers to rapidly accruing charges, would be counterproductive. However, Grubb (2009) and Grubb and Osborne (2010) present compelling evidence that cellular phone customers have biased beliefs about their likely usage. Moreover it is not clear how competitive the market for cellular-phone service is. As a result, the welfare impact of price-posting regulation is ambiguous and caution should be applied in adopting the FCC's considered bill-shock regulation.

When consumers underestimate their demand for an add-on service or good, the combination of consumer inattention and penalty fees can be highly profitable for firms. In fact, they can enable firms to earn more in profit than the entire social surplus from a transaction and even profit from selling a product with negative social value. In these cases consumers are exploited in the sense that they are worse off than had they never done business with the firm. It is ambiguous whether price-posting regulation would increase or decrease welfare, but such changes in total welfare could be overshadowed by much larger changes in the distribution of surplus and the elimination of consumer exploitation. In both monopoly and competitive markets, price-posting regulation eliminates consumer exploitation and can increase consumer surplus by orders of magnitude more than the entire social surplus of the transaction.

This is one explanation for the high revenues ($20Bn in 2009) from overdraft charges for ATM and one-time debit-card transactions, which is consistent with the fact that Bank of America cancelled its $2.2Bn service when required by the Federal Reserve Board to ask consumers to opt-in (Martin 2010, Sidel and Fitzpatrick 2010). Moreover, it suggests that he Federal Reserve Board's regulation will substantially benefit consumers, and that banks will not be able to recoup the lost overdraft revenue simply by raising monthly fees on accounts. (Although they may of course find other equally profitable penalty fees to exploit.) It also suggests that the bill-shock regulation under consideration by the FCC could have substantial benefits for consumers, in particular as applied to fees such as roaming charges, which typically are the same across calling plans and are not used for purposes of price discrimination.

7 Figures

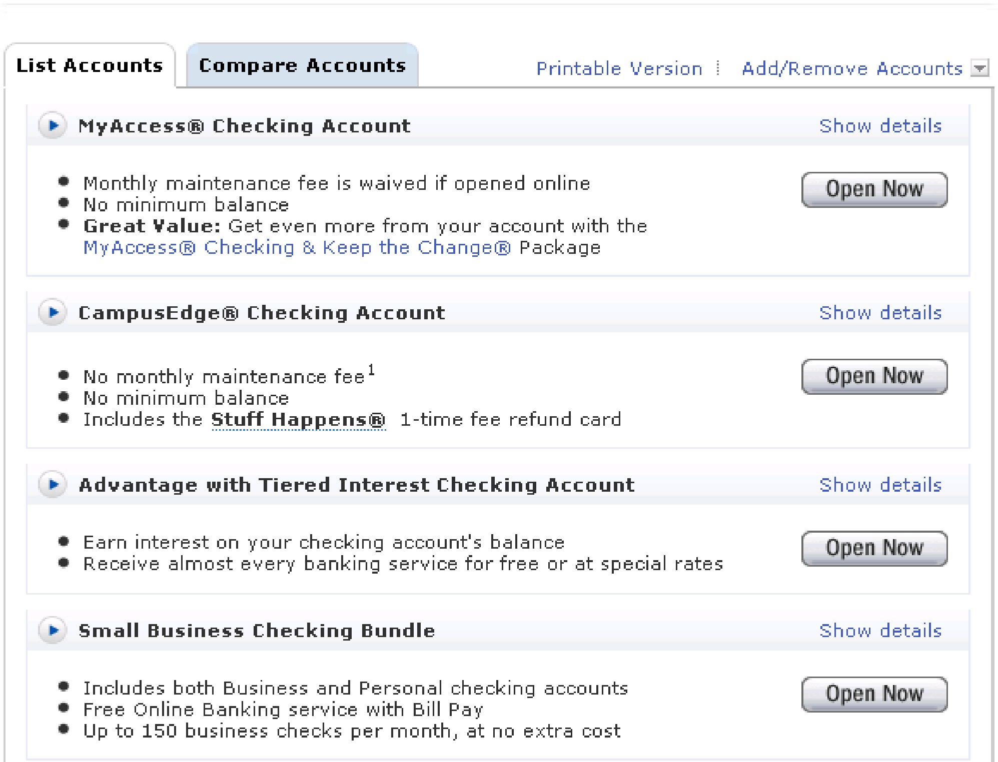

Figure 1: Bank of America's menu of 4 checking accounts, offered online at www.bankofamerica.com on March 1, 2010.

Fee Category	Fee Name / Description	Fee Amount	Accounts Qualifying for Waiver of this Fee
Overdraft Items (an overdraft item)	Overdraft Item Fee	$35.00 each item	N/A
	NSF: Returned Item Fee	$35.00 each item	N/A
	Extended Overdrawn Balance Charge	$35.00 – charged when we determine your account is overdrawn for 5 or more consecutive business days.	N/A

Figure 2: The overdraft fees associated with Bank of America's checking accounts shown in Figure 1. They are the same across all accounts. Source www.bankofamerica.com, March 1, 2010.

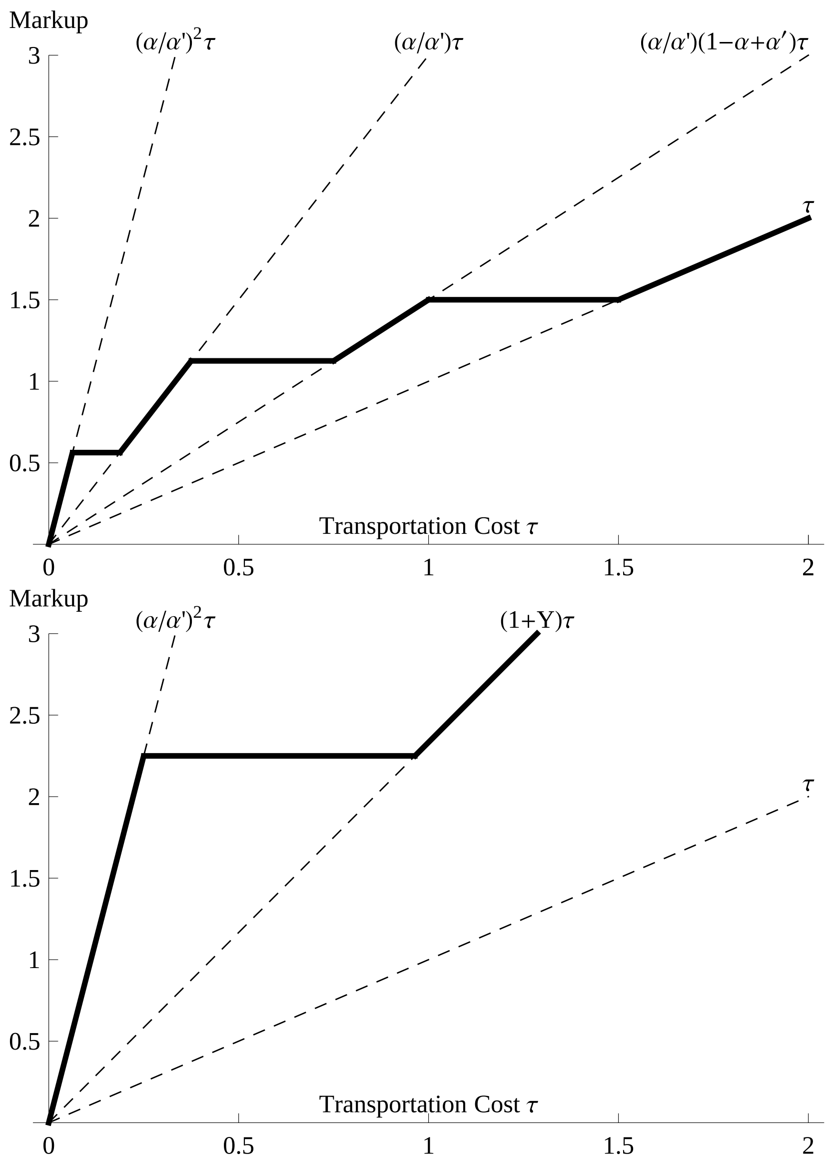

Figure 3: Firm markup as a function of transportation cost τ in a Hotelling duopoly with the no-free-lunch constraint and consumers who receive Bernoulli taste shocks and underestimate demand ($\alpha' < \alpha$). Top panel: attentive consumers. Bottom panel: inattentive consumers. The figure is plotted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

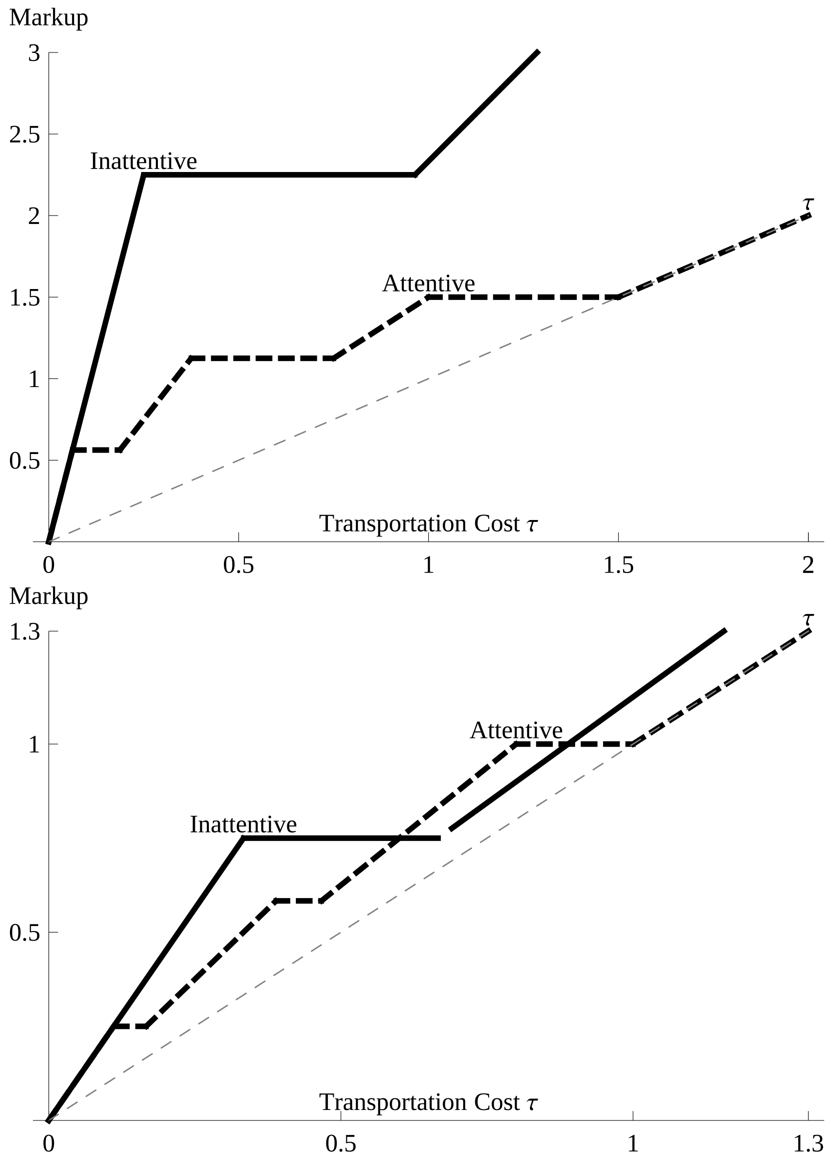

Figure 4: Firm markup as a function of transportation cost τ in a Hotelling duopoly with the no-free-lunch constraint and consumers who receive Bernoulli taste shocks and underestimate demand ($\alpha' < \alpha$). Solid line: attentive consumers. Dashed line: inattentive consumers. Top panel depicts 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 full market coverage.

References

Agarwal, Sumit, John C. Driscoll, Xavier Gabaix, and David I. Laibson, "Learning in the Credit Card Market," *NBER Working Paper*, 2008, (W13822). http://ssrn.com/paper=1093660.

Armstrong, Mark and John Vickers, "Competitive Price Discrimination," *RAND Journal of Economics*, 2001, *32* (4), 579–605.

Aumann, Robert J., Sergiu Hart, and Motty Perry, "The Absent-Minded Driver," *Games and Economic Behavior*, 1997, *20* (1), 102–116.

___ , ___ , and ___ , "The Forgetful Passenger," *Games and Economic Behavior*, 1997, *20* (1), 117–120.

Ausubel, Lawrence M. and Haiyan Shui, "Time inconsistency in the credit card market," Working Paper 2005.

Bank of America, "Personal Schedule of Fees," Technical Report March 1 2010. https://www3.bankofamerica.com/efulfillment/documents/91-11-3000ED.20100201.htm.

Baron, David P. and David Besanko, "Regulation and Information in a Continuing Relationship," *Information Economics and Policy*, 1984, *1* (3), 267–302.

Battigalli, Pierpaolo, "Dynamic Consistency and Imperfect Recall," *Games and Economic Behavior*, 1997, *20* (1), 31–50.

Borenstein, Severin, "To What Electricity Price Do Consumers Respond? Residential Demand Elasticity Under Increasing-Block Pricing," Preliminary Draft April 30 2009.

Broadbent, Donald E., *Perception and Communication*, New York: Pergamon Press, 1958.

Bubb, Ryan and Alex Kaufman, "Consumer Biases and Firm Ownership," Working Paper 2009.

Bulow, Jeremy I., John D. Geanakoplos, and Paul D. Klemperer, "Multimarket Oligopoly: Strategic Substitutes and Complements," *The Journal of Political Economy*, 1985, *93* (3), 488–511.

Cardon, James H. and Igal Hendel, "Asymmetric Information in Health Insurance: Evidence from the National Medical Expenditure Survey," *The RAND Journal of Economics*, 2001, *32* (3), 408–427.

Courty, Pascal and Hao Li, "Sequential Screening," *The Review of Economic Studies*, 2000, *67* (4), 697–717.

Dash, Eric and Nelson D. Schwartz, "Banks Seek to Keep Profits as New Oversight Rules Loom," July 15 2010.

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.

—— **and** —— , "Paying Not to Go to the Gym," *The American Economic Review*, 2006, *96* (3), 694–719.

Diamond, Peter A., "A model of price adjustment," *Journal of Economic Theory*, 1971, *3* (2), 156–168. doi: DOI: 10.1016/0022-0531(71)90013-5.

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, "Comment Sought on Measures Designed to Assist U.S. Wireless Consumers to Avoid "Bill Shock"," Public Notice May 11 2010. CG Docket No. 09-158, http://hraunfoss.fcc.gov/edocs_public/attachmatch/DA-10-803A1.pdf.

Federal Reserve Board, "Federal Register notice: Regulation E final rule," Technical Report November 11 2009.

—— , "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.

Gabaix, Xavier and David Laibson, "Shrouded Attributes, Consumer Myopia, and Information Suppression in Competitive Markets," *Quarterly Journal of Economics*, 2006, *121* (2), 505–540.

Gaynor, Martin S., Yunfeng Shi, Rahul Telang, and William B. Vogt, "Cell Phone Demand and Consumer Learning - An Empirical Analysis," *SSRN eLibrary*, 2005.

Genachowski, Julius, "Prepared Remarks of Chairman Julius Genachowski," 2010. http://reboot.fcc.gov/fcc-s-consumer-empowerment-agenda.

Gilboa, Itzhak, "A Comment on the Absent-Minded Driver Paradox," *Games and Economic Behavior*, 1997, *20* (1), 25–30.

Grove, Adam J. and Joseph Y. Halpern, "On the Expected Value of Games with Absentmindedness," *Games and Economic Behavior*, 1997, *20* (1), 51–65.

Grubb, Michael D., "Selling to Overconfident Consumers," *American Economic Review*, 2009, *99* (5), 1770–1807.

—— **and Matthew Osborne**, "Cellular service demand: Tariff choice, usage uncertainty, biased beliefs, and learning," Working Paper 2010.

Hadar, Josef and William R. Russell, "Rules for Ordering Uncertain Prospects," *American Economic Review*, 1969, *59* (1), 25.

Halpern, Joseph Y., "On Ambiguities in the Interpretation of Game Trees," *Games and Economic Behavior*, 1997, *20* (1), 66–96.

Huang, Ching-I., "Estimating Demand for Cellular Phone Service Under Nonlinear Pricing," *Quantitative Marketing and Economics*, 2008, *6* (4), 371–413.

Laibson, David, "Golden Eggs and Hyperbolic Discounting," *The Quarterly Journal of Economics*, 1997, *112* (2), 443–477.

Lambrecht, Anja, Katja Seim, and Bernd Skiera, "Does Uncertainty Matter? Consumer Behavior under Three-Part Tariffs," *Marketing Science*, 2007, *26* (5), 698–710.

Liebman, Jeffrey B. and Richard Zeckhauser, "Schmeduling," Working Paper October 2004.

Lipman, Barton L., "More Absentmindedness," *Games and Economic Behavior*, 1997, *20* (1), 97–101.

Martin, Andrew, "Bank of America to End Debit Overdraft Fees," March 10 2010.

McAfee, R. Preston and Vera L. te Velde, "Dynamic Pricing in the Airline Industry," in T.J. Hendershott, ed., *Handbook on Economics and Information Systems*, Elsevier Handbooks in Information Systems 2007.

Miravete, Eugenio J., "Screening Consumers Through Alternative Pricing Mechanisms," *Journal of Regulatory Economics*, 1996, *9* (2), 111–132.

—— , "The Welfare Performance of Sequential Pricing Mechanisms," *International Economic Review*, 2005, *46* (4), 1321–1360.

O'Donoghue, Ted and Matthew Rabin, "Choice and Procrastination," *Quarterly Journal of Economics*, 2001, *116* (1), 121–160.

Pavan, Alessandro, Ilya Segal, and Juuso Toikka, "Dynamic Mechanism Design: Incentive Compatibility, Profit Maximization and Information Disclosure," Working Paper 2009.

Piccione, Michele and Ariel Rubinstein, "The Absent-Minded Driver's Paradox: Synthesis and Responses," *Games and Economic Behavior*, 1997, *20* (1), 121–130.

—— **and** —— , "On the Interpretation of Decision Problems with Imperfect Recall," *Games and Economic Behavior*, 1997, *20* (1), 3–24.

Reiss, Peter C. and Matthew W. White, "Household Electricity Demand, Revisited," *The Review of Economic Studies*, 2005, *72* (3), 853–883.

Riordan, Michael H. and David E. M. Sappington, "Awarding Monopoly Franchises," *American Economic Review*, 1987, *77* (3), 375–387.

Rochet, Jean-Charles and Lars A. Stole, "Nonlinear Pricing with Random Participation," *The Review of Economic Studies*, 2002, *69* (1), 277–311.

Saez, Emmanuel, "Do Taxpayers Bunch at Kink Points?," Working Paper June 2002.

Sandroni, Alvaro and Francesco Squintani, "Overconfidence, Insurance and Paternalism," *American Economic Review*, 2007, *97* (5), 1994–2004.

Sidel, Robin and Dan Fitzpatrick, "End Is Seen to Free Checking," June 16 2010.

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," July 2010.

Stole, Lars A., "Price Discrimination and Competition," in Mark Armstrong and Robert K. Porter, eds., *Handbook of Industrial Organization*, Vol. Volume 3, Elsevier, 2007, pp. 2221–2299.

Weyl, Eric G. and Michal Fabinger, "Pass-Through as an Economic Tool," *SSRN eLibrary*, 2009.

Wilson, Robert B., *Nonlinear Pricing*, New York, NY: Oxford University Press, 1993.

Wyatt, Edward, "F.C.C. Wants to Stop Cellphone Bill Shock," October 13 2010.

A Proofs

A.1 Derivation of equation (2)

Given $v_2^* = p_2 + q_1 p_3$, the expected utility from choosing first period threshold v_1^* is:

$$U(v_1^*) = v_0 - p_0 + \int_{v_1^*}^{1} \left(v_1 - p_1 + \int_{p_2+p_3}^{1} (v_2 - p_2 - p_3) f(v_2) \, dv_2 \right) f(v_1) \, dv_1 + F(v_1^*) \int_{p_2}^{1} (v_2 - p_2) f(v_2) \, dv_2.$$

The first order condition,

$$\frac{dU}{dv_1^*} = f(v_1^*) \left(-v_1^* + p_1 + \int_{p_2}^{p_2+p_3} (v_2 - p_2) f(v_2) \, dv_2 + (1 - F(p_2 + p_3)) p_3 \right) = 0,$$

yields equation (2). Moreover, this identifies the global maximum since for $v_1^* > p_1 + \int_{p_2}^{p_2+p_3} (v_2 - p_2) f(v_2) \, dv_2 + (1 - F(p_2 + p_3)) p_3$, $\frac{dU}{dv_1^*} < 0$ and vice-versa.

A.2 Proof of Proposition 2

Firm profits can be written as $\Pi = G(U)(S - U)$. For any fixed utility offer U, profits are maximized by choosing marginal prices p_1, p_2, and p_3 to achieve first-best surplus, while adjusting the fixed fee p_0 to keep U constant. The offered utility U is set via the fixed fee p_0 to balance rent extraction versus participation, as in a basic monopoly pricing problem. Given attentive consumers and continuous taste shocks, $p_1 = p_2 = c$ and $p_3 = 0$ are the unique marginal prices which achieve S^{FB}. Given inattentive consumers and continuous taste shocks, any marginal prices which implement $v^* = c$ are optimal. These include all marginal prices which satisfy $p_3 \geq 0$ and equation (5) at $c = v^*$ since equation (5) is sufficient as well as necessary for incentive compatibility given $p_3 \geq 0$.

A.3 Proof of Proposition 3

Note, in the proof I write the firm's problem as a choice of marginal prices p_{2s} and p_{4s} where $p_{4s} = p_{2s} + p_{3s}$ rather than p_{2s} and the penalty fee p_{3s}.

I. First consider half the parameter space: $\mu_H^* > \mu_L^*$. Relax the upward incentive constraint $U_L \geq U_{LH}$ (IC-L).

(1). IC-L slack implies marginal cost pricing ($v_{1H}^* = p_{2H} = c$ and $p_{3H} = 0$) and first-best allocation for the high type ($q_t(H, v^t) = q^{FB}(v_t)$). Proof: Suppose not. Then setting $\{v_{1H}^*, p_{2H}, p_{4H}\}$ equal to $\{c, c, c\}$ while keeping U_H constant keeps IC-H and participation unaffected without violating IC-L since it has been relaxed. However, it increases surplus and hence profit from type H - a contradiction.

(2). The upward incentive constraint $U_H \geq U_{HL}$ (IC-H) binds. Moreover, it will bind with equality given either ZOOM (where $\partial\Pi/\partial U_H = -\beta$ for all $U_H > 0$) or HOO (where decreasing marginal revenue assumption, $U_s + \frac{G_s(U_s)}{g_s(U_s)}$ increasing, implies concavity). (Note that if IC-H it were relaxed then there would be marginal cost pricing $P_s(q_1, q_2) = p_{0s} + c(q_1 + q_2)$ such that $U_H = S_H^{FB} - p_{0H}$ and $U_{HL} = S_H^{FB} - p_{0L}$. Thus IC-H would be equivalent to $S_H^{FB} - p_{0H} \geq S_H^{FB} - p_{0L}$ which implies $S_H^{FB} - p_{0H} \geq S_L^{FB} - p_{0L}$, or $\mu_H^* \leq \mu_L^*$ at optimal offers $\{U_H, U_L\}$ which is a contradiction.)

(3). The downward incentive constraint (IC-H) is $U_H \geq U_{HL}$ which is convenient to re-express as $U_H \geq U_L + (U_{HL} - U_L)$. Let $Z = (U_{HL} - U_L)$. Equation (7) can be simplified by integrating by parts to:

$$U_{s\hat{s}} = v_0 - p_{0\hat{s}} + \int_{v_{1s\hat{s}}^*}^1 (v - p_{1\hat{s}})\, dF_s(v) + F_s(v_{1s\hat{s}}^*) \int_{p_{2\hat{s}}}^1 (1 - F_s(v))\, dv + (1 - F_s(v_{1s\hat{s}}^*)) \int_{p_{4\hat{s}}}^1 (1 - F_s(v))\, dv. \tag{36}$$

Thus the expression for Z can be re-written as:

$$\begin{aligned} Z = & \int_{v_{HL}}^1 (v - p_1) f_H(v)\, dv - \int_{v_L}^1 (v - p_{1L}) f_L(v)\, dv \\ & + \int_{p2L}^1 (F_H(v_{HL})(1 - F_H(v)) - F_L(v_L)(1 - F_L(v)))\, dv \\ & + \int_{p4L}^1 ((1 - F_H(v_{HL}))(1 - F_H(v)) - (1 - F_L(v_L))(1 - F_L(v)))\, dv \end{aligned} \tag{37}$$

where from equation (6) evaluated at $\hat{s} = s = L$:

$$p_{1L} = v_{1L} - \int_{p_{2L}}^{p_{4L}} (1 - F_L(v))\, dv.$$

Given (1) and (2), the firm's problem can be reduced to:

$$\max_{U_L, v_L^*, p_{2L}, p_{3L}} \left\{ \begin{array}{c} (1-\beta) G_L(U_L)(S_L(v_L^*, p_{2L}, p_{4L}) - U_L) \\ +\beta G_H(U_H(U_L, v_L^*, p_{2L}, p_{4L}))\left(S_H^{FB} - U_H(U_L, v_L^*, p_{2L}, p_{4L})\right) \end{array} \right\}$$

where $U_H(U_L, v_L^*, p_{2L}, p_{4L}) = U_L + Z(v_L^*, p_{2L}, p_{4L})$.

By the envelope condition, for any $x \in \{v_L^*, p_{2L}, p_{4L}\}$

$$\frac{dU_H}{dx} = \frac{\partial U_H}{\partial x} + \frac{\partial U_H}{\partial v_{HL}} \frac{dv_{HL}}{dx} = \frac{\partial U_H}{\partial x}.$$

Thus, the first order condition for any $x \in \{v_L^*, p_{2L}, p_{4L}\}$ is:

$$\frac{d\Pi}{dx} = \frac{\partial \Pi}{\partial S_L}\frac{dS_L}{dx} + \frac{\partial \Pi}{\partial U_H}\frac{\partial U_H}{\partial x} = 0.$$

The derivatives in the first term are: $\partial \Pi / \partial S_L = (1-\beta) G_L(U_L)$,

$$\frac{dS_L}{dv_L} = f_L(v_L)\left(\int_{p_{2L}}^{p_{4L}} (v-c) f_L(v)\, dv - (v_L - c)\right),$$

$dS_L/dp_{2L} = -F_L(v_L^*)(p_{2L}-c) f_L(p_{2L})$, and $dS_L/dp_{4L} = -(1-F_L(v_L^*))(p_{4L}-c) f_L(p_{4L})$. Components of the derivative of the second term in profits are: $dp_1/dv_L = 1$, $dp_1/dp_2 = (1-F_L(p_2))$, $dp_1/dp_4 = -(1-F_L(p_4))$, $\partial Z/\partial p_1 = (F_H(v_{HL}) - F_L(v_L))$,

$$\frac{dZ}{dp_2} = \frac{\partial Z}{\partial p_2} + \frac{\partial Z}{\partial p_1}\frac{dp_1}{dp_2} = F_H(v_{HL})(F_H(p_2) - F_L(p_2)),$$

$$\frac{dZ}{dp_4} = \frac{\partial Z}{\partial p_4} + \frac{\partial Z}{\partial p_1}\frac{dp_1}{dp_4} = (1-F_H(v_{HL}))(F_H(p_4) - F_L(p_4)),$$

and finally since by the envelope condition $\partial Z/\partial v_L = 0$,

$$\frac{dZ}{dv_L} = \frac{\partial Z}{\partial v_L} + \frac{\partial Z}{\partial p_1}\frac{dp_1}{dv_L} = \frac{\partial Z}{\partial p_1} = (F_H(v_{HL}) - F_L(v_L)).$$

Putting all these pieces together gives the first order conditions

$$\frac{d\Pi}{dv_L} = (1-\beta) G_L(U_L) f_L(v_L)\left(\int_{p_2}^{p_4} (v-c) f_L(v)\, dv - (v_L - c)\right) + \frac{\partial \Pi}{\partial U_H}(F_H(v_{HL}) - F_L(v_L)) = 0 \tag{38}$$

$$\frac{d\Pi}{dp_2} = -(1-\beta) G_L(U_L) F_L(v_L)(p_2 - c) f_L(p_2) + \frac{\partial \Pi}{\partial U_H} F_H(v_{HL})(F_H(p_2) - F_L(p_2)) = 0 \tag{39}$$

$$\frac{d\Pi}{dp_4} = -(1-\beta) G_L(U_L)(1-F_L(v_L))(p_4 - c) f_L(p_4) + \frac{\partial \Pi}{\partial U_H}(1-F_H(v_{HL}))(F_H(p_4) - F_L(p_4)) = 0 \tag{40}$$

which can be rearranged to derive equations (9)-(11).

(4) By inspection of equations (10)-(11), FOSD implies $p_{2L} > c$ and $p_{4L} > 0$. (The fact that IC-H is binding implies that $\partial \Pi / \partial U_H < 0$.) It follows from equation (6) that

$$v_{HL} - v_L = \int_{p_2}^{p_4} (F_L(v) - F_H(v))\, dv \tag{41}$$

Now we know that $(F_L(v) - F_H(v)) \geq 0$ by FOSD, so the difference $(v_{HL} - v_L)$ has the same sign as the difference $(p_4 - p_2)$.

I claim that $p_4 > p_2$. Proof: Suppose not and $p_2 > p_4$. Then $v_{HL} < v_L$ by equation (41) and $F_H(v_{HL}) < F_H(v_L) \leq F_L(v_L)$ by FOSD. This implies $\frac{1-F_H(v_{HL})}{1-F_L(v_L)} > 1 > \frac{F_H(v_{HL})}{F_L(v_L)}$ which in turn implies that $\int_a^b \frac{d\Pi}{dp_4} > 0$ follows from $\int_a^b \frac{d\Pi}{dp_2} \geq 0$ for any $b > a$ given equations (39)-(40). Now the fact that $p_2 > p_4$ is optimal implies that $\int_{p4}^{p2} \frac{d\Pi}{dp_2} \geq 0$ so it must also be true that $\int_{p4}^{p2} \frac{d\Pi}{dp_4} > 0$, contradicting optimality of p_4.

Similarly I claim that $F_L(v_L) - F_H(v_{HL}) > 0$. Proof: Suppose not and $\frac{1-F_H(v_{HL})}{1-F_L(v_L)} < 1 < \frac{F_H(v_{HL})}{F_L(v_L)}$. Then by a similar comparison of derivatives as above it follows that $p_2 > p_4$ which contradicts the prior result. Given $F_L(v_L) - F_H(v_{HL}) > 0$ and $p_4 > p_2$ equation (38) implies that $v_L > c$. Thus all distortions are downwards.

(5) The final step is to show that the relaxed IC-L constraint is satisfied. This follows from the fact that quantities are monotonic in the ex ante signal: $q_t\left(H, v^t\right) \geq q_t\left(L, v^t\right)$. To show that IC-L is satisfied, it is sufficient to show that

$$U_H - U_{HL} \geq U_{LH} - U_L(v_{HL}), \tag{42}$$

where by $U_L(v_{HL})$ I mean the expected utility of type L who chooses contract L but uses the optimal first-period threshold of a deviating high-type. By binding IC-H this implies $U_L(v_{HL}) \geq U_{LH}$ and since $U_L \geq U_L(v_{HL})$ this is guarantees IC-L.

To show equation (42) I compute

$$U_H - U_{LH} + U_L(v_{HL}) - U_{HL}$$

from equation (7) adjusting for the different threshold in the case of $U_L(v_{HL})$. After several lines of algebra and integration by parts, this can be shown to equal

$$\int_{p_{2L}}^{p_{4L}} \left(\begin{array}{c} (1 - F_H(v_{HL}^*))(1 - F_H(v)) \\ -(1 - F_L(v_{HL}^*))(1 - F_L(v)) \end{array} \right) dv + \int_c^{p_{2L}} (F_L(v) - F_H(v))\, dv,$$

which is positive by FOSD and $p_{4L} > p_{2L}$.

II. The result for $\mu_H^* < \mu_L^*$ follows by a symmetric argument, where I start by relaxing IC-H and showing that IC-L must bind with equality. **III.** Finally the result for $\mu_H^* = \mu_L^*$ follows because the optimal solution when both IC constraints are relaxed is a single contract which obviously satisfies incentive compatibility.

A.4 Proof of Proposition 4

(1) $\tau_H = \tau_L = \tau$: Firm A's residual demand from consumers of type s is $G_s\left(U_s^A\right) = \frac{1}{2\tau_s}\left(U_s^A - U_s^B + \tau_s\right)$. In the proposed symmetric equilibrium, this implies $\mu_s^* = \tau_s$. Proposition 3 implies firm offers are best responses to each other. There are no other symmetric pure strategy equilibria, since with any set of symmetric offers $\mu_s^* = \tau_s$.

(2) If $\tau_H \neq \tau_L$, then all equilibria are inefficient: Suppose not, and in equilibrium we have efficient allocations. Then $p_{3s}^i = 0$ and $p_{1s}^i = p_{2s}^i = c$. This means that we need to have $p_{0L}^i = p_{0H}^i$. As a result $\mu_H^i = \mu_L^i = \mu^i$. These statements hold for any offer in B's mixed strategy. A's expected market share in segment s is therefore $\frac{1}{2\tau_s}\left(E\left[\mu^B\right] - \mu_s^A + \tau_s\right)$, and A's best response markup is $\mu_s^{*A} = \frac{1}{2}\left(E\left[\mu^B\right] + \tau_s\right)$. Thus $\mu_L^{*A} \neq \mu_H^{*A}$ and by Proposition 3 A's best response includes an inefficient contract.

(3a) If $\tau_H > \tau_L$, then in all symmetric equilibria, high types receive first-best allocations, while low types' allocation is distorted downwards below first best: We know that in a symmetric pure strategy equilibrium that for both firms, either $\mu_L^* = \mu_H^*$, $\mu_L^* < \mu_H^*$ or $\mu_L^* > \mu_H^*$. Part (2) rules out $\mu_L^* = \mu_H^*$ if $\tau_H > \tau_L$. All that remains is to rule out $\mu_L^* > \mu_H^*$. This is ruled out by the assumption that the pass-through-rate (PTR) of demand is less than 1,[27] which implies markups are strategic complements (Weyl and Fabinger 2009, Bulow, Geanakoplos and Klemperer 1985): Let μ_s^{**} be the optimal markup unconstrained by ex ante IC at current allocation (i.e. that solves $\frac{\partial \Pi}{\partial U_s} = \beta_s g_s\left(U_s\right)\left(S_s - U_s - \frac{G_s(U_s)}{g_s(U_s)}\right) = 0$ at the current allocation). The fact that $\frac{G_s(U_s)}{g_s(U_s)}$ is increasing implies that $\mu_s^{**}\left(S_s\right)$ is increasing in S_s and $\mu_s^{**} \leq \mu_s^*$. Also, if $S_s - U_s < \frac{G_s(U_s)}{g_s(U_s)}$ then $\mu_s^{**} < \frac{G_s(U_s)}{g_s(U_s)}$. In any symmetric equilibrium, $\frac{G_s(U_s)}{g_s(U_s)} = \tau_s$, so if $\frac{\partial \Pi}{\partial U_s} < 0$ then $\mu_s < \mu_s^{**} < \tau_s$ and vice versa. Supposing $\mu_L^* > \mu_H^*$, then by Proposition 3, low contracts are first best. Hence $\mu_L^{**} = \mu_L^*$ while $\mu_H^{**} \leq \mu_H^*$. Also, $\frac{\partial \Pi}{\partial U_L} < 0$ and $\frac{\partial \Pi}{\partial U_H} > 0$, so $\mu_L^{**} < \tau_L$ and $\mu_H^{**} > \tau_H$. Putting these together with $\tau_L < \tau_H$ gives

$$\mu_L^* = \mu_L^{**} < \tau_L < \tau_H < \mu_H^{**} \leq \mu_H^*$$

which contradicts $\mu_L^* > \mu_H^*$.

(3b) $\tau_H < \tau_L$ follows a symmetric argument.

[27] In Section 4, for the heterogeneous outside-options case of the model, I assume that $U_s + \frac{G_s(U_s)}{g_s(U_s)}$ is strictly increasing. This corresponds to a decreasing marginal revenue assumption. The stronger assumption, $\frac{G_s(U_s)}{g_s(U_s)}$ increasing, is equivalent to $G\left(U\right)$ log concave and pass-through-rate less than 1. The later is implied by the assumption that consumers are uniformly distributed so that $G_s\left(U_s^A\right)/g_s\left(U_s^A\right) = \left(U_s^A - U_s^B + \tau_s\right)$.

A.5 Proof of Proposition 6

Proposition 6 is stated for either of two restrictions: (1) $p_{3s} \leq p^{\max}$ or (2) $p_{3s} \leq v_s/(1-F_s(v_s))$. Both can be written as $p_{3s} \leq h(v_s)$ for some $h(v_s)$ that is strictly positive and non-decreasing. All but the last step of the proof work with the restrictions in this general form.

I. First consider half the parameter space: $\mu_H^* \geq \mu_L^*$.

By equation (18), IC-H is given by equation (43):

$$U_H \geq U_{HL} = U_L + 2\int_{v_{HL}}^{\bar{v}} (v - v_L)\, dF_H(v) - 2\int_{v_L}^{\bar{v}} (v - v_L)\, dF_L(v) - p_{3L}\left(F_L(v_L) - F_H(v_{HL})\right)^2 \tag{43}$$

Relax IC-L. There are two cases: either (1) IC-H is slack or (2) IC-H binds.

Case (1), IC-H is slack.

(a) Show that IC-L is satisfied. Since IC-L is relaxed by increasing p_{3H}, it is sufficient to check at $p_{3H} = 0$. If both IC-L and IC-H are slack, then $v_L = v_H = c$ and at $p_{3H} = 0$, $P_H(q_1, q_2) = T + c(q_1 + q_2)$, so that $U_H = S_H^{FB} - T$ and $U_{LH} = S_L^{FB} - T$. Thus IC-L, $U_L \geq U_{LH}$, is equivalent to $S_H^{FB} - U_H \geq S_L^{FB} - U_L$, or $\mu_H^* \geq \mu_L^*$ at optimal offer $\{U_H, U_L\}$ which is satisfied by assumption.

(b) Substituting $v_L = c$ into equation (43), gives

$$U_H \geq U_{HL} = \left(U_L - S_L^{FB}\right) + 2\int_{v_{HL}}^{\bar{v}} (v - c)\, dF_H(v) - p_{3L}\left(F_L(c) - F_H(v_{HL})\right)^2.$$

Noting that p_{3L} can be set to the maximum $h_L(c)$, $2\int_{v_{HL}}^{\bar{v}} (v-c)\, dF_H(v) = S_H^{FB} - 2\int_c^{v_{HL}} (v-c)\, dF_H(v)$, and by definition at optimal utility offers, $\left(S_H^{FB} - \hat{U}_H\right) - \left(S_L^{FB} - \hat{U}_L\right) = \mu_H^* - \mu_L^*$, IC-H simplifies to:

$$\left(\mu_H^* - \mu_L^*\right) \leq 2\int_c^{v_{HL}} (v - c)\, dF_H(v) + h_L(c)\left(F_L(c) - F_H(v_{HL})\right)^2.$$

Further, $v_{HL} = c + h_L(c)\left(F_L(c) - F_H(v_{HL})\right)$ is uniquely defined by the FOC from equation (13) for v_{HL}, where $\bar{p}_L$ is given by equation (17) and $p_{3L} = h_L(c)$. Note, if instead $p_{3L} = 0$, then IC-H reduces to $\left(\mu_H^* - \mu_L^*\right) = 0$. So we need $p_{3L} > 0$ for any $\mu_H^* > \mu_L^*$ even if IC-H doesn't bind at optimal prices - because it would bind at $p_{3L} = 0$.

Case (2), IC-H binds. If equation (20) is not satisfied, then IC-H cannot be relaxed. Moreover, it will bind with equality given either ZOOM (where $\partial\Pi/\partial U_H = -\beta$ for all $U_H > 0$) or HOO (where decreasing marginal revenue assumption, $U_s + \frac{G_s(U_s)}{g_s(U_s)}$ increasing, implies concavity).

(a) Show $v_H = c$, derive FOC for v_L, and show $v_L \geq v_H$: (i) $v_H = c$: If IC-L is relaxed, then it is optimal to choose $v_H = c$. Moving from $\hat{v}_H$ a little bit towards c holding U_H fixed increases profits from high types, leaves IC-H unaffected, and we are ignoring IC-L. (Note that at $p_{3H} = 0$, we will also have $v_{LH} = c$.) (ii) FOC for v_L: The profit maximization problem is

$$\max_{U_L, v_L, p_{3L}} \left((1-\beta) G_L(U_L)(S_L - U_L) + \beta G_H(U_H)\left(S_H^{FB} - U_H\right) \right)$$

$$\begin{aligned}
\text{s.t. } U_H &= U_L + 2\int_{v_{HL}}^{\bar{v}} (v - v_L)\, dF_H(v) - 2\int_{v_L}^{\bar{v}} (v - v_L)\, dF_L(v) - p_{3L}\left(F_L(v_L) - F_H(v_{HL})\right)^2 \\
S_L &= 2\int_{v_L}^{\bar{v}} (v - c)\, dF_L(v) \\
p_{3L} &= h(v_L)
\end{aligned}$$

By the envelope condition, $\frac{dU_H}{dv_L} = \frac{\partial U_H}{\partial v_L}$, so the FOC for v_L is:

$$\frac{d\Pi}{dv_L} = \frac{\partial \Pi}{\partial S_L}\frac{dS_L}{dv_L} + \frac{\partial \Pi}{\partial U_H}\left(\frac{\partial U_H}{\partial v_L} + \frac{\partial U_H}{\partial p_{3L}} h'(v_L)\right)$$

Taking derivatives and substituting equation (19) for $\frac{\partial U_H}{\partial p_{3L}}$ gives:

$$\begin{aligned}
\frac{d\Pi}{dv_L} &= -2(1-\beta) G_L(U_L)(v_L - c) f_L(v_L) \\
&\quad - \frac{\partial \Pi}{\partial U_H}\left[2\left(F_L(v_L) - F_H(v_{HL})\right)(1 + p_{3L} f_L(v_L)) + \left(F_L(v_L) - F_H(v_{HL})\right)^2 h'(v_L) \right].
\end{aligned}$$

The FOC $\frac{d\Pi}{dv_L} = 0$ simplifies to equation (21), or for non-negative marginal prices, $h_s(v_s) = v_s/(1 - F_s(v_s))$, to:

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

Similar to case (1), $v_{HL} = v_L + h_L(v_L)(F_L(v_L) - F_H(v_{HL}))$, follows from equations (17) and (13). Since IC-H is binding, $\frac{\partial \Pi}{\partial U_H} \leq 0$. (ZOOM: $\frac{\partial \Pi}{\partial U_H} = -\beta$, HOO: $\frac{\partial \Pi}{\partial U_H} = \beta g_H(U_H)\left(S_H^{FB} - U_H - \frac{G_H(U_H)}{g_H(U_H)}\right)$.) Moreover, $F_L(v_L) - F_H(v_{HL}) \geq 0$, so $v_L \geq c = v_H$. (Why is $F_L(v_L) - F_H(v_{HL}) \geq 0$? Suppose not. Then by $v_{HL} = v_L + h(v_L)(F_L(v_L) - F_H(v_{HL}))$ and $h(v_L) > 0$ we would have $v_{HL} < v_L$ and hence $F_L(v_L) - F_H(v_{HL}) > F_L(v_L) - F_H(v_L) \geq 0$ by FOSD. Contradiction.)

(b) Show that IC-L is satisfied: Suppose that $\{U_L, v_L, p_{3L}, v_H, p_{3H}\}$ is the relaxed solution, with IC-H binding so that equation (43) holds with equality. Now consider the alternative contract menu $\left\{U_L, v_L, \hat{p}_{3L} = 0, \hat{U}_H, v_H, \hat{p}_{3H} = 0\right\}$ with $\hat{U}_H = \left(2\int_{v_{HL}}^{\bar{v}} (v - v_L)\, dF_H(v) - 2\int_{v_L}^{\bar{v}} (v - v_L)\, dF_L(v) + U_L\right)$ (preserving IC-H with equality). In this case IC-H equality, FOSD, and $v_L \geq v_H$ imply IC-L. This is the standard logic - high types are willing to pay more ex ante for a decrease in marginal price than are low types. If high-types are just indifferent to the upgrades, then low-types won't find it worthwhile. Next move back to the original contract menu, in two steps. First adjust the p_{3s}

keeping $\hat{U}_H$ fixed. We know that this relaxes the IC constraints, so it is still IC. Second decrease U_H back to IC-H binding. The decrease in U_H relaxes IC-L still further. So it is still satisfied.

II. Now consider the other half of the parameter space, $\mu_H^* \leq \mu_L^*$. The results follow by a nearly symmetrical argument. The only important difference is that for $\mu_H^* - \mu_L^* < -X_L$, the first order condition for v_H,

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

may call for $v_H > c$ if $h_H'(v_H)$ is sufficiently positive, which would violate the relaxed IC-H condition. However, the proposition is stated for $h_H(v_H) = p^{\max}$ or for $h_H(v_H) = \frac{v_H}{1-F_H(v_H)}$ rather than for general $h_H(v_H)$. In the former case there is no issue, since $h_H'(v_H) = 0$. In the latter case, there is an additional step to show that $v_H < c$. Given $p_{3H} = h_H(v_H) = \frac{v_H}{1-F_H(v_H)}$, the first order condition for v_H can be re-written as

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

In this form, it is apparent by inspection that $v_H < c$, despite $h' > 0$.

A.6 Proof of Proposition 7

(1) Show proposed equilibrium exists by construction: Impose $p_{3s} \leq h_s(v_s) = v_s/(1-F_s(v_s))$. Assume that each firm offers $p_{3s} = h_s(c)$, $v_L = v_H = c$, and $U_s = S_s^{FB} - \tau_s$. In this case, $U_s = \hat{U}_s$ and $\mu_s = \mu_s^* = \tau_s$. As a result, $(\mu_H^* - \mu_L^*) = \tau(H - L)$. For τ sufficiently small, this satisfies the condition for first-best allocations in Proposition 6, which verifies that the proposed offers are best responses. If the constraint $p_{3s} \leq h_s(v_s)$ were relaxed (no such constraint was imposed in the corollary) this would still be an equilibrium.

(2) Show that no other symmetric pure strategy equilibrium exist: There are three possibilities: (a) $(\mu_H^* - \mu_L^*) < -X_L$, (b) $(\mu_H^* - \mu_L^*) > X_H$, and (c) $(\mu_H^* - \mu_L^*) \in [-X_L, X_H]$. Given (c), the proposed equilibrium is unique. A symmetric equilibrium in case (a) is ruled out by $\tau_H > \tau_L$ and pass-through-rate less than 1 following a similar argument that was used in the proof of Proposition 4.[28] I rule out a symmetric equilibrium in case (b) by showing that there would exist a profitable deviation:

[28] There I showed that PTR less than 1 implies (i) that $\mu_s^{**} \leq \mu_s^*$ and (ii) that in any symmetric equilibrium if $\frac{\partial \Pi}{\partial U_s} < 0$ then $\mu_s < \mu_s^{**} < \tau_s$ and vice versa. If $(\mu_H^* - \mu_L^*) < -X_L$, then by Proposition 6, low contracts are first best. Hence $\mu_L^{**} = \mu_L^*$ while $\mu_H^{**} \leq \mu_H^*$. Also, $\frac{\partial \Pi}{\partial U_L} < 0$ and $\frac{\partial \Pi}{\partial U_H} > 0$, so $\mu_L^{**} < \tau_L$ and $\mu_H^{**} > \tau_H$. Putting these together with $\tau_L < \tau_H$ gives $\mu_L^* = \mu_L^{**} < \tau_L < \tau_H < \mu_H^{**} \leq \mu_H^*$ which contradicts $\mu_L^* > \mu_H^*$. See also footnote 27.

Suppose a symmetric equilibrium satisfied $(\mu_H^* - \mu_L^*) > X_H$. Then by Proposition 6, IC-H binds and IC-L is slack. By symmetry, at the equilibrium utility offers: $\frac{G_H}{g_H} - \frac{G_L}{g_L} = (\tau_H - \tau_L)$. I construct a profitable menu deviation in three steps, ignoring IC-H until the end. (i) Change U_s to the unconstrained optimum at current S_s, which increases profits. This means lowering U_H and raising U_L, and relaxing IC-L. Given the pass-through-rate less than 1 assumption (see footnote 27), this means lowering $\frac{G_H}{g_H} - \frac{G_L}{g_L}$. (ii) Change S_L to S_L^{FB} for L type, which increases profits and does not affect IC-L. (We already have $S_H = S_H^{FB}$ by $(\mu_H^* - \mu_L^*) > X_H$.) (iii) Change U_L to $\hat{U}_L$, which increases profits now that $S_L = S_L^{FB}$. (We already have $U_H = \hat{U}_H$ from step (i).) The change in U_L follows an increase in S_L, so is an increase in U_L by decreasing marginal revenue, and hence relaxes IC-L. By pass-through-rate less than 1, this lowers $\frac{G_H}{g_H} - \frac{G_L}{g_L}$. The new contract has strictly higher profits and still satisfies IC-L. The new contract menu offers unconstrained optimal markups and first-best allocations, so $\hat{\mu}_s = \frac{G_s(\hat{U}_s)}{g_s(\hat{U}_s)} = \mu_s^*$. As a result

$$\mu_H^* - \mu_L^* = \frac{G_H(\hat{U}_H)}{g_H(\hat{U}_H)} - \frac{G_L(\hat{U}_L)}{g_L(\hat{U}_L)} \leq \frac{G_H\,(U_H)}{g_H\,(U_H)} - \frac{G_L\,(U_L)}{g_L\,(U_L)} = \tau\,(H - L)$$

(where U_s means original utility offer, and $\hat{U}_s$ is the unconstrained optimal utility offer used in the new menu) and by Proposition 6, IC-H is satisfied for small τ. Thus this deviation was strictly profitable, and the proposed contract menus cannot have been an equilibrium.

(3) Total welfare result: With price-posting regulation, equilibrium pricing matches the attentive case, and Proposition 4 implies that allocations are inefficient in all equilibria for any $\tau > 0$. Thus PPR strictly reduces welfare.

(4) Distributional result: Without PPR, $\frac{\partial \Pi}{\partial U_L} = -\frac{\partial \Pi}{\partial U_H} = 0$. With PPR, Proposition 4 implies that in any symmetric pure strategy equilibrium IC-H binds and $\frac{\partial \Pi}{\partial U_L} = -\frac{\partial \Pi}{\partial U_H} > 0$. This implies high types win, $U_H^{PPR} > S_H^{FB} - \tau_H = \hat{U}_H$, but low types $U_L^{PPR} < S_L^{PPR} - \tau_L < S_L^{FB} - \tau_L = \hat{U}_L$ are losers. Firms still split both segments equally, but now make less on high types $S_H^{FB} - U_H^{PPR} < \tau_H$, but more on low types $S_L^{PPR} - U_L^{PPR} > \tau_L$. On average firms lose money. The first order condition under PPR, $\frac{\partial \Pi}{\partial U_L} = -\frac{\partial \Pi}{\partial U_H} > 0$, and symmetry ($G_s/g_s = \tau_s$) imply that

$$\frac{1}{2}\left(S_L^{PPR} - U_L^{PPR} - \tau_L\right)(1-\beta) = -\frac{\tau_L}{\tau_H}\frac{1}{2}\left(S_H^{FB} - U_H^{PPR} - \tau_H\right)\beta < -\frac{1}{2}\left(S_H^{FB} - U_H^{PPR} - \tau_H\right)\beta.$$

The inequality shows that the profit gain on low types (LHS) is less than the profit loss on high types (RHS).

A.7 Proof of Proposition 8

Consumer perceived and true utilities are

$$U = v_0 - p_0 + \int_{v_1}^{1} (v - p_1)\, dF(v) + F(v_1) \int_{p_2}^{1} (1 - F(v))\, dv + (1 - F(v_1)) \int_{p_4}^{1} (1 - F(v))\, dv,$$

and

$$U^* = v_0 - p_0 + \int_{v_1}^{1} (v - p_1)\, dF^*(v) + F^*(v_1) \int_{p_2}^{1} (1 - F^*(v))\, dv + (1 - F^*(v_1)) \int_{p_4}^{1} (1 - F^*(v))\, dv,$$

(the analogs of equation (36)) where $v_1 = p_1 + \int_{p_2}^{p_4} (1 - F^*(v))\, dv$ (the analog of equation (3)). Surplus is

$$S = v_0 + \int_{v_1}^{1} (v - c)\, dF(v) + F(v_1) \int_{p_2}^{1} (v - c)\, dF(v) + (1 - F(v_1)) \int_{p_4}^{1} (v - c)\, dF(v).$$

The firm's problem can be written as

$$\max_{U^*, v_1, p_2, p_4} G(U^*) (S - U^* + (U^* - U)),$$

where $p_1 = v_1 - \int_{p_2}^{p_4} (1 - F^*(v))\, dv$ and the difference between perceived and true expected utility is:

$$\begin{aligned} U^* - U \quad = \quad & \int_{v_1}^{1} (v - p_1) (f^*(v) - f(v))\, dv \\ & + F^*(v_1) \int_{p_2}^{1} (1 - F^*(v))\, dv + (1 - F^*(v_1)) \int_{p_4}^{1} (1 - F^*(v))\, dv \\ & - F(v_1) \int_{p_2}^{1} (1 - F(v))\, dv - (1 - F(v_1)) \int_{p_4}^{1} (1 - F(v))\, dv \end{aligned}$$

Let $\Delta = U^* - U$. First order conditions for $x \in \{v_1, p_2, p_4\}$ have the form:

$$\frac{d\Pi}{dx} = G(U^*) \left(\frac{\partial S}{\partial x} + \frac{\partial \Delta}{\partial x} + \frac{\partial \Delta}{\partial p_1} \frac{dp_1}{dx} \right) = 0.$$

Components of these derivatives are $dp_1/dv_1 = 1$, $dp_1/dp_2 = (1 - F^*(p_2))$, $dp_1/dp_4 = -(1 - F^*(p_4))$, $\partial \Delta / \partial p_1 = (F^*(v_1) - F(v_1))$, $\partial \Delta / \partial p_2 = F(v_1)(1 - F(p_2)) - F^*(v_1)(1 - F^*(p_2))$, $\partial \Delta / \partial p_4 = (1 - F(v_1))(1 - F(p_4)) - (1 - F^*(v_1))(1 - F^*(p_4))$, $\partial \Delta / \partial v_1 = -(v_1 - p_1)(f^*(v_1) - f(v_1))$, $\partial S / \partial p_2 = -F(v_1)(p_2 - c) f(p_2)$, $\partial S / \partial p_4 = -(1 - F(v_1))(p_4 - c) f(p_4)$, $\partial S / \partial v_1 = f(v_1) \int_{p_2}^{p_4} (v - c)\, dF(v) - (v_1 - c) f(v_1)$. Combining these pieces and canceling terms yields the following first-order conditions:

$$\frac{d\Pi}{dp_2} = G\left(U^*\right) F\left(v_1\right)\left(-\left(p_2 - c\right) f\left(p_2\right) + F^*\left(p_2\right) - F\left(p_2\right)\right) = 0,$$

$$\frac{d\Pi}{dp_4} = G\left(U^*\right)\left(1 - F\left(v_1\right)\right)\left(-\left(p_4 - c\right) f\left(p_4\right) + F^*\left(p_4\right) - F\left(p_4\right)\right) = 0.$$

These two first-order conditions imply that $p_2 = p_4$ and $v_1 = p_1$, so the third condition simplifies to

$$\frac{d\Pi}{dv_1} = G\left(U^*\right)\left(-\left(v_1 - c\right) f\left(v_1\right) + \left(F^*\left(v_1\right) - F\left(v_1\right)\right)\right) = 0.$$

As a result

$$v_1 = p_1 = p_2 = p_4 = p = c + \frac{F^*\left(p\right) - F\left(p\right)}{f\left(p\right)},$$

and optimal profits are given by equation (24). FOSD, $F^*\left(p\right) \geq F\left(p\right)$, implies $p \geq c$. The assumption $F^*\left(c\right) > F\left(c\right)$ implies $p > c$.

A.8 Proof of Proposition 9

The firm's problem: Perceived and true expected utilities are:

$$U^* = v_0 - p_0 + 2\int_{v^*}^{1} v dF^*\left(v\right) - \left(p_1 + p_2\right)\left(1 - F^*\left(v^*\right)\right) - p_3\left(1 - F^*\left(v^*\right)\right)^2, \tag{44}$$

and

$$U = v_0 - p_0 + 2\int_{v^*}^{1} v dF\left(v\right) - \left(p_1 + p_2\right)\left(1 - F\left(v^*\right)\right) - p_3\left(1 - F\left(v^*\right)\right)^2, \tag{45}$$

respectively. Expected surplus is:

$$S = \int_{v^*}^{1}\left(v - c\right) dF\left(v\right). \tag{46}$$

Substituting equations (26) and (27) into equation (45), yields true expected utility as a function of U^*, v^*, and p_3:

$$U = U^* + 2\int_{v^*}^{1}\left(\frac{F^*\left(v\right) - F\left(v\right)}{f\left(v\right)}\right) f\left(v\right) dv - p_3\left(F^*\left(v^*\right) - F\left(v^*\right)\right)^2. \tag{47}$$

The firm's profit function in equation (28) is then obtained by substituting equations (46) and (47) into the expression $\Pi = G\left(U^*\right)\left(S - U\right)$.

The proof: (1) First note that given FOSD, the sign of cross partial derivative $\partial^2\Pi/\partial p_3 \partial v^*$

equals the sign of $(f^*(v^*) - f(v^*))$:

$$\frac{\partial^2 \Pi}{\partial p_3 \partial v^*} = 2(F^*(v^*) - F(v^*))(f^*(v^*) - f(v^*)).$$

Given $F < F^*$ for all $v \in (0,1)$, there is an interval $[0,x)$ for which $f^* > f$. (In many natural cases f^* will cross f once from above at x). Profits are strictly super modular in p_3 and v^* ($\partial^2\Pi/\partial p_3 \partial v^* > 0$) over the interval $(0,x)$. Moreover, x is independent of γ. Let v^A be the solution $v^A = c + \frac{F^*(v^A) - F(v^A)}{f(v^A)}$ with attentive consumers. The limit of v^A as γ approaches zero is c. Therefore, if $c < x$ then for sufficiently small γ, $v^A < x$. Given the constraint $p_3 \leq p^{\max}$, strict super modularity on $(0,x)$ and $v^A < x$ imply $v^* > v^A$. This follows (from Edlin and Shannon (1998)) because the change in the firm's maximization problem from attentive to inattentive customers is identical to the change when customers are inattentive but p_3 exogenously increases from zero to $p^{\max}$. Note: given the constraint $p_3 \leq h(v^*)$ for $h(v^*)$ non-decreasing, the result continues to hold. If $h(v^*)$ is strictly increasing, the constraint simply creates an additional incentive to raise v^* when consumers are inattentive: to relax the constraint on p_3.

(2) For $c = 1$, $v^A = 1$ and allocations are first best with attentive customers. However, fix any $v^* \in (0,1)$ and for $p^{\max}$ sufficiently large,

$$2\int_{v^*}^{1}\left(v - 1 - \frac{F^*(v) - F(v)}{f(v)}\right) f(v)\, dv + p_3 (F^*(v^*) - F(v^*))^2 > 0,$$

which implies with inattentive customers there is overconsumption ($v^* < 1$) and total welfare is strictly lower. By continuity, this is true for c in a neighborhood around $c = 1$.

(3) (a) Let $\lambda(x) = \left(\hat{F}(x) - F(x)\right)/f(x)$. For sufficiently small γ, $\gamma\lambda'(x) < 1$. In this case the firm's profit function in the attentive case is strictly quasi-concave and the first order condition which characterizes the attentive solution, $v^A = c + \gamma\lambda\left(v^A\right)$, has a unique solution. Also, by the implicit function theorem, v^A is a continuous increasing function of c: $dv^A/dc = \left(1 - \gamma\lambda'(x)\right) > 0$. Moreover it varies from $v^A(c = 0) = 0$ to $v^A(c = 1) = 1$. The inverse is $c\left(v^A\right) = v^A - \gamma\lambda\left(v^A\right)$. (b) Define $x = \arg\max_v \{F^*(v) - F(v)\}$ to be the set of values which maximize disagreement. Define the largest point in the set to be $x^* = \sup\{x\}$. Note that $x^* < 1$. Let $c^* = x^* - \gamma\lambda(x^*)$ be the marginal cost for which $v^A(c^*) = x^*$. Such a value c^* exists by the argument in (a).

The inattentive solution is the same as the attentive solution at c^*, since disagreement is already maximized at x^*. Thus inattention does not change the distortion at c^*. However, for marginal costs c in a neighborhood above c^*, where $v^A(c)$ is slightly above x^*, the inattentive solution will be between x^* and the attentive solution. Hence $c^* < c < x^* < v^*(c) < v^A(c)$ for c in a neighborhood above c^*. It is clear that there is a local maximum to the inattentive problem between x^* and

$v^A(c)$. Reducing v^* below v^A initially has a second order negative effect on the first term in the firm's profit function but a first-order positive effect on the disagreement term. As v^* approaches x^*, the sign of the effects is unchanged but the orders are reversed. There could be no global maximum below x^* since any $v^* < x^*$ is dominated by x^* for both terms in the profit function. Given the assumption that disagreement $(F^*(v) - F(v))$ has a finitely many peaks (expressed in the text as the densities crossing finitely many times), I can always take $c > c^*$ close enough to c^* such that disagreement is larger at $v^A(c)$ than any higher v. In this case $v^* = v^A(c)$ dominates any higher choice of v^* for both terms in the profit function ruling out global maxima above $v^A(c)$.

(4) Constant-marginal price regulation has an identical effect to price-posting regulation because optimal pricing to attentive consumers derived in Proposition 8 features constant-marginal prices. (Also inattentive consumers behave the same as attentive consumers when marginal price is constant.)

A.9 Proof of Lemma 1

Given Bernoulli taste shocks, an attentive consumer's strategy is described by the tuple $\{b_0, b_1, b_{10}, b_{11}, b_{00}, b_{01}\}$. The pair $\{b_0, b_1\}$ describe the probabilities of first period purchase conditional on realizing $v_1 = 0$ or $v_1 = 1$ respectively. Following a first period purchase, the pair $\{b_{10}, b_{11}\}$ describe the probabilities of second period purchase conditional on a realized value of $v_2 = 0$ or $v_2 = 1$ respectively. The pair $\{b_{00}, b_{01}\}$ describe the corresponding second period purchase probabilities conditional on no purchase in period 1. Let $p_4 = p_2 + p_3$. Incentive compatibility constraints are straight forward in the second period. For instance, $b_{01} = 0$ requires $p_4 \geq 1$, $b_{01} \in (0,1)$ requires $p_4 = 1$, and $b_{01} = 1$ requires $p_4 \leq 1$. In the first period, purchases are made only if $v_1 \geq v_1^*$ where

$$v_1^* = p_1 + (1-\beta)(\max\{0, -p_2\} - \max\{0, -p_4\}) + \beta(\max\{0, 1-p_2\} - \max\{0, 1-p_4\}). \quad (48)$$

The expression simplifies substantially if $p_2, p_4 \in [0,1]$ (as is shown to be optimal below) in which case

$$v_1^* = p_1 + \alpha'(p_4 - p_2). \quad (49)$$

Then surplus and an attentive consumer's true and perceived expected-utilities are given by equations (50)-(52) as a function of prices and the strategy:

$$\begin{aligned} S = {} & v_0 + (1-\alpha) b_0 (-c) + \alpha b_1 (1-c) \\ & + (1 - (1-\alpha) b_0 - \alpha b_1)((1-\alpha) b_{00}(-c) + \alpha b_{01}(1-c)) \\ & + ((1-\alpha) b_0 + \alpha b_1)((1-\alpha) b_{10}(-c) + \alpha b_{11}(1-c)) \end{aligned} \quad (50)$$

$$
\begin{aligned}
U &= v_0 - p_0 + (1-\alpha)\, b_0\, (-p_1) + \alpha b_1 (1-p_1) \\
&\quad + (1 - (1-\alpha)\, b_0 - \alpha b_1)\, ((1-\alpha)\, b_{00}\, (-p_2) + \alpha b_{01} (1-p_2)) \\
&\quad + ((1-\alpha)\, b_0 + \alpha b_1)\, ((1-\alpha)\, b_{10}\, (-p_4) + \alpha b_{11} (1-p_4))
\end{aligned} \tag{51}
$$

$$
\begin{aligned}
U^* &= v_0 - p_0 + \left(1-\alpha'\right) b_0\, (-p_1) + \alpha' b_1 (1-p_1) \\
&\quad + \left(1 - \left(1-\alpha'\right) b_0 - \alpha' b_1\right) \left(\left(1-\alpha'\right) b_{00}\, (-p_2) + \alpha' b_{01} (1-p_2)\right) \\
&\quad + \left(\left(1-\alpha'\right) b_0 + \alpha' b_1\right) \left(\left(1-\alpha'\right) b_{10}\, (-p_4) + \alpha' b_{11} (1-p_4)\right)
\end{aligned} \tag{52}
$$

Firm profits are $\Pi = G(U^*)(S - U^* + \Delta)$, where the perception gap $\Delta = U^* - U$ is the difference between perceived and true expected utility.

Consider the firm maximizing profits by choosing the prices, and in cases of consumer indifference, the allocation.

1. It is optimal for the firm to induce efficient allocations in the second period and charge prices $p_2, p_4 \in [0,1]$.

   (a) It is optimal for the firm to induce $b_{01} = b_{11} = 1$ and charge $p_2, p_4 \leq 1$. Proof: Suppose $b_{11} < 1$. Then incentive compatibility implies $p_4 \geq 1$. Consider first reducing p_4 to $p_4' = 1$ if it happens to be higher and second changing the allocation to $b_{11}' = 1$ while keeping all else fixed. If $p_4 > 1$ then $b_{10} = b_{11} = 0$ and the initial reduction in p_4 maintains incentive compatibility of the allocation and satisfies NFL constraints without changing payoffs. Increasing b_{11} to 1 changes neither U^* nor U since consumers are indifferent to purchasing at a price equal to value. However it does increase surplus and does so strictly if there is positive probability of first period purchase. Thus this is a profitable deviation. A similar argument applies to b_{01} and p_2.

   (b) It is optimal for the firm to induce $b_{10} = 0$ and charge $p_4 \geq 0$. Proof: Suppose $b_{10} > 0$. Then incentive compatibility requires $b_{11} = 1$ and $p_4 \leq 0$. Consider the following changes: First, if $p_4 < 0$ then increase p_4 to $p_4' = 0$ and reduce p_1 by the same amount. Second reduce b_{10} to 0.

   The joint price change keeps v_1^* constant, and hence maintains incentive compatibility of the first period allocation (see equation (48)). Moreover, if $p_4 < 0$ then $b_{10} = 1$ and the joint price change does not affect payoffs because the consumer pays p_4 if and only if she pays p_1 and the sum is constant. The NFL constraints involving p_1 and p_4 are still satisfied. First $p_0 + p_1 + p_4 \geq 0$ is still satisfied because $p_1 + p_4$ was held constant. Second, since $p_4' = 0$, this implies the other constraint $p_0 + p_1 \geq 0$ holds. Thus the joint price change maintains NFL constraints, incentive compatibility, and does not affect payoffs.

The reduction of b_{10} to 0 increases surplus and does so strictly if first period purchases have positive probability. Moreover, it does not affect perceived consumer payoff U^* or the perception gap Δ since $p_4' = 0$. (Notice b_{10} only enters U and U^* in the product $b_{10}p_4$). Thus profits increase by the same amount as surplus and this is a profitable deviation.

(c) It is optimal for the firm to induce $b_{00} = 0$ and charge $p_2 \geq 0$. Proof: Suppose $b_{00} > 0$. Then incentive compatibility requires $b_{01} = 1$ and $p_2 \leq 0$. Consider the following changes: First, if $p_2 < 0$ then increase p_2 to $p_2' = 0$ and increase p_1 and reduce p_0 by the same amount: $p_1' = p_1 - p_2$, $p_0' = p_0 + p_2$. Second reduce b_{00} to 0.

The joint price change keeps v_1^* constant, and hence maintains incentive compatibility of the first period allocation (see equation (48)). Moreover, if $p_2 < 0$ then $b_{00} = b_{01} = 1$ and the joint price change does not affect payoffs.[29] The NFL constraints involving p_0, p_1, and p_2 are all still satisfied. First $p_0 + p_2 \geq 0$ is satisfied because the sum $p_0 + p_2$ is held constant. Second, $p_0 \geq 0$ is implied by $p_0 + p_2 \geq 0$ since $p_2' = 0$. Third, $p_0 + p_1 \geq 0$ and $p_0 + p_1 + p_4 \geq 0$ are satisfied because the sum $p_0 + p_1$ is held constant. Thus the joint price change maintains incentive compatibility of the allocation, satisfies NFL constraints, and does not affect payoffs.

The reduction of b_{00} to 0 increases surplus and does so strictly if first period purchases have probability less than 1. Moreover, it does not affect U^* or the perception gap because $p_2' = 0$. Thus profits increase by the same amount as surplus and this is a profitable deviation.

2. It is optimal for the firm to charge a non-negative penalty fee: $p_4 \geq p_2$. Proof: Suppose not and $p_4 < p_2$. By (1) we can consider $0 \leq p_2 < p_4 \leq 1$. In this case the expression for v_1^* is given by equation (49). Consider raising p_4 by $(p_2 - p_4)$ to $p_4' = p_2$ and reducing p_1 by $\alpha' (p_2 - p_4)$ so that v_1^* is held constant. The allocation remains incentive compatible and there is no change in surplus. Moreover, the price change leaves U^* constant as the relative sizes of the opposing price changes (the change in p_1 is smaller by factor α') are offset by the relative probabilities they are perceived to be paid (since the second period allocation is

[29] If the consumer does not buy in the first period she will buy in the second period ($b_{00} = b_{01} = 1$) and pay an additional $|p_2|$ because $p_2' = 0$. On the other hand, if she does buy in the first period she will still pay an additional $|p_2|$ due to the increase in p_1. However both changes are equally offset by the reduction in the fixed fee.

efficient from part 1, the perceived probability p_4 is paid is smaller than that of p_1 by factor α'.) However the true utility delivered and hence the perception gap both change because the relative likelihood the two prices are paid depends on α rather than α'. Plugging in efficient second period allocations, the perception gap is initially:

$$\Delta = \left(\alpha - \alpha'\right) \left(\begin{array}{c} -b_1 + (b_1 - b_0)\, p_1 - (1 - p_2) \\ + \left(b_0 + (\alpha + \alpha')\,(b_1 - b_0)\right)(p_4 - p_2) \end{array} \right) \tag{53}$$

After adjusting prices to $p_1' = p_1 - \alpha'\,(p_2 - p_4)$ and $p_4' = p_2$, this becomes

$$\Delta' = \left(\alpha - \alpha'\right)\left(-b_1 + (b_1 - b_0)\left(p_1 - \alpha'\,(p_2 - p_4)\right) - (1 - p_2)\right).$$

The difference is

$$\Delta' - \Delta = \left(\alpha - \alpha'\right)(p_2 - p_4)\left(b_0 + \alpha\,(b_1 - b_0)\right)$$

which is non-negative since $p_2 > p_4$ by assumption and incentive compatibility requires $b_1 \geq b_0$. Thus this is a profitable deviation (strictly profitable if there are any purchases in the first period).

3. It is optimal for the firm to induce the efficient allocation in the first period.

    (a) It is optimal for the firm to induce $b_1 = 1$. Proof: Suppose not and $b_1 < 1$. Incentive compatibility requires $b_0 = 0$ and $v_1^* = p_1 + \alpha'\,(p_4 - p_2) \geq 1$. Suppose $v_1^* > 1$. Then $b_0 = b_1 = 0$ and I can reduce p_1 to $1 - \alpha'\,(p_4 - p_2) \geq 1 - \alpha' > 0$ so that $v_1^* = 1$ without disrupting incentive constraints or effecting payoffs or violating NFL constraints. (Constraint $p_0 + p_1 \geq 0$ is redundant to $p_0 \geq 0$ since p_1 is positive. Constraint $p_0 + p_1 + p_4 \geq 0$ is implied by $p_4 \geq p_2$ from part (2) and p_1 positive since $p_0 + p_2 + p_1 \geq p_0 + p_2 \geq 0$.) So I can safely consider $v_1^* = 1$. Now consider raising b_1 to 1. Since $v_1^* = 1$, the consumer is indifferent and U^* is unaffected. Surplus is strictly increased. The perception gap is initially described by equation (53). After the increase in b_1, this becomes

$$\Delta' = \left(\alpha - \alpha'\right) \left(\begin{array}{c} -1 + (1 - b_0)\, p_1 - (1 - p_2) \\ + \left(b_0 + (\alpha + \alpha')\,(1 - b_0)\right)(p_4 - p_2) \end{array} \right)$$

$$\begin{aligned} \Delta' &= \left(\alpha - \alpha'\right)\left((1 - b_0)\, p_1 - 1\right) - (1 - p_2) \\ &\quad + (p_4 - p_2)\left(\alpha - \alpha'\right)\left(b_0 + \left(\alpha + \alpha'\right)(1 - b_0)\right) \end{aligned}$$

and the difference is

$$\Delta' - \Delta = \left(\alpha - \alpha'\right)(1 - b_1)\left(\left(\alpha + \alpha'\right)(p_4 - p_2) - (1 - p_1)\right).$$

Substituting $p_1 = 1 - \alpha'(p_4 - p_2)$ into this expression yields

$$\Delta' - \Delta = \left(\alpha - \alpha'\right)(1 - b_1)(p_4 - p_2)\,\alpha$$

which is nonnegative since $p_4 \geq p_2$ by part (2). Thus there is a strict increase in profits at least as high as the increase in surplus. Hence $b_1 < 1$ could not have been optimal.

(b) It is optimal for the firm to induce $b_0 = 0$. Proof: Suppose not and $b_0 > 0$. Incentive compatibility requires $b_1 = 1$ and $v_1^* = p_1 + \alpha'(p_4 - p_2) \leq 0$. Suppose that $v_1^* < 0$. Then $b_0 = b_1 = 1$ and I can increase p_1 to $-\alpha'(p_4 - p_2)$ such that $v_1^* = 0$ without disruption incentive compatibility. If I increase p_0 by the same amount, then payoffs (U, U^*, and Π) remain constant. Moreover NFL constraints are still satisfied. The constraints involving p_2 and p_4 are redundant since $p_2, p_4 \geq 0$. The constraint $p_0 + p_1 \geq 0$ is unaffected because the sum remains constant. Moreover, it implies $p_0 \geq 0$ since $p_1 = -\alpha'(p_4 - p_2)$ is nonpositive by part (2). Thus the joint price change maintains incentive compatibility, NFL constraints, and does not affect payoffs. So I can safely consider $v_1^* = 1$ and $p_1 = -\alpha'(p_4 - p_2)$.

Note that $p_0 + p_1 \geq 0$ implies $p_0 \geq -p_1 = \alpha'(p_4 - p_2)$. Now consider increasing p_1 to zero, increasing p_2 to p_4, and reducing p_0 by $\alpha'^2(p_4 - p_2)$. NFL constraints are all satisfied. The preceding note shows $p_0 \geq 0$. This implies $p_0 + p_1 \geq 0$ since p_1 is now 0. The remaining constraints with p_2 and p_4 are slack since $p_2 = p_4$ is strictly positive. Increasing p_1 to 0 lowers U^* by $\alpha'^2(p_4 - p_2)$. Increasing p_2 to p_4 lowers U^* by $(1 - \alpha')\alpha'(p_4 - p_2)$. The total reduction is $\alpha'(p_4 - p_2)$. Lowering p_0 by $\alpha'(p_4 - p_2)$ exactly offsets this change so that U^* is in fact held constant. Note that the change in p_1 and p_2 ensure that v_1^* remains equal to zero and incentive compatibility is maintained. Surplus is unchanged but profits are effected via the perception gap. Substituting $b_1 = 1$ and $p_1 = -\alpha'(p_4 - p_2)$ into equation (53) yields an expression for the initial perception gap:

$$\Delta = -\left(\alpha - \alpha'\right)(2 - p_2 - (p_4 - p_2)(\alpha + b_0(1 - \alpha))).$$

After the price change, substituting $p_1' = 0$, $p_2' = p_4$, and $b_1 = 1$ into equation (53) yields an expression for the new perception gap:

$$\Delta' = \left(\alpha - \alpha'\right)(-1 - (1 - p_4)) \tag{54}$$

$$\Delta' = -\left(\alpha - \alpha'\right)(2 - p_4).$$

Thus the difference is

$$\Delta' - \Delta = \left(\alpha - \alpha'\right)(p_4 - p_2)(1 - (\alpha + (1 - \alpha) b_0)),$$

which is nonnegative. Thus this price change weakly increases profits.

Finally, lower b_0 to 0. This strictly increases surplus, and does not further effect the perception gap $\Delta' = -(\alpha - \alpha')(2 - p_4)$ because the penalty fee is zero. Thus profits strictly increase. Hence $b_0 > 0$ was not optimal.

A.10 Proof of Proposition 10

The proof follows from the description in the text except for four details.

(1) As stated in the text, reducing p_2 has a higher bang-for-the-buck than either p_1 or p_4. However, either p_1 or p_4 must be reduced at the same time to maintain incentive compatibility. Lowering p_2 maintains $v_1^* = 1$ if p_4 is reduced equally or p_1 is reduced a proportion α' as much (or a convex combination of such reductions in both p_1 and p_4). Reducing p_1 in tandem with p_2 has the highest bang-for-the-buck of these options because p_1 has higher bang-for-the-buck than p_4 and need be reduced only at rate $\alpha' < 1$. The bang-for-the-buck of simultaneously reducing p_2 and p_1 while maintaining $v_1^* = 1$ is

$$\gamma_{12} = -\frac{dU/dp_2 + \alpha' dU/dp_1}{dU^*/dp_2 + \alpha' dU^*/dp_1} = -\left(\alpha/\alpha'\right)\left(1 - \left(\alpha - \alpha'\right)\right).$$

Now the relevant bang-for-the-buck coefficients can be completely ranked:

$$\gamma_0 > \gamma_{12} > \gamma_1 > \gamma_4.$$

This implies that the procedure described in the text identifies the optimal price. This procedure leads to the four qualitative pricing regions in the proposition as a function of U^*.

(2) The boundary values of U^* between qualitative pricing regions are found by evaluating equation (30) at the boundary prices. For instance, between region (2) and (3) the boundary prices are $p_0 = p_2 = 0$, $p_4 = 1$, $p_3 = 1 - \alpha'$. At these prices, $U^* = v_0 + 2\alpha'(1 - \alpha')$. Within

a region, all prices but one are at a boundary. For instance in region (2) all other prices are a function of p_2. Plugging these prices into equation (30) and inverting for p_2 yields the expression $p_2 = 1-(U^* - v_0)/\alpha'$. In this way, the details of all five pricing regions presented in the proposition are derived.

(3) The derived prices in the proposition can be plugged into equation (29) to find U. Substituting these expressions along with $S^{FB} = v_0 + 2\alpha(1-c)$ provide the markup $\left(S^{FB} - U\right)$. The markups μ and corresponding derivatives $d\mu/dU^*$ for the four pricing regions are:

$$
\begin{array}{lll}
\text{Region} & \text{Markup} & d\mu/dU^* \\
1 & 2\alpha(1-c) - (U^* - v_0) & -1 \\
2 & 2\alpha(1-c) - (\alpha/\alpha')(1-\alpha+\alpha')(U^* - v_0) & -(\alpha/\alpha')(1-\alpha+\alpha') \\
3 & 2\alpha(1-c) + \alpha(\alpha-\alpha') - (\alpha/\alpha')(U^* - v_0) & -(\alpha/\alpha') \\
4 & 2\alpha(1-c) + 2(\alpha/\alpha')(\alpha-\alpha') - (\alpha/\alpha')^2(U^* - v_0) & -(\alpha/\alpha')^2
\end{array}
\tag{55}
$$

(4) The non-exploitation result $U \geq 0$ follows by brute force by calculating U in each region and comparing to zero. However it follows more directly by noting first that U is increasing in U^* and second that $U = U^*$ in region 1 when $U^* = 0$.

A.11 Proof of Proposition 11

For sufficiently small transportation cost τ, there will be full market coverage in equilibrium, with each firm receiving positive market share. In this case, if firms A and B offer perceived expected utilities of U^A and U^B respectively, market share of firm A is: $G\left(U^A, U^B\right) = \frac{1}{2\tau}\left(U^A - U^B + \tau\right) \geq 0$. Profits are

$$\Pi^A = G\left(U^A, U^B\right)\mu\left(U^A\right)$$

where $\mu\left(U^A\right)$ is the markup derived in Proposition 10 given by equation (55).The profit function is concave (with a kinks at the boundaries between pricing regions), and hence firm A's best response is a continuous function of U^B. Away from the kinks, $d^2\Pi^A/dU^{A2} = g\left(U^A, U^B\right) d\mu/dU^A < 0$, and at the kink $d\Pi^A/dU^A$ decreases. This follows since

$$\frac{d\Pi^A}{dU^A} = g\left(U^A, U^B\right)\mu\left(U^A\right) + G\left(U^A, U^B\right)\frac{d\mu}{dU^A},$$

and while $G\left(U^A, U^B\right)$ and $\mu\left(U^A\right)$ are continuous and nonnegative, $d\mu/dU^A$ decreases at kink points as shown in equation (**??**). The slope $d\mu/dU^A$ decreases precisely because firms order price cuts from highest bang-for-the buck to lowest.

The optimal U^A either solves the first order condition $\mu\left(U^A\right) = -\left(U^A - U^B + \tau\right) d\mu/dU^A$, or is located at a kink at the boundary between pricing regions. In the attentive case, there are seven sub-cases corresponding to the four pricing regions and three kinks. Substituting $\mu\left(U^A\right)$ and $d\mu/dU^A$ from equation (55), the first order conditions for the four pricing regions are:

Region	First Order Condition
1	$U^A = \frac{1}{2}\left(S^{FB} + U^B - \tau\right)$
2	$U^A = \frac{1}{2}\left(U^B - \tau + v_0\right) + \alpha'\left(1-c\right)/\left(1-\alpha+\alpha'\right)$
3	$U^A = \frac{1}{2}\left(U^B - \tau + v_0\right) + \frac{1}{2}\alpha'\left(\alpha-\alpha'\right) + \alpha'\left(1-c\right)$
4	$U^A = \frac{1}{2}\left(U^B - \tau + v_0\right) + \left(\alpha'/\alpha\right)\left(\alpha-\alpha'\right) + \left(\alpha'/\alpha\right)\alpha'\left(1-c\right)$

(56)

At the boundaries between pricing regions, offered utility is: Boundary 1/2, $U^A = v_0$; Boundary 2/3, $U^A = v_0 + 2\alpha'\left(1-\alpha'\right)$; Boundary 3/4, $U^A = v_0 + \alpha'\left(2-\alpha'\right)$.

By inspection, the best response by A has slope dU^A/dU^B of either zero (at a boundary point between pricing regions) or 1/2 (within a pricing region where a first order condition holds). Since $dU^A/dU^B \in [0,1)$ (and as already noted $U^A\left(U^B\right)$ is continuous), there is a unique pure strategy equilibrium, which is symmetric. This is true for both attentive and inattentive cases.

Each first order condition in equation (56) has a corresponding symmetric solution for the offered U^* and corresponding markup. Each is relevant for the range of transportation costs for which the solution U^* actually lies within the relevant pricing region. These are given in equation (61) below:

Region	Symmetric Solution, $U^* =$	Relevant Range
1	$v_0 - \tau + 2\alpha\left(1-c\right)$	$\tau > 2\alpha\left(1-c\right)$
2	$v_0 - \tau + 2\alpha'\left(1-c\right)/\left(1-\alpha+\alpha'\right)$	$\frac{2\alpha'(1-c)}{(1-\alpha+\alpha')} - \alpha' \leq \tau \leq \frac{2\alpha'(1-c)}{(1-\alpha+\alpha')}$
3	$v_0 - \tau + \alpha'\left(\alpha-\alpha'\right) + 2\alpha'\left(1-c\right)$	$\alpha'\left(\alpha-2c\right) \leq \tau \leq \alpha'\left(\alpha-2c\right) + \alpha'\left(1-\alpha'\right)$
4	$v_0 - \tau + \left(\alpha'/\alpha\right)\left(\alpha - c\alpha'\right)$	$0 \leq \tau \leq \alpha'^2 - 2\alpha'^2 c/\alpha$

(57)

Plugging the values of U^* derived in equation (61) into equation (55) yields markups as a function of transportation cost as described in the proposition. Markups are constant between pricing regions. Denote these boundary regions 1/2, 2/3, and 3/4 respectively. In these regions offered utilities and markups (derived from substituting offered utilities into equation (55)) are given by equation (58):

$$\begin{array}{llll} \text{Region} & U^* = & \mu = & \text{Relevant Range} \\ 1/2 & v_0 & 2\alpha\left(1-c\right) & \frac{2\alpha'(1-c)}{(1-\alpha+\alpha')} \leq \tau \leq 2\alpha\left(1-c\right) \\ 2/3 & v_0 + 2\alpha'\left(1-\alpha'\right) & 2\alpha\left(1-c\right) - \alpha\left(1-\alpha+\alpha'\right) & \alpha'\left(\alpha-2c\right) + \alpha'\left(1-\alpha'\right) \leq \tau \leq \frac{2\alpha'(1-c)}{(1-\alpha+\alpha')} - \alpha' \\ 3/4 & v_0 + \alpha'\left(2-\alpha'\right) & 2\alpha\left(1-c\right) - \alpha\left(2-\alpha\right) & \alpha'^2 - 2\alpha'^2 c/\alpha \leq \tau \leq \alpha'\left(\alpha-2c\right) \end{array} \tag{58}$$

A.12 Proof of Lemma 2

Solving equation (31) for p_0 yields:

$$p_0 = -U^* + v_0 + 2\left(1-\alpha'\right) b_0\left(-\bar{p}\right) + 2\alpha' b_1\left(1-\bar{p}\right) - \left(\left(1-\alpha'\right) b_0 + \alpha' b_1\right)^2 p_3.$$

Substituting this for p_0 into equation (32) gives:

$$\Pi = G\left(U^*\right)\left(\begin{array}{c} -U^* + v_0 + 2b_0\left(-\left(\alpha-\alpha'\right)\bar{p} - \left(1-\alpha\right)c\right) + 2b_1\left(\left(\alpha-\alpha'\right)\bar{p} + \alpha' - \alpha c\right) \\ +\left(\left(\left(1-\alpha\right)b_0 + \alpha b_1\right)^2 - \left(\left(1-\alpha'\right)b_0 + \alpha' b_1\right)^2\right) p_3 \end{array}\right).$$

There are four alternatives to the efficient allocation to consider:

1. $b_0 = b_1 = 1$: Profits and the fixed fee are:

$$\Pi_1 = G\left(U^*\right)\left(-U^* + v_0 + 2\left(\alpha' - c\right)\right),$$

$$p_0 = -U^* + v_0 + 2\alpha' - 2\bar{p} - p_3.$$

If $U^* \leq v_0 + 2\alpha'$, then this allocation can be implemented without violating the NFL constraint with prices $p_1 = p_2 = p_3 = 0$ and $p_0 = -U^* + v_0 + 2\alpha'$. If $U^* > v_0 + 2\alpha'$, then this allocation is not implementable without violating the NFL constraint. This follows from the fact that $p_0 + 2\bar{p} + p_3 \geq 0$ is equivalent to $U^* \leq v_0 + 2\alpha'$. However, the efficient allocation could be implemented with identical prices, also satisfying the NFL constraint for $U^* \leq v_0 + 2\alpha'$, but yielding strictly higher profit,

$$\Pi = G\left(U^*\right)\left(-U^* + v_0 + 2\left(\alpha' - \alpha c\right)\right),$$

by saving production cost $2\left(1-\alpha\right)c$. Thus $b_0 = b_1 = 1$ is never optimal.

2. $b_0 = b_1 = 0$: Profits and the fixed fee are:

$$\Pi_2 = G(U^*)(-U^* + v_0),$$

$$p_0 = -U^* + v_0.$$

If $U^* \leq v_0$ then this allocation is implementable without violating the NFL with prices $p_0 = -U^* + v_0$, $p_1 = p_2 = 1$, and $p_3 = 0$. If $U^* > v_0$, then this allocation is not implementable without violating the NFL constraint. However, the efficient allocation can be implemented with identical prices, strictly raising profits by $2\alpha(1-c)$ from the additional sales. Thus $b_0 = b_1 = 0$ is never optimal.

3. $b_0 \in (0, 1)$, $b_1 = 1$: For this allocation to be implemented, b_0 must satisfy first and second order conditions of the consumers' problem:

$$\frac{dU^*}{db_0} = -2(1-\alpha')\left(\bar{p} + ((1-\alpha')b_0 + \alpha')p_3\right) = 0,$$

and

$$\frac{d^2U^*}{db_0^2} = -2(1-\alpha')^2 p_3 \leq 0.$$

This requires that $p_3 \geq 0$ and $\bar{p} = -((1-\alpha')b_0 + \alpha')p_3$. At these prices, the three NFL constraints, (a) $p_0 \geq 0$, (b) $p_0 + \bar{p} \geq 0$, and (c) $p_0 + 2\bar{p} + p_3 \geq 0$ are:

$$\max\left\{\frac{U^* - v_0 - 2\alpha'}{((1-\alpha')b_0 + \alpha')^2}, \frac{U^* - v_0 - 2\alpha'}{(1-\alpha')^2(1-b_0)^2}\right\} \leq p_3 \leq \frac{2\alpha' + v_0 - U^*}{(1-\alpha')(1-b_0)((1-\alpha')b_0 + \alpha')}$$

If $p_3 \geq 0$, the upper bound on penalty fees can only be satisfied if $U^* \leq v_0 + 2\alpha'$, in which case the lower bound is always satisfied. Moreover, profits are increasing in penalty fee p_3,

$$\Pi_3 = G(U^*)\left(v_0 - U^* + 2(\alpha' - \alpha c) - 2b_0(1-\alpha)c + (\alpha - \alpha')^2(1-b_0)^2 p_3\right),$$

so the optimal penalty fee satisfies the upper bound with equality:

$$p_3 = \frac{(2\alpha' + v_0 - U^*)}{(1-\alpha')(1-b_0)((1-\alpha')b_0 + \alpha')}.$$

Given these prices, profits are strictly decreasing in b_0,

$$\frac{d\Pi_3}{db_0} = G(U^*)\left(-2(1-\alpha)c - \frac{(\alpha - \alpha')^2(1-b_0)}{((1-\alpha')b_0 + \alpha')}p_3\right) < 0,$$

for all $p_3 \geq 0$ and hence any NFL implementable allocation with $b_0 \in (0,1)$ is always dominated by the efficient allocation.

4. $b_0 = 0$, $b_1 \in (0,1)$: For this allocation to be implemented, b_1 must satisfy first and second order conditions of the consumers' problem:

$$\frac{dU^*}{db_1} = +2\alpha' (1-\bar{p}) - 2(\alpha')^2 b_1 p_3 = 0,$$

and

$$\frac{d^2U^*}{db_1} = -2(\alpha')^2 p_3 \leq 0.$$

This requires $p_3 \geq 0$ and $\bar{p} = 1 - \alpha' b_1 p_3$. At these prices, the three NFL constraints, (a) $p_0 \geq 0$, (b) $p_0 + \bar{p} \geq 0$, and (c) $p_0 + 2\bar{p} + p_3 \geq 0$ are:

$$\max\left\{\frac{U^* - v_0}{\alpha'^2 b_1^2}, \frac{U^* - v_0 - 2}{(1-\alpha' b_1)^2}\right\} \leq p_3 \leq \frac{1 + v_0 - U^*}{\alpha' b_1 (1 - \alpha' b_1)}$$

All three constraints can be satisfied only if $U^* \leq v_0 + \alpha' b_1$. (This is equivalent to $\frac{U^* - v_0}{\alpha'^2 b_1^2} \leq \frac{1+v_0-U^*}{\alpha' b_1(1-\alpha' b_1)}$, while $\frac{U^*-v_0-2}{(1-\alpha' b_1)^2} \leq \frac{1+v_0-U^*}{\alpha' b_1(1-\alpha' b_1)}$ is equivalent to the weaker condition $U^* \leq 1 + v_0 + \alpha' b_1$.) Otherwise, this allocation is not implementable without violating NFL. Profits are strictly increasing in p_3,

$$\Pi_4 = G(U^*)\left(-U^* + v_0 + 2b_1\alpha(1-c) + b_1^2(\alpha - \alpha')^2 p_3\right),$$

so the optimal penalty fee p_3 will equal the upper bound:

$$p_3 = \frac{1 + v_0 - U^*}{\alpha' b_1 (1 - \alpha' b_1)}.$$

Given these prices, profits are strictly increasing in b_1,

$$\frac{d\Pi_4}{db_1} = G(U^*)\left(2\alpha(1-c) + \frac{b_1(\alpha-\alpha')^2}{1-\alpha' b_1} p_3\right) > 0,$$

so any NFL implementable allocation with $b_1 \in (0,1)$ is dominated by the efficient allocation.

A.13 Proof Proposition 12

I begin by solving the firm's problem assuming that the firm does not disclose $\{t, q^{t-1}\}$ to consumers. The final step is to show that this is optimal.

NFL says prices can be no lower than $p_0 = p_1 = p_2 = p_3 = 0$, and hence offered perceived utility U^* can be no higher than $v_0+2\alpha'$. Optimal pricing need only be characterized for $U^* \in [0, v_0 + 2\alpha']$. By Lemma 2, the firm will induce the efficient allocation, $b_0 = 0$, $b_1 = 1$. As usual, profits are $\Pi = G(U^*)\,\mu(U^*)$. In this case, markups and fixed fees are:

$$\mu(U^*) = -U^* + v_0 + 2\left(\left(\alpha - \alpha'\right)\bar{p} + \alpha' - \alpha c\right) + \left(\alpha^2 - \alpha'^2\right) p_3,$$

$$p_0 = -U^* + v_0 + 2\alpha'(1 - \bar{p}) - \alpha'^2 p_3.$$

Incentive compatibility requires that the expected marginal price be between zero and one: $0 \leq \bar{p} + \alpha' p_3 \leq 1$, or alternatively that the penalty fee be between: $-\bar{p}/\alpha' \leq p_3 \leq (1 - \bar{p})/\alpha'$. The three NFL constraints, (a) $p_0 \geq 0$, (b) $p_0 + \bar{p} \geq 0$, and (c) $p_0 + 2\bar{p} + p_3 \geq 0$ are:

$$\frac{U^* - v_0 - 2\alpha'(1 - \bar{p}) - 2\bar{p}}{1 - \alpha'^2} \leq p_3 \leq \frac{2\alpha'(1 - \bar{p}) + v_0 - U^*}{\alpha'^2} + \min\left\{0, \frac{\bar{p}}{\alpha'^2}\right\}.$$

There are two cases to consider.

Case I, $U^* < \alpha' + v_0$: Impose the NFL upper bound $p_3 \leq \left(2\alpha'(1 - \bar{p}) + \bar{p} + v_0 - U^*\right)/\alpha'^2$ and the IC upper bound $p_3 \leq (1 - \bar{p})/\alpha'$, but relax the other three constraints. At $\bar{p} = -\frac{\alpha' - (U^* - v_0)}{1 - \alpha'}$, both constraints are the same and the optimal penalty fee would be the upper bound $p_3 = \frac{1 - (U^* - v_0)}{(1 - \alpha')\alpha'}$. For larger $\bar{p}$, the IC upper bound is tighter and the optimal penalty is $p_3 = (1 - \bar{p})/\alpha'$. In this case, profits are,

$$\Pi = G(U^*)\left(-U^* + v_0 + 2\left(\left(\alpha - \alpha'\right)\bar{p} + \alpha' - \alpha c\right) + \left(\alpha^2 - \left(\alpha'\right)^2\right)(1 - \bar{p})/\alpha'\right),$$

and $d\Pi/d\bar{p} = -G(U^*)(\alpha - \alpha')^2/\alpha' < 0$, so it is optimal to reduce $\bar{p}$ towards $\bar{p} = -\frac{\alpha' - (U^* - v_0)}{1 - \alpha'}$. For $\bar{p}$ below $-\frac{\alpha' - (U^* - v_0)}{1 - \alpha'}$, the NFL upper bound is binding, and as shown under case 1, it is optimal to increase $\bar{p}$. Thus the optimal prices are those given by equation (34) in the proposition. The assumption $U^* < v_0 + \alpha'$ ensures $\bar{p}$ is negative, and hence the alternative NFL upper bound is satisfied. Substituting for prices, the NFL lower bound reduces to $U^* \leq v_0 + \alpha' + 1$, which is satisfied given $U^* < v_0 + \alpha'$. The IC lower bound is always satisfied when the upper bound is satisfied with equality. Substituting for prices, the markup is

$$\mu(U^*) = 2\alpha(1 - c) + Y - (1 + Y)(U^* - v_0). \tag{59}$$

Case II, $U^* \in [v_0 + \alpha', v_0 + 2\alpha']$: Relax the incentive constraint and the NFL lower bound on the penalty fee. Since profits are increasing in both $\bar{p}$ and p_3, for any fixed $\bar{p}$, the penalty fee p_3

will be set at the NFL upper bound. If $\bar{p} \geq 0$, this implies $p_3 = (2\alpha' (1-\bar{p}) + v_0 - U^*) / \alpha'^2$,

$$\Pi = G(U^*) \left(2\alpha(\bar{p}-c) + 2\alpha^2 (1-\bar{p}) / \alpha' - \left(\alpha/\alpha'\right)^2 (U^* - v_0)\right),$$

and $d\Pi/d\bar{p} = -2\alpha(\alpha - \alpha')/\alpha' < 0$. Thus profits increases as $\bar{p}$ is reduced towards zero. If $\bar{p} \leq 0$, this implies $p_3 = (2\alpha'(1-\bar{p}) + \bar{p} + v_0 - U^*)/\alpha'^2$,

$$\Pi = G(U^*) \left(2\alpha(\bar{p}-c) - \bar{p} + \left(2\alpha'(1-\bar{p}) + \bar{p} - (U^* - v_0)\right) \left(\alpha/\alpha'\right)^2\right),$$

and $d\Pi/d\bar{p} = \left(\alpha^2 (1-2\alpha') - \alpha'^2 (1-2\alpha)\right)/\alpha'^2 > 0$. Thus profits increase as $\bar{p}$ is increased towards zero. As a result, optimal prices are $\bar{p} = p_0 = 0$ and $p_3 = (2\alpha' + v_0 - U^*)/\alpha'^2$. Substituting for prices, the IC constraint is equivalent to the assumption $U^* \in [v_0 + \alpha', v_0 + 2\alpha']$ and hence is satisfied. Similarly, the NFL lower bound is equivalent to $U^* \leq v_0 + 2\alpha'$ and so is satisfied. Substituting for prices, the markup is

$$\mu(U^*) = 2\alpha(1-c) + 2\left(\alpha/\alpha'\right)\left(\alpha - \alpha'\right) - \left(\alpha/\alpha'\right)^2 (U^* - v_0). \tag{60}$$

Comparing the markups derived above to those derived in the proof of Proposition 10 in equation (55) shows that markups are weakly higher in the inattentive case for all U^*. For $U^* \in [v_0 + \alpha'(2-\alpha'), v_0 + 2\alpha']$ the contracts and markups are identical. For $U^* \in [0, v_0 + \alpha'(2-\alpha'))$, the markup up is strictly higher in the inattentive case. To show the latter, denote the inattentive markup by μ^I and the attentive markup by μ^A. For $U^* \in [0, v_0]$,

$$\mu^I - \mu^A = \frac{(\alpha - \alpha')^2}{\alpha'(1-\alpha')} (1 - (U^* - v_0)),$$

which is strictly positive because $(U^* - v_0) < 0$. For $U^* \in [v_0, v_0 + \alpha']$,

$$\mu^I - \mu^A = \frac{(\alpha - \alpha')^2}{\alpha'(1-\alpha')} (1 - (U^* - v_0)) + \frac{1}{\alpha'} (1-\alpha)\left(\alpha - \alpha'\right)(U^* - v_0),$$

which is strictly positive because $(U^* - v_0) \in [0, 1)$. For $U^* \in [v_0 + \alpha', v_0 + \alpha'(2-\alpha'))$,

$$\mu^I - \mu^A = \alpha\left(\alpha - \alpha'\right)\left(\alpha'\left(2-\alpha'\right) + v_0 - U^*\right)/\alpha'^2,$$

which is strictly positive because $U^* < v_0 + \alpha'(2-\alpha')$.

This comparison shows that firms always weakly prefer nondisclosure of $\left\{q^{t-1}, t\right\}$ and do so strictly for $U^* \in [0, v_0 + \alpha'(2-\alpha'))$.

A.14 Proof of Proposition 13

The argument closely follows that of Proposition 11. The difference is that there are two pricing regions rather than four, and the associated first-order conditions are: (1) For $U^A < v_0 + \alpha'$,

$$U^A = \frac{1}{2}\left(U^B + v_0 - \tau\right) + \frac{2\alpha\,(1-c) + Y}{2\,(1+Y)}.$$

(2) For $U^A > v_0 + \alpha'$,

$$U^A = \frac{1}{2}\left(U^B + v_0 - \tau\right) + \alpha'\left(1 - c\alpha'/\alpha\right).$$

Otherwise profits are maximized at a kink for $U^A = v_0 + \alpha'$.

The corresponding symmetric solutions to the first order conditions are:

Region	Symmetric Solution, $U^* =$	$\mu =$	Relevant Range
1	$v_0 - \tau + (2\alpha\,(1-c) + Y)\,/\,(1+Y)$	$(1+Y)\,\tau$	$\tau > (2\alpha\,(1-c) + Y)\,/\,(1+Y) - \alpha'$
2	$v_0 - \tau + 2\alpha'\,(1 - c\alpha'/\alpha)$	$(\alpha/\alpha')^2\,\tau$	$0 \leq \tau \leq (\alpha'/\alpha)\,(\alpha - 2c\alpha')$

(61)

For intermediate values of $\tau \in [(\alpha'/\alpha)\,(\alpha - 2c\alpha')\,, (2\alpha\,(1-c) + Y)\,/\,(1+Y) - \alpha']$,

$$U^* = v_0 + \alpha' \tag{62}$$

and

$$\mu = \left(\alpha/\alpha'\right)\left(\alpha - 2c\alpha'\right). \tag{63}$$

Markups are derived by plugging the values of U^* for the relevant regions into equations (59)-(60).

The preference for surprise penalty fees follows from Proposition 12. The condition for a strict preference, $U^* < v_0 + \alpha'\,(2 - \alpha')$, corresponds to region 2, and therefore (substituting U^* from region 2) that $\tau > (\alpha')^2\,(\alpha - 2c)\,/\alpha$.

A.15 Proof of Corollary 3

(1) Full market coverage result: A sufficient condition for full market coverage is that the equilibrium offered utilities U^* characterized in the proofs of Propositions 11 and 13 under the assumption of full market coverage satisfy $U^* > \tau/2$. Inspection of equations (61), (62), (61), and (58) show that for all levels of τ, $U^* \geq v_0 - \tau$. Note that $v_0 - \tau > \tau/2$ is equivalent to $\tau < (2/3)\,v_0$ which is true by assumption. Therefore there is full market coverage in equilibrium.

(2) Sufficiently large bias result (α'/α is sufficiently small): First, consider the inattentive case. By Proposition 13, the minimum equilibrium markup with full market coverage is $(1+Y)\,\tau$. Taking

α'/α small implies taking α' to zero. Since $\lim_{\alpha' \to 0} Y = \infty$, taking α' to zero holding $\tau > 0$ fixed implies that the lower bound on markup, $(1+Y)\tau$, tends to infinity. Since all served consumers are exploited whenever the markup exceeds first best surplus of $S^{FB} = v_0 + 2\alpha(1-c)$, this implies all consumers are exploited for sufficiently large bias. This exploitation must be eliminated by PPR because Proposition 10 guarantees that attentive consumers are not exploited. Since there is still full market coverage this must be due to a strict reduction in markup.

(3) Severe and mild bias results: Denote the inattentive markup for competitive region 1/2 given by equation (63) as $\mu^I_{1/2}$, and extend the notation (superscript "I "for inattentive, or "A" for attentive and subscript for competitive region) for other markups as well. I now rank pairs of the five markups $\mu^I_{1/2}$, $\mu^A_{2/3}$, $\mu^A_{1/2}$, μ^I_1, and μ^A_2 which are given by equations (58) and (63) and Propositions 11 and 13. First, $\mu^I_{1/2} > \mu^A_{2/3}$, because

$$\mu^I_{1/2} - \mu^A_{2/3} = \left(\alpha/\alpha'\right)\left(1-\alpha'\right)\left(\alpha-\alpha'\right) > 0.$$

Second, $\mu^I_{1/2} > \mu^A_{1/2}$ given $\alpha'/\alpha < 1/2$ but $\mu^I_{1/2} < \mu^A_{1/2}$ given $\alpha'/\alpha > 1/2$. This follows because

$$\mu^I_{1/2} - \mu^A_{1/2} = \left(\alpha/\alpha'\right)\left(\alpha - 2\alpha'\right).$$

Third, $\mu^I_1 > \mu^A_2$ given $\alpha'/\alpha < (2\alpha-1)/\alpha^2$ but $\mu^I_1 < \mu^A_2$ given $\alpha'/\alpha > (2\alpha-1)/\alpha^2$. This follows because

$$\mu^I_1 - \mu^A_2 = (1+Y)\tau - \left(\alpha/\alpha'\right)\left(1-\alpha+\alpha'\right)\tau,$$

which can be simplified to

$$\mu^I_1 - \mu^A_2 = \frac{\alpha-\alpha'}{\alpha'(1-\alpha')}\left(2\alpha - 1 - \alpha\alpha'\right)\tau.$$

Putting the last two markup rankings together implies that given severe bias $\left(\alpha'/\alpha < \max\left\{1/2, (2\alpha-1)/\alpha^2\right\}\right)$ $\mu^I_{1/2} > \mu^A_{1/2}$ or $\mu^I_1 > \mu^A_2$. Similarly, given mild bias $\left(\alpha'/\alpha > \max\left\{1/2, (2\alpha-1)/\alpha^2\right\}\right)$ $\mu^I_{1/2} < \mu^A_{1/2}$ and $\mu^I_1 < \mu^A_2$.

Comparing equilibrium markups derived in Propositions 11 and 13, for attentive and inattentive cases respectively, shows that markups coincide for $\tau \leq \left(\alpha'^2/\alpha\right)(\alpha-2c)$. (Inattentive region 2 extends to higher τ than does attentive region 4 since $\left(\alpha'^2/\alpha\right)(\alpha-2c) < (\alpha'/\alpha)(\alpha-2c\alpha')$.) For τ slightly above this range, inattentive markups are strictly higher.

Given $\mu^I_{1/2} > \mu^A_{2/3}$ and $\mu^I_1 > \mu^A_1$, severe bias ($\mu^I_{1/2} > \mu^A_{1/2}$ or $\mu^I_1 > \mu^A_2$) is sufficient to ensure that inattentive and attentive markups never cross again (inattentive markup remains strictly higher) for $\tau > \left(\alpha'^2/\alpha\right)(\alpha-2c)$. (Note that for $c \geq \alpha/2$, $\left(\alpha'^2/\alpha\right)(\alpha-2c) \leq 0$, so $\tau > \left(\alpha'^2/\alpha\right)(\alpha-2c)$

is implied by $\tau > 0$.) For mild bias, the two markups will intersect at some $\tau > \left(\alpha'^2/\alpha\right)(\alpha - 2c)$. The ranking $\mu^A_{2/3} < \mu^I_{1/2} < \mu^A_{1/2}$ implies that the first intersection will be at τ_1, where $\mu^I_{1/2} = \mu^A_2$. The ranking $\mu^A_2 > \mu^I_1 > \mu^A_1$ implies that the second intersection will be at τ_2, where $\mu^I_1 = \mu^A_{1/2}$. Taking the expressions for these markups from equations (58) and (63) and Propositions 11 and 13 and solving for τ_1 and τ_2 yields:

$$\{\tau_1, \tau_2\} = \left\{ \frac{1 - \alpha + \alpha'}{\alpha - 2c\alpha'}, \frac{2\alpha(1-c)}{1+Y} \right\}.$$

For mild bias, the attentive markup is strictly higher in the interval (τ_1, τ_2).