Penalty Pricing:

Optimal Price-Posting Regulation with Inattentive Consumers

Michael D. Grubb
MIT Sloan School of Management
Cambridge, MA 02142

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 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 base model which serves as a benchmark for the rest of the paper. The base model 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 base 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-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 base 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, 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 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. But note that these assumptions about beliefs and competition 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"

[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.

(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 base 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). Section 5 enriches the base 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

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 base 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.

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 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.

[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 Agrawal's finding that accruing one credit card late penalty fee reduces the likelihood of incurring one in the following month.

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 rst 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, the solution to my benchmark model with attentive consumers corresponds to a repetition of the Courty and Li (2000) solution and is nearly a special case of Pavan et al.'s (2009) model, although I assume two ex ante types at the contracting stage rather than a continuum. I go further, however, by solving 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 su cient competition in nonlinear price-schedules leads to two-part-tari pricing at marginal cost and rst-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 nd an analogous result in my attentive model with competitive sequential screening. The rst-best-allocation result (although not the two-part-tari -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 di er 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 rather than xed fees. Moreover, inattention exacerbates the softening of competition due to biased beliefs and makes consumers even worse o . Ellison (2005) shows that shrouded add-on fees can soften price competition without biased beliefs, if the consumers most price sensitive to cuts in xed 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 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]

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

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:

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

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: $\mathbf{q}(\mathbf{v};\mathbf{p})$. 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 $S^i = v_0 + E\left[\sum_{t=1}^{2}(v_t - c)\, q_t\left(\mathbf{v};\mathbf{p}^i\right)\right]$ 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 $c \in (0,1)$ per unit served. Thus firm i's expected profits are

$\pi^i = G\left(U^i;U^{-i}\right)$ .849i [(.849 Td s2)]TJ//F19 7.9.9091 T0.9091 Tf 10.3d [(q)]T[

cost $c \in (0,1)$ per unit served. Thus firm i's expected profits

3.2 Consumer Strategies

The rst step in analyzing the game is to solve the consumers problem. As I do so below, I suppress the rm 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 $v \quad q^{t \quad 1}; t$ which is a function of the date t and the vector of past usage choices $q^{t \quad 1}$. Let the period one and 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 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.

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^* : $q^I(v; p) = (1_{v_1 \geq v^*}; 1_{v_2 \geq v^*})$. 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:*

$$v^* = \frac{p_1 + p_2}{2} + (1 - F(v^*))\, p_3. \tag{4}$$

Equation (4) 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 (4) 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 $U(b^*; b^*)$:

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

mentation stage, the resulting payo $U(b\ ;b)$ is maximized at $b = b$.

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

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

$$\frac{dU(b\ ;b)}{db(v}$$

Definition 2 *Constant-Marginal-Price Regulation (CMPR) is the requirement that firms charge a constant marginal price as a function of usage:* $p_3 = 0$*.*

It will be a recurring result throughout the paper that firms optimally offer attentive consumers two-part tariffs with zero penalty fees. Thus 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.

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

4 Unbiased but ex ante Heterogeneous Consumers

In this section, I relax the assumption of ex ante homogeneity imposed in the base 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 1 firms. At the contracting stage ($t = 0$), each consumer privately receives one of two private signals $s \in \{L, H\}$, where $\Pr(s = H) =$. 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 $[0;$

consumption choices thereafter is $U^i_{s\hat{s}} = E\ u\ \mathbf{q}\ \mathbf{v}; \mathbf{p}^i_{\hat{s}}\ ; \mathbf{v}; \hat{s}\ j\, s$. De ne $U^i_s \quad U^i_{ss}$ to be the expected base utility of a consumer who chooses the intended contract from rm i

types' allocation is distorted downwards below first best. For $_H < _L$, low-types receive first best, while high types' allocation is distorted upwards.

The knife-edge efficiency-result in Proposition 3 and Corollary 2 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 $_L < _H$, 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

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 (4):

$$v_{s\hat{s}} = \frac{p_{1\hat{s}} + p_{2\hat{s}}}{2} + p_{3\hat{s}} \left(1 - F_s\left(v_{s\hat{s}}\right)\right). \tag{5}$$

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

$$U_{s\hat{s}} = v_0 - p_{0\hat{s}} + 2\int_{v_{s\hat{s}}}^{1} v dF_s(v) - \left(p_{1\hat{s}} + p_{2\hat{s}}\right)\left(1 - F_s\left(v_{s\hat{s}}\right)\right) - p_{3\hat{s}}\left(1 - F_s\left(v_{s\hat{s}}\right)\right)^2, \tag{6}$$

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

$$S_s = \int_{v_s}^{1} (v - c)\, dF_s(v). \tag{7}$$

Dene

De ne s[1 TDe ne 1

the base fee p_{0s} is given by equation (8):

$$p_{0s} = -U_s + v_0 + 2\int_{v_s}^{1} v dF_s(v) - 2p_s(1 - F_s(v_s)) - p_{3s}(1 - F_s(v_s))^2. \quad (8)$$

Second, it is convenient to think of the firm first choosing consumer threshold v_s and then choosing the best marginal prices 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[12] for p_s to satisfy the first order condition:

$$p_s = v_s - p_{3s}(1 - F_s(v_s)). \quad (9)$$

The firm's problem can be written as:

$$\max_{\substack{U_L, v_L, p_{3L} \\ U_H, v_H, p_{3H}}} ((1-\alpha) G_L(U_L)(S_L(v_L) - U_L) + \alpha G_H(U_H)(S_H(v_H) - U_H))$$

$$\text{s.t. } U_s \geq U_{s\hat{s}} \;\forall s, \hat{s} \in \{L, H\},$$

$$v_s \in \arg\max_x 2\int_x^1 v f_s(v)\, dv - 2p_s(1 - F_s(x)) - p_{3s}(1 - F_s(x))^2$$

where $U_{s\hat{s}}$, S_s, p_{0s}, and p_s are given by equations (6) through (9).

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 (9) 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 4 *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 (8-9) into equation (6) yields

$$U_{s\hat{s}} = U_{\hat{s}} + 2\int_{v_{s\hat{s}}}^{1} (v - v_{\hat{s}})\, dF_s(v) - 2\int_{v_{\hat{s}}}^{1} (v - v_{\hat{s}})\, dF_{\hat{s}}(v) - p_{3\hat{s}}(F_{\hat{s}}(v_{\hat{s}}) - F_s(v_{s\hat{s}}))^2. \quad (10)$$

By the envelope condition:

$$\frac{d}{dp_{3\hat{s}}} U_{s\hat{s}} = \frac{\partial}{\partial p_{3\hat{s}}} U_{s\hat{s}} = -(F_{\hat{s}}(v_{\hat{s}}) - F_s(v_{s\hat{s}}))^2 \leq 0.^{13} \quad (11)$$

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

■

Proposition 4 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.[14]

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 $_L = {}_H$. With inattentive consumers this is no longer true. Slack ex ante incentive-constraints and first-best allocations are a feature for $({}_H - {}_L)$ in an interval around zero. This can be achieved because strictly positive penalty fees relax the ex ante incentive constraints when *incentive-constraint (IC-H) binds and the low-type's allocation is distorted downwards below rst 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 rst order condition:*

$$v_L = c + \frac{}{1} \frac{F_L(v_L) \quad F_H(v_{HL})}{f_L(v_L)} \frac{@ \quad =@U_H}{G_L(U_L)} \quad (1 + p_{3L} f_L(v_L)) + \frac{1}{2}(F_L(v_L) \quad F_H(v_{HL}))\, h_L^{0}(v_L) \quad , \tag{13}$$

where $v_{HL} = v_L + p_{3L}(F_L(v_L) \quad F_H(v_{HL}))$*. (3) If* $_H \quad _L < \quad X_L$*, then the low type receives the rst best allocation* $v_L = c$ *and any weakly positive penalty fee* $p_{3L} \quad 0$ *is optimal on the low contract. However the upward incentive-constraint (IC-L) binds and the high type's allocation is distorted upwards above rst 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 rst order condition:*

$$v_H = c \quad \frac{1}{} \frac{F_L(v_{LH}) \quad F_H(v_H)}{f}$$

Corollary 3 *Let duopolists compete on a uniform Hotelling line, high types have transportation costs $_H = H$ strictly higher than low types $_L = L$, and marginal cost c be strictly positive. If > 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 $_s = _s$. Moreover, the set of equilibrium prices includes $p^i_{1s} = p^i_{2s} = 0$ and $p^i_{3s} = c/(1 - F_s(c))$. (2) Price-posting regulation (or constant-marginal-price regulation) would strictly decrease welfare and firm profits. Low types would be losers while high types would be winners.*

The intuition behind the result in Corollary 3 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 (IC-H) 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 equilibria, 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.[15]

In contrast with fairly-competitive markets, sufficient market power implies that penalty fees and inattentive consumers do not produce efficient outcomes. Corollary 4 illustrates this for the zero outside option monopoly.

Corollary 4 *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: $v_L > c$.*

Proof. By assumption, $G_s(U_s) = 1_{U_s \geq 0}$ and 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 $\frac{\partial \Pi / \partial U_H}{G_L(U_L)} = 1$. 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

[15] Shifts in markups in each segment are already inversely weighted by shares of each segment and $(1 -)$ 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.

the same number of purchases, but at FOSD higher valuations. Thus $U_{HL} > U_L = U_H = 0$, which violates IC-H. ■

When there is su cient market power the impact of regulation becomes ambiguous. Let the rm be a monopolist serving consumers with zero outside option. Without a binding revenue rasing requirement, a regulator with su cient information and authority would optimally set marginal price equal to marginal cost to achieve e cient allocations. In this case inattention and price-posting regulation have no e ect 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 rm would: inattention allows revenues to be more e ciently 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 bene cial. Proposition 6 gives su cient conditions for such price-posting regulation to be bene cial and su cient conditions for price-posting regulation to be harmful.

Proposition 6 *Let the rm be a monopolist serving consumers with zero outside option (ZOOM). Suppose that there is an exogenous restriction that* p_{3s} $h_s(v_s)$ *for* $h_s(v_s) > 0$ *and* $h_s^{0}(v_s)$ 0*. Assume that* f_H *crosses* f_L *once from below at* $v = c > 0$*. (1) If* $c < c$ *and* f_H *is weakly decreasing above* c*, then for* > 0 *su ciently small, price-posting regulation improves welfare. (2) If* $c > c$ *and* $h_s(v_s) = p^{\max} > 0$*, then either for* $p^{\max}$ *su ciently small or for* f_H *weakly increasing above* c*, price-posting regulation reduces welfare.*

4.4 Constraints on penalty fees

Corollary 3 shows that, given su cient competition, case (1) of Proposition 5 applies, ex ante incentive constraints are slack, and nite penalty fees are optimal. Thus with su cient competition, restrictions on penalty fees do not bind, and the precise form of restriction does not matter. Hence Corollary 3 and the result it highlights { that in competitive markets the combination of penalty fees and consumer inattention can be socially valuable { are robust to a variety of restrictions on penalty fees.

When equation (12) isn't satis ed 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 in nity base marginal prices and positive in nity 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 pro ts 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 $p_{0s} + p_{1s} + p_{2s} + p$

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 attention) with zero value for the service. The existence of such potential customers imposes what 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 $p_{0s} + p_{1s} + p_{2s} + p_{3s} \geq 0$. 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 $v_{Otherwise,}$

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 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. 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. 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

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)). Return to the assumption in the base model that consumers all have the same distribution of taste shocks F. Now, however, assume that consumers believe that the distribution is $\hat{F}$, which like the true distribution F is continuous and strictly increasing on $[0;1]$. Moreover, assume that F first-order-stochastically-dominates $\hat{F}$ so that consumers underestimate their demand for the add-on services.[17]

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

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

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

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

to consumer beliefs:

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

The fraction $G\left(\hat{U}^i;\hat{U}^{-i}\right)$ 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:

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

Thus firm i's expected profits are

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

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

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

5.1.1 Attentive benchmark

Proposition 7 characterizes optimal pricing in the attentive case.[18]

Proposition 7 *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 above marginal cost,*

$$p = c + \frac{\hat{F}(p) - F(p)}{f(p)} > c,$$

and allocations are inefficiently low. All consumers are weakly better off than choosing their outside options, and all transactions generate positive surplus.

Proposition 7 shows the potential for biased beliefs to reduce welfare in the absence of inattention by distorting consumption downwards. It 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.

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

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,

$$\hat{v} = \frac{p_1 + p_2}{2} + p_3 \left(1 - \hat{F}(\hat{v})\right) \tag{16}$$

which substitutes consumer beliefs in place of the true distribution of tastes in equation (4). As before, I focus on symmetric pricing $p_1 = p_2 = p$ 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 $\hat{U}$ rather than setting fixed fee p_0. In this case the fixed fee p_0 is given by equation (17):

$$p_0 = -\hat{U} + v_0 + 2\int_{\hat{v}}^{\infty} v\, d\hat{F}(v) - 2p\left(1 - \hat{F}(\hat{v})\right) - p_3\left(1 - \hat{F}(\hat{v})\right)^2. \tag{17}$$

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

$$p = \hat{v} - p_3\left(1 - \hat{F}(\hat{v})\right). \tag{18}$$

Using equations (17) and (18), firm profits can be written as a function of perceived expected-utility $\hat{U}$, penalty p_3, and consumer threshold $\hat{v}$:

$$\pi = G(\hat{U}) \left[-\hat{U} + 2\int_{\hat{v}}^{\infty} \left(v - c - \frac{\hat{F}(v) - F(v)}{f(v)} \right) f(v)\, dv + p_3\left(\hat{F}(\hat{v}) - F(\hat{v})\right)^2 \right]. \tag{19}$$

Note that profits increase linearly in the penalty fee p_3. Thus the optimal penalty fee will be positive, in which case the local incentive constraint of equation (18) is sufficient for $\hat{v}$ to be globally optimal. Moreover, without any additional constraints, firms optimally choose $p_3 = \infty$ and $\hat{v} \in (0, \infty)$. 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 $p^{\max}$. The rm's problem can then be written as:

$$\max_{U ;v ;p_3} G(U) \quad U + 2 \int_{v}^{1} v \quad c \quad \frac{F (v) \quad F(v)}{f(v)} \quad f(v)\,dv + p_3 (F (v) \quad F (v$$

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 and zero otherwise. Consumers underestimate their demand and believe that v_t equals 1 with probability $^{0} <$. 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 benchmark

(2) Conditional on o ering U 2 $[v_0; v_0 + 2 \ ^{0}]$, optimal prices and markup are:

$$p_3 = p_0 = 0, \; p_1 = p_2 = 1 \quad (U \quad v_0) = 2 \ ^{0},$$

$$(U \ ) = \ S^{FB} \quad U \qquad = \ ^{0} \quad 1 \ (U \quad v_0) \ . \qquad (21)$$

(3) O ering U $> v_0 + 2 \ ^{0}$ is not feasible under NFL.

The no-free-lunch constraint requires that all payments from consumers to the rm be non-negative. Increasing perceived utility U while holding the allocation xed at rst best entails lowering prices. Hence the no-free-lunch constraint is increasingly di cult to satisfy as the o ered U rises. This explains the three pricing regions in Proposition 9. Absent the NFL constraint, it would always be optimal to charge the maximum marginal price that induces rst-best consumption (p

The rst result is that it will be optimal for rms to set prices which induce the e cient allocation $fb_0; b_1g = f0; 1g$.

Lemma 1 *Given Bernoulli taste shocks, inattentive consumers who underestimate demand (^{0} <), c 2* (0; 1)*, and the no-free-lunch constraint, rms set prices which induce the e cient allocation: consumers buy if and only if $v_t = 1$.*

To induce the e cient allocation, the rm must set expected marginal price conditional on a purchase to be between zero and one: 0 p + $^0 p_3$ 1. Applying Lemma 1, the rm's problem can thus be reduced to the following:[19]

$$
\begin{array}{rcl}
\max\limits_{U \ ;p;p_3} & = & G(U \ ) \ \ p_0 + 2 \ \ (p \quad c) + \ \ ^2 p_3 \\
\text{such that} & : & \\
\text{IC:} & & 0 \quad p + \ \ ^0 p_3 \quad 1 \\
\text{NFL:} & & p_0 \quad 0,\ p_0 + p \quad 0,\ p_0 + 2p + p_3 \quad 0 \\
\text{Fixed Fee} & : & p_0 = \quad U \ \ + v_0 + 2 \ \ ^0 (1 \quad p) \qquad ^{02} p_3
\end{array}
$$

Proposition 10 characterizes optimal prices given a xed perceived-expected-utility U

$$(U\ ) = \ S^{FB} \quad U \qquad = {}^{0\,2} \quad 1 \ \ (U \quad v_0) + 2 \qquad {}^{0} \ = {}^{0} \ . \tag{25}$$

(3) Offering $U \ > v_0 + 2\ ^0$ is not feasible under NFL.

Propositions 9 and 10 characterize optimal prices and markup $(U\ )$ as a function of perceived expected-utility U . Corollary 5 applies Propositions 9 and 10 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 5 *Assume a zero-outside-option monopoly, the no-free-lunch constraint, Bernoulli taste-shocks, consumers who underestimate demand ($^0 <$), and $c\ 2\ (0;1)$. 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 ($= S^{FB}$, $CS = 0$). Let $Y \quad (\quad {}^0)^2 = ({}^0 (1 \quad {}^0))$. If consumers are inattentive, the monopolist charges $p_1 = p_2 = \ \ p_0 = \ \ (v_0 + \ {}^0) = (1 \quad {}^0)$ and $p_3 = (v_0 + 1) = ({}^0 (1 \quad {}^0))$. While still inducing efficient consumption, the monopolist now captures more than the entire first best surplus ($= S^{FB} + (1 + v_0)\,Y$) 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 9 and 10 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 5, 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. The no-free-lunch constraint, however, restricts fixed fees to be non-negative. Once firms reduce fixed fees to zero, they are forced to compete on either base marginal charges or penalty fees. This softens price competition and raises profits, because consumers underweight the chance of paying both base marginal charges and penalty fees and hence are less price sensitive to them than to fixed fees. The effect is larger for penalty fees, used with inattentive consumers, than with base marginal charges, used with attentive consumers. As a result, the no-free-lunch constraint implies that the redistributive effects of price-posting regulation persist under competition.

		Inattentive	Attentive (PPR)	Redistribution
Monopoly		S^{FB} + (1 + v_0) Y	S^{FB}	(1 + v_0) Y
(ZOOM)	U	(1 + v_0) Y	0	
Duopoly (Hotelling)		(= ^{0})^2	(= ^{0})	(= ^{0}) (= ^{0} 1)
For < ^{0}(1 2c)	U	S^{FB} (= ^{0})^2	S^{FB} (= ^{0})	

Table 1: Summary of surplus distribution results from Corollaries 5 and 6. Pro ts, and consumer surplus under zero outside option monopoly and Hotelling duopoly with and without price posting regulation.

Corollary 6 applies Propositions 9 and 10 to a fairly competitive Hotelling duopoly, solves for equilibrium utility o ers U , and compares attentive and inattentive cases to evaluate the e ect of price-posting regulation.

Corollary 6

		Inattentive	Attentive (PPR)	Redistribution
Monopoly	$fp_0; p; p_3g$	$f{:}1; \quad {:}1; 11{:}1g$	$f0; 1; 0g$	
(ZOOM)		$8{:}1$	1	$7{:}1$

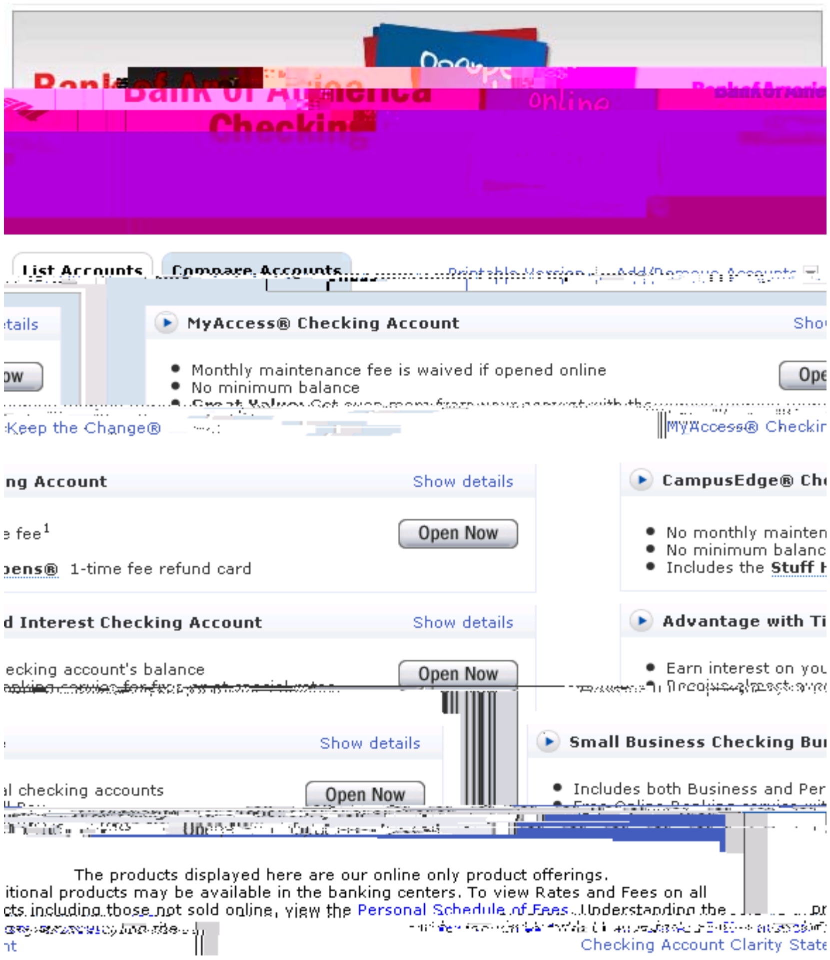

Figure 1: Bank of America's menu of 4 checking accounts, o ered online at `www.bankofamerica.com` on March 1, 2010.

Other Account Fees			
Fee Category	**Fee Name / Description**	**Fee Amount**	**Accounts Qualifying for Waiver of this Fee**
Overdraft Items (an overdraft item)			

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.

References

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.

Edlin, Aaron S. and Chris Shannon, \Strict monotonicity in comparative statics," *Journal of Economic Theory*, 1998, *81* (1), 201{219.

Eliaz, K r 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 nal rule," Technical Report November 11 2009.

—— , \Federal Reserve announces nal 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.

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 Overcon dent Consumers," *American Economic Review*, 2009, *99* (5), 1770{1807.

—— **and Matthew Osborne**, \Cellular service demand: Tari 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.

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

Liebman, Je rey 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," Technical Report, The New York Times 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.

Stole, Lars A.

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:

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

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 $v > p_1 + \int_{p_2}^{p_2+p_3} (v_2 - p_2) f(v_2)\, dv_2 + (1 - F(p_2 + p_3))\, p_3$, $\frac{dU}{dv} < 0$ 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 (4) at $c = v^*$ since equation (4) is sufficient as well as necessary for incentive compatibility given $p_3 \geq 0$.

A.3 Proof of Proposition 3

The results in the paper are stated for the case $T = 2$. However, the proofs in this section are written for the more general case $T \geq 1$.

At time 0, consumers receive signals $s \in \{L, H\}$ ($\Pr(s = H) =$ [illegible]) and choose a tariff $\hat{s}$. At time $t \in \{1, 2, ..., T\}$ consumers realize taste shock $v_t \mid s \sim^{iid} F_s(v_t)$ make report $\hat{v}_t$ and receive allocation $q_t(\hat{s}, \hat{v}^t) \in [0, 1]$ (the probability of receiving the unit), where $\hat{v}$ is the vector of reports to date $[\hat{v}_1, ..., \hat{v}_t]$. At time T, consumers pay $P(\hat{s}, \hat{v}^T)$. Define $U_t(s, \hat{s}, v^t, \hat{v}^t)$ to be expected utility at time t conditional on realizations (s, v^t) and reports $(\hat{s}, \hat{v}^t)$ to date as well as a plan to report truthfully from $t+1$ onwards. Consumers utility is quasi-linear and time separable, with unit demand each period, so that $U_T\left(s,\hat{s},v^T,\hat{v}^T\right) = \sum_{t=1}^T v_t q_t\left(\hat{s},\hat{v}^t\right) - P\left(\hat{s},\hat{v}^T\right)$. Moreover let

$$U_t\left(s,v^t,\hat{s},\hat{v}^t\right) = E\left[v_t q_t\left(\hat{s},\hat{v}^t\right) + \sum_{\tau=t+1}^{T} v_\tau q_\tau\left(\hat{s},\hat{v}^t,v_{t+1},\ldots,v_\tau\right) - P\left(\hat{s},\hat{v}^t,v_{t+1},\ldots,v_T\right) \mid s,v^t,\hat{s},\hat{v}^t\right]$$

for $t \geq 1$ and let $U_0(s,\hat{s})$ be the expected utility of someone who has signal s, reports $\hat{s}$, and reports all v^T truthfully:

$$U_0(s,\hat{s}) = E\left[\sum_{t=1}^T v_t q_t\left(\hat{s},v^t\right) - P\left(\hat{s},v^T\right) \mid s,\hat{s}\right].$$

Let $U_{s\hat{s}} = U_0(s,\hat{s})$ and $U_s = U_0(s,s)$ be the expected utility of someone who plans to be entirely truthful conditional on realization of signal s. Let $G_s(U_s)$ be the fraction of consumers of type s with outside option below U_s. Let costs be $C\left(q^T\right) = c\sum_{t=1}^T q_t$, so that surplus is $\sum_{t=1}^T (v_t - c)\, q_t$. Define S_s to be the expected surplus from a type s consumer who reports truthfully: $S_s = E\left[\sum_{t=1}^T (v_t - c)\, q_t\left(s,v^t\right) \mid s\right]$.

Invoking the revelation principle, the monopolist's problem may then be written as:

$$\max_{\substack{q^T\left(s,v^T\right)\in[0,1] \\ P\left(s,v^T\right)}} (1-\alpha)\, G_L(U_L)(S_L - U_L) + \alpha\, G_H(U_H)(S_H - U_H)$$

such that

1. Truthful history IC $t \geq 1$: $U_t\left(s,s,v^t,v^t\right) \geq U_t\left(s,s,v^t,\ldots\right.$

compatibility. ■

Lemma 3 *Local IC: A necessary condition for incentive compatibility is* $\frac{d}{dv_t}U_t\left(s;s;v^t\right) = \frac{\partial}{\partial v_t}U_t\left(s;s;v^t\right) = q_t\left(s;v^t\right)$.

Proof. This follows from conditional independence of v_t and application of an envelope theorem, which is valid by Proposition 1 of Pavan et al. (2009), since my setting fits within the Pavan et al. (2009) framework for $t \geq 1$. ■

I begin by solving a relaxed problem. I impose monotonicity ($q_t\left(s;v^t\right)$ non-decreasing in v_t), local incentive compatibility for $t \geq 1$ ($\frac{d}{dv_t}U_t\left(s;s;v^t\right) = q_t\left(s;v^t\right)$), and an ex ante incentive constraints IC-H ($U_0\left(H;H\right) \geq U_0(H;L)$) and IC-L: $U_0\left(L;L\right) \geq U_0(L;H)$. However I relax all other incentive constraints. In particular, I am only checking incentive compatibility against one step deviations, rather than multiple step deviations. After solving the relaxed problem, I will need to check (1) incentive compatibility against multiple step deviations and (2) for global incentive compatibility of v^T reporting to confirm that the relaxed solution solves the original problem.

By the envelope condition $\frac{d}{dv_t}U_t\left(s;s;v^t\right) = \frac{\partial}{\partial v_t}U_t\left(s;s;v^t\right) = q_t\left(s;v^t\right)$ and the FTC,

$$U_t\left(s;s;v^t\right) = U_t\left(s;s;v^{t-1};\underline{v}_t\right) + \int_{\underline{v}_t}^{v_t} q_t\left(s;v^{t-1};x\right)dx.$$

Moreover, since

$$U_T\left(s;\hat{s};v^t;\underline{v}_{t+1}...\underline{v}_T\right) = \sum_{\tau=0}^{t} v_\tau q_\tau\left(\hat{s};v^\tau\right) + \sum_{\tau=t+1}^{T} \underline{v}_\tau q_\tau\left(\hat{s};v^t;\underline{v}_{t+1};...;\underline{v}_\tau\right) - P\left(\hat{s};v^t;\underline{v}_{t+1}...\underline{v}_T\right),$$

it is true that

$$\begin{aligned}
\frac{d}{dv_t}U_T\left(s;\hat{s};v^t;\underline{v}_{t+1}...\underline{v}_T\right) &= q_t\left(\hat{s};v^t\right) + \sum_{\tau=0}^{t} v_\tau \frac{d}{dv_t}q_\tau\left(\hat{s};v^\tau\right) + \sum_{\tau=t+1}^{T} \underline{v}_\tau \frac{d}{dv_t}q_\tau\left(\hat{s};v^t;\underline{v}_{t+1};...;\underline{v}_\tau\right) \\
&\quad - \frac{d}{dv_t}P\left(\hat{s};v^t;\underline{v}_{t+1}...\underline{v}_T\right) \\
&= \frac{d}{dv_t}U_T\left(\hat{s};\hat{s};v^t;\underline{v}_{t+1}...\underline{v}_T\right).
\end{aligned}$$

Thus, by FTC

$$U_T\left(s;\hat{s};v^t;\underline{v}_{t+1}...\underline{v}_T\right) = U_T\left(s;\hat{s};v^{t-1};\underline{v}_t...\underline{v}_T\right) + \int_{\underline{v}_t}^{v_t} q_t\left(\hat{s};v^{t-1};x\right)dx,$$

and

$$U_T\left(s;\hat{s};v^T\right) = U_T\left(s;\hat{s};v^{T-1};\underline{v}_T\right) + \int_{\underline{v}_T}^{v_T} q_T\left(\hat{s};v^{T-1};x\right)dx.$$

Now by iterated expectations,

$$U_{T-1}\left(s,\hat{s},v^{T-1}\right)=\int_{\underline{v}_T}^{\bar{v}_T}U_T\left(s,\hat{s},v^T\right)f_s\left(v_T\right)dv_T.$$

Substituting in the envelope condition and integrating by parts gives

$$U_{T-1}\left(s,\hat{s},v^{T-1}\right)=U_T\left(s,\hat{s},v^{T-1},\underline{v}_T\right)+\int_{\underline{v}_T}^{\bar{v}_T}q_T\left(\hat{s},v^T\right)\left(1-F_s\left(v_T\right)\right)dv_T.$$

This can now be repeated recursively to yield

$$U_0\left(s,\hat{s}\right)=U_T\left(s,\hat{s},\underline{v}_1,\underline{v}_2,\ldots,\underline{v}_T\right)+\sum_{t=1}^{T}\int_{\underline{v}_t}^{\bar{v}_t}q_t\left(\hat{s},v^t\right)\left(1-F_s\left(v_t\right)\right)dv_t. \tag{26}$$

Equation (26) pins down U_{HL} as a function of the allocation $q^T\left(L,v^T\right)$ and U_L, and thus IC-H is:

$$U_H\geq U_L+\sum_{t=1}^{T}\int_{\underline{v}_t}^{\bar{v}_t}q_t\left(L,v^t\right)\left(F_L\left(v_t\right)-F_H\left(v_t\right)\right)dv_t$$

Similarly, IC-L is

$$U_L\geq U_H-\sum_{t=1}^{T}\int_{\underline{v}_t}^{\bar{v}_t}q_t\left(H,v^t\right)\left(F_L\left(v_t\right)-F_H\left(v_t\right)\right)dv_t$$

or together:

$$\sum_{t=1}^{T}\int_{\underline{v}_t}^{\bar{v}_t}q_t\left(L,v^t\right)\left(F_L\left(v_t\right)-F_H\left(v_t\right)\right)dv_t\leq U_H-U_L\leq\sum_{t=1}^{T}\int_{\underline{v}_t}^{\bar{v}_t}q_t\left(H,v^t\right)\left(F_L\left(v_t\right)-F_H\left(v_t\right)\right)dv_t$$

Note that given FOSD, monotonicity of $q_t\left(s,v^t\right)$ in s, $q_t\left(H,v^t\right)\geq q_t\left(L,v^t\right)$, implies implies IC-L follows from binding IC-H and vice versa. Also, given FOSD, IC-H implies $U_H\geq U_L$. Finally, either binding IC-H or binding IC-L implies $\frac{dU_H}{dU_L}=1$.

Lemma 4 *(1) IC-L slack implies* $q_t\left(H,v^t\right)=q^{FB}\left(v_t\right)$ *and (2) IC-H slack implies* $q_t\left(L,v^t\right)=q^{FB}\left(v_t\right)$.

Proof. (1) Suppose not. Then moving $q_t\left(H,v^t\right)$ towards $q^{FB}\left(v_t\right)$ a little bit while it2155[(A))-461(IC-0.9091

Case (1), $_L = \ _H$. Relax both IC-L and IC-H. Then allocations are rst best and unconstrained optimal markups $_L$, $_H$ are charged on each contract. Since both allocations and markups are the same, the L and H contracts are the same, and hence IC-L and IC-H are satis ed.

Case (2) $_H > \ _L$. Relax IC-L. By Lemma 4, $q_t \; H; v^t \; = q^{FB}(v_t)$.

Lemma 5 *Relaxed IC-L and* $_H > \ _L$ *imply IC-H is binding such that* $U_H = U_{HL}$ *in the relaxed problem.*

Proof. Suppose IC-H is slack: $U_H > U_{HL}$. Given IC-L is relaxed, Lemma 4 implies $q_t \; H; v^t \; = q_t \; L; v^t \; = q^{FB}(v_t)$g

The $q_t\left(L; v^t\right)$ first order condition is:

$$\frac{d}{dq_t\,(L; v^t)} = \frac{@}{@q_t\,(L; v^t)} + \frac{@}{@U_H}\frac{dU_H}{dq_t\,(L; v^t)} = \frac{@}{@q_t\,(L; v^t)} + \frac{@}{@U_H}\int_{\underline{v}_t}^{\overline{v}_t} \left(F_L\,(v_t) \quad F_H\,(v_t)\right) dv_t$$

$$= (1 \quad)\,G_L\,(U_L) \int_{\underline{v}_t}^{\overline{v}_t} (v_t \quad c)\, f_L\,(v_t)\, dv_t + \frac{@}{@U_H}\int_{\underline{v}_t}^{\overline{v}_t} \left(F_L\,(v_t) \quad F_H\,(v_t)\right) dv_t$$

$$= (1 \quad)\,G_L\,(U_L) \int_{\underline{v}_t}^{\overline{v}_t} v_t \quad c \quad \frac{}{1}\,\frac{F_L\,(v_t) \quad F_H\,(v_t)}{f_L\,(v_t)}\,\frac{@ \quad =@U_H}{G_L\,(U_L)} \quad f_L\,(v_t)\, dv_t.$$

Define v_t^A such that:

$$v_L^A = c + \frac{}{1}\,\frac{F_L \quad v_L^A \qquad F_H \quad v_L^A}{f_L \quad v_L^A}\,\frac{@ \quad =@U_H}{G_L\,(U_L)}. \tag{27}$$

Maximizing point-wise implies that $q_t\left(L; v^t\right) = 1$ if

$$v_t \quad c \quad \frac{}{1}\,\frac{F_L\,(v_t) \quad F_H\,(v_t)}{f_L\,(v_t)} \quad @ \quad =@U \qquad (U_L)$$

(assuming small enough that L types served) and also $G_H(U_H) = 1$ since IC-H implies IR. (H types can always mimic L types, but would have stochastically higher values for the same consumption by FOSD. Alternatively, $\sum_{t=1}^{T}\int_{\underline{v}_t}^{\bar{v}_t} q_t\left(L; v^t\right)\left(F_L(v_t) \quad F_H(v_t)\right) dv_t \quad 0$ by FOSD and $q \in [0;1]$). In this case $\frac{\partial \Pi}{\partial U_H} = $ and the first order condition for $q_t\left(L; v^t\right)$ reduces to

$$v_L^{CL} = c + \frac{}{1 \quad } \frac{F_L\left(v_L^{CL}\right) \quad F_H\left(v_L^{CL}\right)}{f_L\left(v_L^{CL}\right)},$$

which is the Courty and Li (2000) solution.

(2) Heterogeneous outside options (G_s differentiable etc.). Then

$$\frac{\partial}{\partial U_L} = (1 \quad)\, g_L(U_L)\,(S_L \quad U_L) \quad (1 \quad)\, G_L(U_L)$$

$$\frac{\partial}{\partial U_H} = \quad g_H(U_H)\left(S_H^{FB} \quad U_H\right) \quad G_H(U_H)$$

and the first order condition for U_L is

$$\frac{\partial}{\partial U_L} = \quad \frac{\partial}{\partial U_H}$$

Case (3)

market share in segment s is therefore $\frac{1}{2\ s}$ E ^{B} $^{A}_{s}$ + $_{s}$, and A's best response markup is $_{s}^{A}$ = $\frac{1}{2}$ E ^{B} + $_{s}$. Thus $_{L}^{A} \neq$ $_{H}^{A}$ and by Proposition 3 A's best response includes an ine cient contract.

(3a) If $_{H}$ > $_{L}$, then in all symmetric equilibria, high types receive rst best allocations, while low types' allocation is distorted downwards below rst best: We know that in a symmetric pure strategy equilibrium that for both rms, either $_{L}$ = $_{H}$, $_{L}$ < $_{H}$ or $_{L}$ > $_{H}$. Part (2) rules out $_{L}$ = $_{H}$ if $_{H}$ > $_{L}$. All that remains is to rule out $_{L}$ > $_{H}$, 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 $\theta_H \geq \theta_L$ at optimal offer $\{U_H; U_L\}$ which is satisfied by assumption.

(b) Substituting $v_L = c$ into equation (28), gives

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

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, $\left(S_H^{FB} - \hat{U}_H\right) - \left(S_L^{FB} - \hat{U}_L\right) = \theta_H - \theta_L$, IC-H simplifies to:

$$(\theta_H - \theta_L) \geq 2\int_c^{v_{HL}} (v - c)\, dF_H(v) + h_L(c)\left(F_L(c) - F_H(v_{HL})\right)^2.$$

Further, $v_{HL} = c + h_L(c)\left(F_L(c) - F_H(v_{HL})\right)$ is uniquely defined by the FOC from equation (5) for v_{HL}, where p_L is given by equation (9) and $p_{3L} = h_L(c)$. Note, if instead $p_{3L} = 0$, then IC-H

Taking derivatives and substituting equation (11) for $\frac{\partial U_H}{\partial p_{3L}}$ gives:

$$\frac{d}{dv_L} = 2(1 \quad) G_L(U_L)(v_L \quad c) f_L(v_L)$$
$$\frac{\partial}{\partial U_H} \Big[2(F_L(v_L) \quad F_H(v_{HL}))(1 + p_{3L} f_L(v_L)) + (F_L(v_L) \quad F_H(v_{HL}))^2 h'(v_L) \Big] .$$

The FOC $\frac{d\Pi}{dv_L} = 0$ simpli es to equation (13), or for non-negative marginal prices, $h_s(v_s) = v_s{=}(1 \quad F_s(v_s))$, to:

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

Similar to case (1), $v_{HL} = v_L + h_L(v_L)(F_L(v_L) \quad F_{=}^{=}$

first order condition for v_H can be re-written as

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

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

A.6 Proof of Corollary 3

(1) Show proposed equilibrium exists by construction: Impose $p_{3s} \quad h_s(v_s) = v_s = (1 \quad F_s(v_s))$. Assume that each firm offers $p_{3s} = h_s(c)$, $v_L = v_H = c$, and $U_s = S_s^{FB} \quad _s$. In this case, $U_s = \hat{U}_s$ and $_s = \ _s = \ _s$. As a result, $(\ _H \quad _L) = \ (H \quad L)$. For sufficiently small, this satisfies the condition for first best allocations in Proposition 5, which verifies that the proposed offers are best responses. If the constraint $p_{3s} \quad h_s(v_s)$ 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) $(\ _H \quad _L) < \ X_L$, (b) $(\ _H \quad _L) > X_H$, and (c) $(\ _H \quad _L) \in [\ X_L; X_H]$. Given (c), the proposed equilibrium is unique. A symmetric equilibrium in case (a) is ruled out by $_H > \ _L$ and pass-through rate less than 1 following a similar argument that was used in the proof of Corollary 2.[20]

allocations, so $\hat{}_{s} = \frac{G_s(\hat{U}_s)}{g_s(\hat{U}_s)} = \ _s$. As a result

$$_H \quad _L = \frac{G_H(\hat{U}_H)}{g_H(\hat{U}_H)} \quad \frac{G_L(\hat{U}_L)}{g_L(\hat{U}_L)} \quad \frac{G_H(U_H)}{g_H(U_H)} \quad \frac{G_L(U_L)}{g_L(U_L)} = \quad (H \quad L)$$

(where U_s means original utility o er, and $\hat{U}_s$ is the unconstrained optimal utility o er used in the new menu) and by Proposition 5, IC-H is satis ed for small

$f_L(v_L) < f_H(v_{HL})$. As $p^{\max}$ goes to zero, the constraint $p_{3L} \leq p^{\max}$ implies that v_{HL} approaches v_L and hence the inequality $f_L(v_L) < f_H(v_{HL})$ holds. Also imposing price-posting regulation could cause low types not to be served at all.

A.8 Proof of Proposition 7

The results in the paper are stated for the case $T = 2$. However, the proofs in this section are written for the more general case $T \geq 1$.

At time $t \in \{1, 2, \ldots, T\}$ consumers realize taste shock $v_t \overset{iid}{\sim} F(v_t)$ make report $\hat{v}_t$ and receive allocation $q_t(\hat{v}^t) \in [0, 1]$ (the probability of receiving the unit), where $\hat{v}$ is the vector of reports

Invoking the revelation principle, the monopolist's problem may then be written as:

$$\max_{\substack{q^T(v^T)\in[0,1] \\ P(v^T)}} G(U)(S - U)$$

such that

1. Truthful history IC $t \geq 1$ $\quad U_t\left(v^t; v^t\right) \geq U_t\left(v^t; [v^{t-1}; \hat{v}_t]\right) \quad \forall t; v^t$ and $\hat{v}_t$
2. Any history IC $\quad U_t\left(v^t; v^t\right) \geq U_t\left(v^t; \hat{v}^t\right) \quad \forall t; v^t$ and $\hat{v}_t$

Lemma 6 *Monotonicity: A necessary condition for incentive compatibility is that $q_t\left(v^t\right)$ be non-decreasing in v_t.*

Proof. Analogous to Lemma 2. ■

Lemma 7 *Local IC: A necessary condition for incentive compatibility is $\frac{d}{dv_t}U_t\left(v^t\right) = \frac{\partial}{\partial v_t}U_t\left(v^t\right) = q_t\left(v^t\right)$.*

Proof. Analogous to Lemma 3. ■

I begin by solving a relaxed problem. I impose monotonicity ($q_t\left(v^t\right)$ non-decreasing in v_t) and local incentive compatibility for $t \geq 1$ ($\frac{d}{dv_t}U_t\left(v^t\right) = q_t\left(v^t\right)$).

This can now be repeated recursively to yield

$$U = U_T(\underline{v}_1, \underline{v}_2, \ldots, \underline{v}_T) + \sum_{t=1}^{T} \int_{\underline{v}_t}^{\bar{v}_t} q_t \quad v^t \quad (1 \quad F \quad (v_t))\, dv_t. \tag{29}$$

Similarly,

$$U = U_T(\underline{v}_1, \underline{v}_2, \ldots, \underline{v}_T) + \sum_{t=1}^{T} \int_{\underline{v}_t}^{\bar{v}_t} q_t \quad v^t \quad (1 \quad F(v_t))\, dv_t. \tag{30}$$

Or, alternatively,

$$U_T(\underline{v}_1, \underline{v}_2, \ldots, \underline{v}_T) = U \qquad \sum_{t=1}^{T} \int_{\underline{v}_t}^{\bar{v}_t} q_t \quad v^t \quad (1 \quad F \quad (v_t))\, dv_t,$$

and

$$U = U \quad + \sum_{t=1}^{T} \int_{\underline{v}_t}^{\bar{v}_t} q_t \quad v^t \quad (F \quad (v_t) \quad F(v_t))\, dv_t.$$

Now we can simplify the doubly relaxed problem, by substituting the local incentive constraints summarized by equations (29-30) for $U_T(\underline{v}_1, \underline{v}_2, \ldots, \underline{v}_T)$ and U ing U

The q_t v^t rst order condition is:

$$\frac{d}{dq_t\,(v^t)} = G\,(U\ ) \quad \frac{dS}{dq_t\,(v^t)} \quad \int_{\underline{v}_t}^{\bar{v}_t} (F \quad (v_t) \quad F\,(v_t))\,dv_t \;!$$

$$= G\,(U\ ) \int_{\underline{v}_t}^{\bar{v}_t} (v_t \quad c)$$

Substituting equations (17) and (18) into equation (33), yields true expected utility as a function of U , v , and p_3:

$$U = U \quad + 2 \int_{v}^{1} \frac{F \quad (v) \quad F(v)}{f(v)} \quad f(v)\, dv \quad p_3 (F \quad (v \ ) \quad F(v \ ))^2 . \tag{35}$$

The rm's pro t function in equation (19) is then obtained by substituting equations (34) and (35) into the expression $\quad = G(U \ )(S \quad U)$.

The proof:

$p_2 = p$, perceived and actual payo s are:

$$U \quad = \quad p_0 + v_0 + 2 \;\; ^{0}(1 \quad p)$$

$$U \quad = \quad p_0 + v_0 + 2 \;\; (1 \quad p)$$

$$= G(U \;) \;\; S^{FB} \quad U \quad 2 \qquad ^{0} \; (1 \quad p)$$

Note that

$$U \quad U \;\; = 2 \qquad ^{0} \; (1 \quad p) \quad 0$$

which implies that there is no exploitation: U 0. Utility o er U is implemented by the following xed fee:

$$p_0 = v_0 \quad U \;\; + 2 \;\; ^{0}(1 \quad p) \, .$$

There are two cases to consider:

(2a) $U \;\; 2 \; [0; v_0]$: Ignoring the NFL constraint, pro ts are increasing in p. Thus the incentive constraint p 1 will bind which implies $p = 1$, $p_0 = v_0 \quad U$ and $(U \;) = S^{FB} \quad U$. Given $U \quad v_0$, this satis es NFL.

(2b) $U \;\; 2 \; (v_0; v_0 + 2 \;\; ^{0}]$: Pro ts are increasing in p, and the NFL constraint p_0 0 will bind before the incentive constraint0

$$p_0 = \quad U \quad + v_0 + 2 \quad {}^{0} \quad 2p \quad p_3.$$

If $U \quad v_0 + 2 \quad {}^{0}$, then this allocation can be implemented without violating the NFL constraint with prices $p_1 = p_2 = p_3 = 0$ and $p_0 = \quad U \quad + v_0 + 2 \quad {}^{0}$. If $U \quad > v_0 + 2 \quad {}^{0}$, then this allocation

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

$$p_3 = \frac{1 + v_0 \quad U}{{}^0 b_1 (1 \quad {}^0 b_1)}.$$

Given these prices, pro ts are strictly increasing in b_1,

$$\frac{d \quad {}_4}{db_1} = G(U \quad) \quad 2 \quad (1 \quad c) + \frac{b_1 (\qquad {}^0)^2}{1 \quad {}^0 b_1} p_3 \quad > 0,$$

so any NFL implementable allocation with b_1 2 $(0;1)$ is dominated by the e cient allocation.

A.12 Proof Proposition 10

NFL says prices can be no lower than $p_0 = p_1 = p_2 = p_3 = 0$, and hence o ered perceived utility U can be no higher than $v_0 + 2 \ {}^0$. Optimal pricing need only be characterized for U 2 $[0; v_0 + 2 \ {}^0]$. By Lemma 1, the rm will induce the e cient allocation, $b_0 = 0$, $b_1 = 1$. In this case, pro ts and xed fees are:

$$= G$$

p below $\frac{{}^{0}\ (U \quad v_0)}{1 \quad {}^{0}}$, the NFL upper bound is binding, and as shown under case 1, it is optimal to increase p. Thus the optimal prices are:

$$p = \quad p_0 = \frac{v_0 + \ {}^{0} \quad U}{1 \quad {}^{0}},\ p_3 = \frac{v_0 + 1 \quad U}{(1 \quad {}^{0})}$$

utilities of U^A and U^B respectively, market share of rm A is: $G\ U^A;U^B\ = \frac{1}{2}\ U^A \quad U^B +$ 0. Pro ts are

$$^A = G\ U^A;U^B \qquad U^A$$

where $\quad U^A$ is the markup derived in Proposition 9 in the attentive case, or Proposition 10 in the inattentive case. In particular, in the attentive case, $\quad U^A$ is given by equation (20) for $U^A\ 2\ [0;v_0]$ and by equation (21) for $U^A\ 2\ (v_0;v_0+2\ ^0]$. In the inattentive case, $\quad U^A$ is given by equation (24) for $U^A\ 2\ [0;v_0+\ ^0]$ and by equation (25) for $U^A\ 2\ (v_0+\ ^0;v_0+2\ ^0]$. In both attentive and inattentive cases, the pro t function is concave (with a kink at $U^A = v_0$ in the attentive case, and with a kink at $U^A = v_0 + \ ^0$ in the inattentive case), and hence rm A's best response is a continuous function of U^B. Away from the kink $d^2\ ^A{=}dU^{A2} = g\ U^A;U^B\ \ d\ {=}dU^A < 0$, and at the kink $d\ ^A{=}dU^A$ decreases. This follows since

$$\frac{d\ ^A}{dU^A} = g\ U^A;U^B \qquad U^A\ + G\ U^A;U^B\ \frac{d}{dU^A},$$

and while $G\ U^A;U^B$ is continuous and nonnegative, in the attentive case $d\ {=}dU^A$ decreases at $U^A = v_0$ (Since $d\ {=}dU^A = \ 1$ for $U^A < v_0$, $d\ {=}dU^A = \ (\ =\ ^0)$ for $U^A > v_0^0$ and $(\ =\ ^0) > 1$), and in the inattentive case $d\ {=}dU^A$ decreases at $U^A = v_0 + \ ^0$ (Since $d\ {=}dU^A = \ (1+Y)$ for $U^A < v_0 + \ ^0$, $d\ {=}dU^A = \ (\ =\ ^0)^2$ for $U^A > v_0 + \ ^0$ and $(\ =$

Since $dU^A = dU^B\ 2\ [0; 1)$, there is a unique pure strategy equilibrium, which is symmetric. This is true for both attentive and inattentive cases.

Attentive case: For an equilibrium with $U \quad > v_0$ and full market coverage, U solves $U \quad = \frac{1}{2}(U \qquad + v_0) + \frac{1}{2}(\ {}^0 = \ )\ S^{FB} \quad v_0$, which yields:

$$U \quad = 2\ {}^0(1 \quad c) + v_0 \qquad .$$

The condition $U \quad > v_0$ is equivalent to $\quad < 2\ {}^0(1 \quad c)$, and $U \quad > \ = 2$ (full market coverage) is equivalent to $\quad < \frac{4}{3}\ {}^0(1 \quad c) + \frac{2}{3}v_0$. The joint assumption $\quad < \frac{4}{3}\ {}^0(1 \quad c)$ and $v_0 \quad 0$ is su cient for both $U \quad > v_0$ and full market coverage. By Proposition 9, the markup is $\quad = \quad (\ = {}^0)$ and prices are $p_3 = p_0 = 0$ and $p_1 = p_2 = c + \ = 2\ {}^0$.

Inattentive case: For an equilibrium with $U \quad > v_0 + \ {}^0$ and full market coverage, U solves $U \quad = \frac{1}{2}(U \ + v_0 \qquad) + \ {}^0 \qquad c(\ {}^0 = \ )^2$, which yields:

$$U \quad = (v_0 \qquad) + 2\ {}^0\ 1 \quad c \quad {}^0 = \qquad .$$

The assumption $\quad < \ {}^0(1 \quad 2c(\ {}^0 = \ ))$ is necessary and su cient for the solution to satisfy $U \quad > v_0 + \ {}^0$. Moreover, it is su cient for full market coverage ($U \quad > \ = 2$) since given $v_0 \quad 0$

$${}^0\ 1 \quad 2c \quad {}^0 = \qquad < \ {}^0\ 1 \quad c \quad {}^0 = \qquad < \frac{4}{3}\ {}^0\ 1 \quad c \quad {}^0 = \qquad \frac{4}{3}\ {}^0\ 1 \quad c \quad {}^0 = \qquad + \frac{2}{3}v_0,$$

and $\quad < \frac{4}{3}\ {}^0(1 \quad c(\ {}^0 = \ )) + \frac{2}{3}v_0$ is equivalent to $U \quad > \ = 2$. By Proposition 10, the markup is $\quad = \quad (\ = {}^0)^2$ and prices are $p = p_0 = 0$ and $p_3 = 2c = \ + \ = {}^{02}$.

The assumption $\qquad {}^0(1 \quad 2c)$ is su cient for both $\quad < \frac{4}{3}\ {}^0(1 \quad c)$ and $\quad < \ {}^0(1 \quad 2c(\ {}^0 = \ ))$.