Using Information about Naivete to Price Discriminate*

Paul Heidhues
ESMT

Botond Kőszegi
Central European University

March 27, 2014

Abstract

We study the welfare effects of an increase in firms' information about consumer naivete in a simple reduced-form model in which competitive firms can introduce distortionary fees that naive consumers ignore when making purchase decisions. We establish that our model captures many markets—including those for credit, banking, hotels, gambling, and mobile phones—in which consumer naivete has been invoked as playing an important role. While in natural analogues with rational consumers (we show) seller information about consumers always increases social welfare, we identify arguably weak conditions under which seller information about consumer naivete *lowers* welfare. When some consumers are naive, a *participation distortion* arises because the additional profits from naive consumers lead competitive firms to lower transparent prices below cost, and an *exploitation distortion* arises from the socially costly hidden fees firms charge. Due to seller information about consumer naivete, (i) the participation distortion increases unless the demand curve is sufficiently concave; (ii) the exploitation distortion increases if the social cost of hidden fees satisfies decreasing absolute convexity; and (iii) the interaction of the two effects further lowers social welfare. We also identify conditions under which perfect seller information about consumer naivete hurts both types of consumers.

Keywords: sophistication, naivete, first-degree price discrimination, third-degree price discrimination, privacy

---

1 Introduction

A growing literature in behavioral economics documents that in a number of consumer retail markets, many individuals misunderstand key fees or other central product features.[1] An important aspect of markets where some but not all consumers are naive is that firms may engage in naivete-based price discrimination—the tailoring of offers to targeted consumers' level of naivete—using external information they have available about individuals. For instance, Gurun, Matvos and Seru (2013) document that the ads lenders directed at less sophisticated populations pushed more expensive mortgages, and Schoar and Ru (2014) find that the offers credit-card companies send to less educated consumers feature more back-loaded payments. More generally, because most theories in behavioral industrial organization imply that—as naive consumers are more profitable—firms' incentive to acquire and use information about naivete can be strong, and it is likely that some information firms have provides clues regarding consumers' naivete, economic logic dictates that naivete-based price discrimination is going on. This possibility is especially relevant since recent technological advances enable firms to collect and analyze more and more detailed information about consumers.

In this paper, we identify the welfare effects of an increase in firms' information about consumer naivete in a simple reduced-form model—designed to capture in one framework the markets for credit, banking, hotels, gambling, and mobile phones—in which competitive firms can take distortionary steps to introduce fees that naive consumers ignore. While in natural analogues with rational consumers (we show) seller information about consumers always increases social welfare, we identify arguably weak conditions under which seller information about consumer naivete *lowers* social welfare. The rough intuition is simple: information about who is naive leads firms to be more aggressive in exploiting naivete, and these steps generate large economic distortions. Furthermore, because naive consumers are hurt by the more aggressive distortionary exploitation and sophisticated consumers benefit less from the price cuts aimed at attracting naive consumers, in many cases information about naivete harms both types of consumers.

Section 2.1 presents the reduced-form model, and discusses the role of our key modeling as-

---

[1] Section 7 elaborates on the empirical claims and arguments in this paragraph.

sumptions. We assume that N competitive firms simultaneously set "anticipated prices" f_n and "additional prices" a_n to a population of consumers with unit demand, of whom a share α is naive and a share $1 - \alpha$ is sophisticated. Naive consumers ignore the additional prices when making purchase decisions, but nevertheless end up paying their chosen product's additional price. Sophisticated consumers, in contrast, understand and avoid the additional price. These assumptions mean that both naive and sophisticated consumers expect to pay only the anticipated price, and choose between products based on this expectation. To isolate the effect of seller information about naivete as distinct from seller information about preferences, we assume that naive and sophisticated consumers have the same distribution of valuations—and hence the same demand curve—for the product. Firm n's marginal cost is $c + k(a_n)$, where the function $k(\cdot)$ captures the distortionary cost of introducing additional prices. We define a competitive equilibrium as a price pair $f(\alpha), a(\alpha)$ such that $f(\alpha), a(\alpha)$ yields zero expected profits, and there are no alternative terms that would be preferred by consumers and earn positive expected profits.

We derive key properties of the competitive equilibrium in Section 2.2. Using these properties, in Section 2.3 we formally show that the above model is a reduced-form variant of—that is, generates the same equilibrium utilities as—a number of theories in which the nature of consumer naivete is explicitly specified and the cost $k(a_n)$ arises endogenously from how a firm attempts to exploit naivete. These applications include settings where (i) firms induce consumers to overborrow to increase the interest they can collect from naive consumers' unexpected willingness to delay loan repayment (e.g., credit and store cards); (ii) firms charge a high price for a service that naive consumers will use more often than they expect, and this high price induces consumers to undertake costly steps to avoid the service (e.g., hotels, bank accounts, mobile phones); (iii) firms exacerbate naive consumers' unexpected spending on a profitable product by oversupplying a complementary good (e.g., casinos); (iv) firms increase naive consumers' unexpected willingness to pay for an upgrade to a base product by distorting the quality of the base product downwards (e.g., computer programs, vacation packages); and (v) firms must pay intermediaries for selling products with additional prices to consumers (e.g., life insurance). In some of these applications, consumers rather than firms bear the exploitation cost—but the equilibrium welfare implications are the same

as in our reduced-form model because firms' savings is transferred to consumers through a lower anticipated price, exactly offsetting consumers' utility loss. And in some of these applications, consumers expect to pay some fees ex post—but these settings also simplify to our reduced-form model once we think of the anticipated price as the total price consumers expect to pay.

Our main interest is in analyzing the effects of naivete-based price discrimination on market outcomes and welfare. Formally, we compare social welfare when firms face a single pool of consumers in which the share of naive consumers is α_0 to that when firms sort consumers into two pools with shares $\alpha_1 > \alpha_0$ and $\alpha_2 < \alpha_0$ of naive consumers. As a benchmark, in Section 2.4 we consider variants of our model in which consumers are rational but heterogeneous in their profitability. We show that both when consumers differ in their marginal cost of service and when they differ in their willingness to pay, seller information either leaves outcomes unaffected, or increases social welfare by allowing firms to make more efficient pricing decisions.

To investigate the effects of naivete-based price discrimination in the most transparent way, we first analyze its effects on two types of inefficiency separately: the "participation distortion" that arises from consumers' socially suboptimal purchase decisions, and the "exploitation distortion" that arises from firms' inefficient expenditures on the additional price. Then, we consider how the interaction of the two distortions modifies our insights.

In Section 3, we isolate the participation distortion by assuming that the additional price is fixed at $\bar{a}$ and $k(\bar{a}) = 0$. In this case, because firms make ex-post profits from naive consumers, they set an anticipated price below marginal cost, so that some consumers with value below production cost end up buying the product. Seller information about consumer naivete leads firms to charge a lower anticipated price for consumers more likely to be naive—leading to the socially harmful entry of consumers into this pool—and to charge a higher anticipated price for consumers more likely to be sophisticated—leading to the socially beneficial exit of consumers from this pool. The net effect depends on the interaction of two forces. First, similarly to classical third-degree price discrimination, a consumer who exits has higher valuation than a consumer who enters, unambiguously lowering social welfare. Second, price discrimination can change the total number of consumers in the market, lowering welfare if more consumers enter than exit and raising welfare if

3

more consumers exit than enter. Under a simple condition on the demand curve that holds unless the curve is sufficiently concave, naivete-based price discrimination lowers social welfare. Furthermore, using a simple back-of-the-envelope calculation for the credit-card market, we illustrate that both the initial participation distortion and the effect of seller information can be massive.

In Section 4, we isolate the exploitation distortion by assuming that all consumers value the product above marginal cost. We show that under weak conditions on the cost function $k(\cdot)$, naivete-based price discrimination lowers social welfare. Intuitively, firms respond to their knowledge about consumer naivete by increasing the distortionary exploitation of consumers more likely to be naive and lowering the distortionary exploitation of consumers more likely to be sophisticated. Since an increase in a pre-existing distortion is more harmful than a similar decrease is beneficial, firms' response tends to lower overall welfare.[2] We also identify conditions under which *both* naive and sophisticated consumers are hurt by perfect seller information, so that—in contrast to classical first-degree price discrimination—first-degree naivete-based price discrimination can be strictly harmful in the Pareto sense.

In Section 5, we allow for both participation and exploitation distortions, and show that their interaction increases the adverse welfare impact of naivete-based price discrimination. On one side, consumers' participation decisions increase the exploitation distortion, as new consumers entering the market face a higher distortionary additional price than do consumers who exit the market. And on the other side, firms' endogenous decisions regarding the additional price increase the participation distortion by further lowering the anticipated price for consumers more likely to be naive—thereby drawing in more consumers with relatively low values for the product—and increasing the anticipated price for consumers more likely to be sophisticated—thereby inducing the exit of more consumers with relatively high values for the product.

Finally, in Section 6 we allow for the possibility that the welfare cost $k(a_n)$ of charging additional prices is different for naive and sophisticated consumers. For example, in the credit-card market naive (but not sophisticated) consumers pay unexpected fees and interest, and this can dis-

---

[2] Precisely, the condition for this to be the case is that $k'(a)/k''(a)$ is increasing in a, a property called (analogously to decreasing absolute risk aversion) "decreasing absolute convexity." Very roughly, decreasing absolute convexity says that the cost function is increasing less slowly than the exponential. This condition is satisfied for most cost functions used in economic analysis.

tort their consumption pattern conditional on participation, imposing an additional cost on them. Because naivete-based price discrimination increases the participation of more naive consumers and decreases the participation of more sophisticated consumers, it increases the share of naive consumers in the market. This "naivification effect" increases the share of consumers who suffer the additional welfare cost, adding another negative term to the welfare impact of naivete-based price discrimination.

In Section 7, we relate our paper to the empirical and theoretical literatures on markets with naive consumers and on price discrimination and privacy. While direct evidence that firms condition offers on information about naivete is limited, we argue based on theoretical and empirical considerations that this is or will soon be a major issue—yet to our knowledge no paper has addressed how it affects market outcomes and welfare. Also recognizing that firms may want to discriminate between consumers of different sophistication, some authors have began to analyze how firms can design contract menus to induce self-selection of consumers according to their naivete (Eliaz and Spiegler 2006, 2008, Heidhues and Kőszegi 2010, for example). It is clear, however, that firms do not rely solely on such screening to discriminate, and in some situations inducing self-selection according to naivete is impossible. For instance, if a naive and a sophisticated consumer have the same valuation for a credit card and both believe they will not incur any fees and interest, then they choose from contract menus in the same way.

In Section 8, we argue briefly that many of our insights would survive under imperfect competition, but also point out some important aspects of markets with naive consumers missing from our framework. Most crucially, the naivete-based price discrimination we analyze in this paper is likely to occur simultaneously with classical preference-based price discrimination as well as with naivete-based second-degree price discrimination, and the interaction of these considerations requires further research.

2   Model and Benchmarks

In this section, we introduce our model, which uses simple reduced-form assumptions to capture the distortionary cost of exploitation. We begin in Section 2.1 with presenting and discussing our

reduced-form model. In Section 2.2, we identify properties of the competitive equilibrium. In Section 2.3, we use these properties to formally establish that our reduced-form model generates the same welfare implications as fully specified models of a number of specific markets in which consumer naivete has been invoked as playing an important role. Finally, in Section 2.4 we show that in natural rational variants of our setup, seller information about consumers weakly increases social welfare.

2.1 Setup

We first present the formal model, and then discuss several notable aspects of it. $N \geq 2$ risk-neutral profit-maximizing firms selling a homogenous good or service simultaneously make offers consisting of an anticipated price $f_n \in \mathbb{R}$ and an additional price $a_n \in \mathbb{R}_+$ to a population of consumers. Each consumer is naive with probability α and sophisticated with probability $1 - \alpha$, and is interested in buying at most one unit of one product. A naive consumer does not take the additional prices into account when making purchase decisions, acting as if the total price of product n was f_n; but if she buys firm n's product, she ends up paying a_n as well. A sophisticated consumer anticipates the additional price, and takes costless steps to avoid paying it. Hence, if a consumer's value for the product is v, she makes purchase decisions under the expectation that her utility from product n will be $v - f_n$, but if she is naive, her actual utility will instead be $v - f_n - a_n$. We assume that both types have a distribution of valuations v that induces the demand curve $D(f) = \text{Prob}[v \geq f]$ for the product, where $D(\cdot)$ is strictly decreasing, twice continuously differentiable, and $D(c) > 0$. When indifferent, a consumer chooses a firm randomly in a way that is independent of whether she is naive. Firm n's cost of serving a customer is $c + k(a_n)$, where the "exploitation cost" function $k(a_n)$ satisfies $k(a) = 0$ for $a \leq \bar{a}$, $k'(\bar{a}) = 0$, $k''(a) > 0$ for $a \geq \bar{a}$, and $\lim_{a \to \infty} k'(a) > 1$.

Throughout, we analyze competitive markets, effectively imposing Bertrand competition in the offers (f_n, a_n). Since all consumers choose between contracts in the same way—based solely on the anticipated price—we can think of the competitive equilibrium as a single zero-profit contract that all consumers choose, and for which there is no profitable alternative contract that consumers would prefer:

Definition 1. Given the share α of naive consumers in a pool, a *competitive equilibrium* is a contract $(f(\alpha), a(\alpha))$ and share of participating consumers $D(f(\alpha))$ such that (i) $(f(\alpha), a(\alpha))$ earns zero profits; and (ii) there is no f', a' with $f' < f(\alpha)$ that makes positive expected profits.

Our main interest is in comparing competitive-equilibrium social welfare without and with price discrimination, where we define social welfare as the population-weighted sum of naive and sophisticated consumers' utilities. The prior probability of a consumer being naive is α_0, and without price discrimination firms base their offers only on this prior. With price discrimination, firms get—and can condition offers on—a binary signal that changes their beliefs about the probability of a consumer being naive from α_0 to either $\alpha_1 > \alpha_0$ or $\alpha_2 < \alpha_0$. We assume that firms receive the same signal. An alternative and for the above model equivalent formulation is that firms face a pool of consumers in which a share α_0 is naive, and without price discrimination they make the same offer to the entire pool, while with price discrimination they make different offers to two subpools with shares α_1 and α_2 of naive consumers.[3],[4]

Our model makes a number of assumptions intended to isolate the main question of this paper: how direct information firms acquire about consumer naivete affects welfare. First, by positing that both types of consumers have the same demand curve $D(f)$, we assume that the heterogeneity in the (ex-ante) willingness to pay for the product is independent of naivete, abstracting away from the question of how information about preferences affects welfare.

Second, by focusing on a pool of consumers who cannot be separated by means of self-selection at the initial stage (because they have the same beliefs and hence choose between terms in the same way), we abstract away from questions of how firms might screen consumers according to sophistication—questions that have been addressed by previous research (Eliaz and Spiegler 2006, 2008, Heidhues and Kőszegi 2010). As in previous research, if consumers have different beliefs at

---

[3] These two formulations have somewhat different practical interpretations. The first formulation is about whether firms acquiring aggregate information about naivete in a pool of consumers increases social welfare. The second is about whether firms offering different terms to more and less naive pools of consumers—or, equivalently, allowing firms to collect and use information about which pool a consumer belongs to—raises social welfare.

[4] Note that for price discrimination to take place, it is not necessary for any single firm to offer contracts to both pools of consumers. The outcomes we identify below also arise if some firms specialize in serving pool 1, and others specialize in serving pool 2. For instance, some mortgage lenders might operate in the market for more deceptive mortgages aimed at less sophisticated populations, and others might stick to more mainstream mortgage contracts aimed at more sophisticated populations.

the initial stage, often it is feasible to induce self-selection among them according to their beliefs. Whenever this is the case, our model can be viewed as applying to a subgroup of consumers with homogenous beliefs, once consumers have been screened according to beliefs. For instance, sophisticated consumers who cannot avoid the additional price will accept an offer with a higher anticipated price and lower additional price than that chosen by the consumers above, so these consumers can be thought of as being in a different market (Heidhues, Kőszegi and Murooka 2012b).

Third, we employ a structurally minimalistic approach by simply positing that firms can charge additional prices and that this results in a distortionary exploitation cost $k(\cdot)$. In Section 2.3, we show how these prices and costs can arise in different economic settings, but from the point of view of welfare their precise source is irrelevant, allowing us to analyze many applications in one framework. Specifically, for instance, while we assume that the exploitation cost $k(a)$ is borne by firms, as Section 2.3 establishes in the context of specific settings, the implications are identical if the cost is borne by consumers. The reason is simple: when the cost is borne by consumers instead of firms, the competitive-equilibrium anticipated price decreases by exactly the amount of the cost, offsetting consumers' utility loss. Similarly, while we assume that the additional price consumers expect to pay is zero, our model accommodates situations in which consumers expect to pay some add-on fees or other extra charges, but naive consumers underestimate how much. In these settings, we think of the anticipated price as the total amount consumers expect to pay ex ante.

It is, however, important to emphasize ways in which our reduced-form model is not sufficiently general to encapsulate all distortions from charging additional prices. Since we assume that sophisticated consumers avoid the additional price, we do not capture settings in which consumers must pay all fees equally. For instance, whether or not a consumer anticipates the management fee of an actively managed mutual fund, she must pay it if she invests in the fund. In addition, our assumption that the exploitation cost is the same for naive and sophisticated consumers implicitly presumes that the relevant distortions arise from the ex-ante considerations of firms and consumers. Since naive and sophisticated consumers have the same ex-ante beliefs and preferences, any distortion that is generated at this stage must apply equally to them. But distortions generated by

ex-post behavior—in particular, the choices that lead naive but not sophisticated consumers to pay an additional price—could be different for the two types of consumers. In Section 6, we consider one simple specification in which the exploitation costs are different across consumer types, but we do not analyze this possibility in generality.

2.2 Characterizing the Competitive Equilibrium

It is easy to see that the competitive-equilibrium contract must solve

$$\max_{(f,a)} \; \alpha(f + a) + (1 - \alpha)f - k(a) - c$$

$$\text{s.t.} \quad f \le f(\alpha).$$

If the competitive-equilibrium contract did not solve this problem, then the solution f^*, a^* would make positive profits, so that for a sufficiently small $\epsilon > 0$ the alternative contract $f' = f^* - \epsilon, a' = a^*$ would be a profitable deviation that violates condition (ii) of Definition 1.

Since a does not enter the constraint above, we must have $k'(a(\alpha)) = \alpha$. Using the zero-profit condition and the facts that sophisticated consumers pay only the anticipated price while naive consumers pay both prices, we get:

Proposition 1. *The competitive-equilibrium contract has $a(\alpha) = (k')^{-1}(\alpha)$ and $f(\alpha) = c + k(a(\alpha)) - \alpha a(\alpha)$, where $a(\alpha)$ is strictly increasing and $f(\alpha)$ is strictly decreasing. The utility of a sophisticated consumer who buys the product is*

$$U_s(\alpha) \equiv v - f(\alpha) = v - c + \alpha a(\alpha) - k(a(\alpha)), \tag{1}$$

while the utility of a naive consumer who buys the product is

$$U_n(\alpha) \equiv v - f(\alpha) - a(\alpha) = v - c - (1 - \alpha)a(\alpha) - k(a(\alpha)). \tag{2}$$

The conditions in Proposition 1 fully determine the competitive-equilibrium welfare implications of our model: any participating consumer receives the utility corresponding to her type above,

9

and since all consumers believe that their utility from participating equals $U_s(\alpha)$, a consumer participates if and only if $v - c + \alpha a(\alpha) - k(a(\alpha)) \geq 0$. In the next subsection, we show that various market models simplify to our reduced-form model by establishing that they have the equilibrium implications identified in Proposition 1.

2.3 Applications

In this section, we show how our reduced-form model captures a number of applications in which previous research has argued consumer naivete plays a role. Most of the existing research aims at understanding the terms firms offer when naive consumers are present, or emphasizes the cross-subsidies from naive to sophisticated consumers that the market generates. Our framework instead focuses on the welfare costs $k(\cdot)$ of exploiting naivete. Hence, we organize the applications according to the source of this cost. Since the criteria of competitive equilibrium—that the contract consumers accept earns zero profits, and that there is no profitable alternative contract that consumers would strictly prefer—are sufficiently straight-forward to apply in our settings, we do not define this concept formally for each new application.

2.3.1 Credit and Store Cards

In our first application, naive consumers with a taste for immediate gratification underestimate their future willingness to pay to put off repayment on a loan, firms exploit this mistake by imposing a high interest rate for delaying repayment, and to increase the interest payments they can collect they induce consumers to overborrow.[5] Formally, issuers of credit or store cards interact with consumers over three periods. In period 0, firms make offers consisting of a loan amount b, interest rate r, and discount d to consumers. The discount could, for instance, take the form of a price reduction if the store card is used or a perk associated with credit-card use. If a consumer takes the loan, in period 1 she chooses an amount $q \in [0, b]$ to repay in that period, leaving $(b - q)(1 + r)$ to be repaid in period 2. Let $\tilde{v}$ be the utility a consumer derives from interacting with an issuer

---

[5] Building on DellaVigna and Malmendier (2004) and Eliaz and Spiegler (2006), Heidhues and Kőszegi (2010) analyze credit contracts when some consumers have a naive taste for immediate gratification, and argue that the framework is consistent with real-life credit-card contracts. The model above is a simplified variant of that in Heidhues and Kőszegi (2010).

without receiving credit, for instance from having the convenience use of a card. Firms' cost of providing a card is c, and they acquire funds at zero interest.

Borrowers have time-inconsistent preferences derived from hyperbolic discounting with a long-run discount factor δ of 1. Self 0 has utility $\tilde{v} + u(b) - q - (b - q)(1 + r) + d$, where $u(\cdot)$ is the gross utility from funds. The assumption that self 0 does not discount the cost of repayment relative to the utility from consumption captures the notion that most of both occurs well after accepting a credit-card offer. We suppose that $u(\cdot)$ is differentiable, $u(0) = 0$, $u'(b) > 0$ and $u''(b) < 0$ for all $b \geq 0$, $\lim_{b \to \infty} u'(b) = 0$, and $u'(0) \geq 1$. In contrast to self 0, self 1 discounts payments in period 2 by a factor of β. For naive consumers, $\beta = \beta_n$, and for sophisticated consumers, $\beta = \beta_s$, where $1 \geq \beta_s > \beta_n > 0$. Both types believe in period 0 that $\beta = \beta_s$. We posit that self 0's utility is relevant for welfare considerations.[6] This means that the efficient level of borrowing, b^e, satisfies $u'(b^e) = 1$.

Since naive consumers delay repayment only if $\beta_n(1 + r) \leq 1$, the maximum amount of interest a firm can collect from naive consumers is $a = (1 - \beta_n)b/\beta_n$. Given this interest rate, sophisticated consumers repay the entire loan b in period 1, and since naive consumers believe that they behave the way sophisticated consumers do, they also think that they will repay early. Furthermore, since a competitive-equilibrium contract must break even, we must have $d = \alpha a - c$. Let $x = \beta_n/(1 - \beta_n)$ and note that $b = xa$. Using this, the equilibrium welfare of a sophisticated consumer is $\tilde{v} + u(b) - b + d = \tilde{v} - c + u(xa) - xa + \alpha a$. Since the competitive-equilibrium contract must maximize sophisticated consumers' utility subject to zero profits,[7] we must have $x - xu'(xa) = \alpha$,

---

[6] Equating welfare with long-run utility follows much of the literature on time inconsistency (Gruber and Kőszegi 2004, DellaVigna and Malmendier 2004, O'Donoghue and Rabin 2006, for instance). As we explain in Heidhues and Kőszegi (2010), although we simplify things by considering a three-period model, in reality time inconsistency seems to be mostly about very immediate gratification that plays out over many short periods. Hence, arguments by O'Donoghue and Rabin (2006) in favor of a long-run perspective apply: in deciding how to weight any particular week of a person's life relative to future weeks, it is reasonable to snub that single week's self—who prefers to greatly downweight the future—in favor of the many earlier selves—who prefer more equal weighting. This is especially so since the utility of consumption from borrowing on a credit card is distributed over many weeks. In addition, the models in Bernheim and Rangel (2004a, 2004b) can be interpreted as saying that a taste for immediate gratification is often a mistake not reflecting true welfare.

[7] If this was not the case, then there would be an alternative contract that makes zero profits and that sophisticated consumers strictly prefer. Starting from this alternative contract, a firm could slightly reduce the discount d it offers consumers and thereby create a profitable contract that consumers still strictly prefer, contradicting competitive equilibrium.

so that $xa \geq b^e$.

Now we define $k(a) = 0$ for $xa < b^e$ and $k(a) = (u(b^e) - b^e) - (u(xa) - xa)$ for $xa \geq b^e$, and $v = \tilde{v} + u(b^e) - b^e$. Then, the above analysis implies that in equilibrium $k'(a) = \alpha$, sophisticated consumers' utility from buying is

$$\tilde{v} - c + u(b) - b + \alpha a = \tilde{v} + u(b^e) - b^e + [(u(b) - b) - (u(b^e) - b^e)] + \alpha a$$
$$= v - c + [(u(xa) - xa) - (u(b^e) - b^e)] + \alpha a$$
$$= v - c - k(a) + \alpha a,$$

and naive consumers' utility from buying is a lower, exactly as in our reduced-form model.

2.3.2 Bank Accounts, Mobile Phones, and Hotels

In our next application, naive consumers underestimate their demand for an additional service, firms exploit this mistake by charging a high price for the service, and this induces consumers to undertake socially inefficient steps to avoid the service or refrain from buying it.[8] Formally, we suppose that firms sell a basic product—e.g., a bank account or mobile-phone contract—and an additional service—e.g., overdraft protection or roaming. We assume that the cost of the basic service is c and that of the additional service is zero, that firm n charges $\tilde{f}_n$ for the basic service and $\tilde{a}_n$ for the additional service, and that consumers observe all prices. Each consumer is interested in contracting with at most one firm, and can only purchase the additional service from the firm from which she bought the basic service. A sophisticated consumer needs the additional service with probability $\theta_s - t$, where θ_s is the baseline probability and t is the avoidance effort—such as arranging for sufficient funds in the banking example or buying a phone card in the mobile-phone example—undertaken by the consumer. Similarly, a naive consumer needs the additional service with probability $\theta_n - t$, where $\theta_n > \theta_s$. Both naive and sophisticated consumers initially believe

---

[8] This mechanism is explicitly discussed by Gabaix and Laibson (2006) in the context of hotels and by Armstrong and Vickers (2012) in the context of bank accounts. One important difference is that in our setting naive and sophisticated have the same initial beliefs regarding their demand for the service, so they both undertake costly avoidance. In addition, Grubb (2009) analyzes mobile-phone contracts under the assumption that consumers are overconfident in predicting how much they will use their phone, and shows that the optimal contract features a high price for high usage. While Grubb does not discuss this implication, the high price can create a welfare loss if it leads consumers to reduce their demand.

that they will need the additional service with the baseline probability θ_s. The cost of avoidance is $\kappa(t)$, where $\kappa(0) = 0$, $\kappa'(t), \kappa''(t) > 0$ for all $t > 0$, $\kappa'(\theta_s) = \infty$, and $\kappa'(t)/\kappa''(t)$ is increasing in t.[9] A consumer's ex-ante valuation for the product (consisting of her valuation for the basic service as well as the additional service if needed) is v.

If a consumer selects the contract $\tilde{f}, \tilde{a}$, her avoidance effort satisfies $\kappa'(t) = \tilde{a}$. Denoting the solution to this equation by $t^*(\tilde{a})$, we transform the model into our reduced-form setup by defining the anticipated price as the overall payment of a sophisticated consumer, the additional price as the unanticipated additional payment by a naive consumer, and the exploitation cost as the inefficient consumer avoidance costs: $f = \tilde{f} + (\theta_s - t^*(\tilde{a}))\tilde{a}$, $a = (\theta_n - \theta_s)\tilde{a}$, and $k(a) = \kappa(t^*(a/(\theta_n - \theta_s))$. Because a competitive-equilibrium contract $(f(\alpha), a(\alpha))$ earns zero profits, we have $f(\alpha) + \alpha a(\alpha) = c$. Combining these considerations, the utilities of sophisticated and naive consumers are exactly the same as in our reduced-form model. Furthermore, since a competitive-equilibrium contract must maximize a sophisticated consumer's utility subject to zero profits,[10] we must also have $k'(a(\alpha)) = \alpha$.

2.3.3 Casinos

In this application, firms can exacerbate naive consumers' unexpected spending on a profitable product by supplying a complementary good, inducing them to overspend on the complementary good.[11] We assume that a casino imposes an entrance fee $\tilde{f}$ and chooses a subsidy level b for extra services—such as free drinks—that increase the amount of impulsive gambling consumers undertake.

---

[9] The second-to-last condition ensures that complete avoidance does not occur in equilibrium, so that the first-order condition describes consumers' avoidance decisions. The last condition ensures that the avoidance cost chosen by consumers, and hence the reduced-form cost of exploitation, is strictly convex in the unanticipated extra payment of naive consumers. This holds for example for any convex cost function $\kappa(\tilde{a}) = \phi\tilde{a}^\gamma$, where $\phi > 0$ and $\gamma > 1$. Furthermore, solving for the equilibrium avoidance costs in this case, one can show that $k'(a)/k''(a) = (\gamma - 1)a$, which is increasing in a. This condition will play a central role in our analysis below.

[10] This follows from an analogous argument to that in Footnote 7.

[11] Consistent with our assumption that some consumers underestimate how much they will gamble, Andrade and Iyer (2008, forthcoming) find in an experiment that subjects gamble more than planned after losses. More broadly, although we are not aware of a systematic investigation of naivete, a large literature studies the determinants and consequences of compulsive and pathological gambling. In a model of casino gambling based on prospect theory, Barberis (2012) also proposes that some consumers underestimate when they will quit. The assumption that it is possible to exacerbate impulsive gambling is similar in spirit to the assumption made by Bernheim and Rangel (2004a) in the context of drug addiction. They assume that consuming a drug increases the probability that the consumer takes the drug in the future irrespective of the consequences.

The analysis below makes clear that in equilibrium $\tilde{f}$ is negative; this can be implemented through subsidizing services, such as hotel rooms or dining, that are not complementary to gambling. The utility of a sophisticated consumer is $\tilde{v} + u(b) - l(b) - \tilde{f}$, where $u(b)$ is the direct utility from the service and $l(b)$ is the consumer's loss (and casino's profit) from the induced impulse gambling. We assume that $l'(b) > 0$, $l''(b) < 0$, and $lim_{b \to \infty} l'(b) = 0$. While a naive consumer believes that she will have the same utility as a sophisticated consumer, her actual utility is $\tilde{v} + u(b) - l(b) - a(b) - \tilde{f}$, where $a(b)$ is the unexpected additional loss from impulse gambling she incurs. We assume that $a'(b) > 0$, $a''(b) < 0$, and $lim_{b \to \infty} a'(b) = 0$.[12] Providing an amount b of the service to a consumer costs the casino κb. Note that the efficient level of service, b^e, therefore satisfies $u'(b^e) = \kappa$.

Because a competitive-equilibrium contract $(\tilde{f}, b)$ earns zero profits, we have $l(b) + \tilde{f} + \alpha a(b) = \kappa b$. Furthermore, as it must be impossible to introduce a profitable contract that consumers prefer, a competitive-equilibrium contract must solve

$$\max_{(\tilde{f}', b')} \quad l(b') + \tilde{f}' + \alpha a(b') - \kappa b'$$

$$\text{s.t.} \quad \tilde{f}' - u(b') + l(b') \leq \tilde{f} - u(b) + l(b).$$

Using that the constraint must hold with equality, we have

$$\alpha = \frac{\kappa - u'(b)}{a'(b)}. \tag{3}$$

Hence, if all consumers are sophisticated, firms provide the efficient level of service ($b = b^e$). As the share of naive consumers increases, however, firms provide excessive service in order to increase the unanticipated gambling losses of naive consumers.

To transform this application into our reduced-form model, we define the anticipated price as the total payment of sophisticated consumers and the additional price as the unanticipated gambling losses of naive consumers: $f = \tilde{f} + l(b)$ and $a = a(b)$. Furthermore, we define the exploitation cost as the welfare loss from the wrong service level. Specifically, since firms can collect the additional price $a(b^e)$ from naive consumers without distorting service provision, $k(a) = 0$ for any $a \leq a(b^e)$.

---

[12] In our model, the losses from impulse gambling are transfers that by themselves have no impact on social welfare. In reality, collecting these revenues could impose an inefficiency as well. For example, since collecting these transfers requires charging high prices for gambling, some consumers may restrict their gambling below efficient, generating a welfare loss similar to that in the previous application.

Denoting the inverse of $a(b)$—i.e., the service necessary to collect an additional price of a—by $b(a)$, for $a \geq a(b^e)$ the exploitation cost is $k(a) = \int_{b^e}^{b(a)}[\kappa - u'(b)]db$. This implies that

$$k'(a) = \frac{\kappa - u'(b)}{a'(b)} = \alpha.$$

Furthermore, defining the gross utility of a consumer as $v = \tilde{v} + u(b^e)$ and the firm's cost as $c = \kappa b^e$, it is easy to check that equilibrium utilities are the same as in our reduced-form model.

2.3.4 Computer Games, Software, and Vacation Packages

Another setting consistent with our reduced-form model is one in which naive consumers underestimate their future willingness to pay for an upgrade to a base product (e.g., better options for a computer program, or additional entertainment and dining for a pre-booked trip), and to increase naive consumers' willingness to pay firms distort the quality of the base product downwards (e.g., by making the baseline program or trip unnecessarily inconvenient or unexciting).

We describe our model in the context of computer games; the same formalism covers the other applications with minor modifications. There are two periods, $t = 1, 2$. Firms choose initial and upgraded quality levels q_1 and q_2, a base fee f, and an upgrade price a. A consumer can buy a game at the beginning of period 1, and can choose whether to upgrade at the beginning of period 2. Sophisticated consumers' utility level is $v_1(q_1)$ in period 1, and—having gotten bored of the game—0 in period 2. Naive consumers also get utility $v_1(q_1)$ in period 1 and expect to be bored by period 2, but to their surprise they find themselves interested in continuing to play. Specifically, naive consumers get utility $v_2(q)$ from playing a game of quality q in period 2.[13] We first assume that the value generated in period 2 does not affect the consumer's true welfare—from a long-run perspective, she does not value continued game playing—and discuss the case in which it does below. We suppose that both value functions are strictly concave, there are satiation levels q_1^e, q_2^e

---

[13] In a slightly different specification, q_1 corresponds to the quality of the basic service level (e.g., included amenities at a hotel), and q_2 corresponds to the service level in an add-on service (e.g., kitesurfing and other activities) that neither type of consumer expects to care about ex ante. In this case, we can specify $v_1(q_1)$ as the utility from the basic service, and $v_2(q_1, q_2)$ as naive consumers' surprise value for the add-on service. So long as q_1 and q_2 are substitutes in v_2, similar arguments to the ones below show that this alternative specification also simplifies to our reduced-form model.

satisfying $v'_t(q^e_t) = 0$ for $t = 1, 2$, and $v'_1(q)/v'_2(q) < 1$ is strictly decreasing for $q \leq q^e_1$. The marginal cost of selling the game to a new customer is c, and the marginal cost of quality is zero.

In period 2, a firm that offers the quality levels $q_1, q_2 \geq q_1$ will charge the additional price $a = v_2(q_2) - v_2(q_1)$. We argue that the competitive-equilibrium quality levels must solve

$$\max_{q_1, q_2} \quad v_1(q_1) + \alpha \underbrace{[v_2(q_2) - v_2(q_1)]}_{=a}. \tag{4}$$

Indeed, suppose toward a contradiction that the quality levels in the competitive equilibrium f, q_1, q_2 do not solve the above problem, and denote a solution by q^*_1, q^*_2. Then, for a sufficiently small $\epsilon > 0$ the offer $f' = f + (v_1(q^*_1) - v_1(q_1)) - \epsilon, q^*_1, q^*_2$ would both be preferred by consumers and earn positive profits, a contradiction.

The problem in (4) yields the first-order conditions

$$\frac{v'_1(q_1)}{v'_2(q_1)} = \alpha \quad \text{and} \quad v'_2(q_2) = 0.$$

The firm thus sets $q_2 = q^e_2$. If there are no naive consumers, the firm chooses $q_1 = q^e_1$, but the larger the share of naive consumers, the lower the quality of the basic package.

To see that this application also simplifies to our reduced-form model, we define the exploitation cost as the distortion in the initial quality level given how much the firm wants to charge for the upgrade: $k(a) = 0$ for $a \leq v_2(q^e_2) - v_2(q^e_1)$, and $k(a) = v_1(q^e_1) - v_1(v_2^{-1}(v_2(q^e_2) - a))$ for $a > v_2(q^e_2) - v_2(q^e_1)$.[14] This implies that in equilibrium $k'(a) = v'_1(q_1)/v'_2(q_2) = \alpha$. Defining $v = v_1(q^e_1)$ and using that the zero-profit condition implies $f = c - \alpha a$, the equilibrium utility levels are exactly as in our reduced-form model.

We conclude our analysis of this application by discussing the possibility that game playing in period 2 affects a naive consumer's welfare. Since this does not modify the firm's problem or consumers' beliefs in period 1, it leaves the equilibrium offer unchanged, and therefore it simply shifts naive consumers' utility by the constant term $x = v_2(q^e_2)$. Depending on the nature of consumers' utility, x could be positive or negative. If game playing in period 1 generates an unanticipated harmful addiction in period 2 that—consistent with the model of Becker and Murphy

---

[14] Note that for $a > v_2(q^e_2) - v_2(q^e_1)$, $v_2^{-1}(v_2(q^e_2) - a)$ is the initial quality level for which the firm can charge an upgrade price of a.

16

(1988)—lowers utility below that of the outside option despite increasing the marginal utility of consumption, then x is negative. In contrast, if game playing in period 2 generates unanticipated genuine enjoyment, then x is positive.[15] We discuss how these alternatives modify our results in Section 6.

2.3.5 Other Possibilities

We discuss two more possible sources of exploitation cost without a formal model.

First, additional expenditures a may be generated by hiding fees or penalties in complicated contracts, where for example the interest rate on future payments depends on the consumer's behavior in other markets or—as in a recent ad by a major German bank—on the outcome of a public lottery.[16] To hide more fees and generate higher unanticipated revenues from naive consumers, a firm has to write longer and more involved contracts. Consumers dislike reading and thinking about such contracts, generating disutility $k(a)$. Naive consumers, despite trying to understand the contract, nevertheless fail to find all traps and to their surprise end up paying an additional price a.

Second, it may be the case that firms have to pay a commission to intermediaries or spend on persuasive advertising to try to induce consumers to purchase a product with an additional price. It seems natural to assume that—due for example to the intermediary's expected future cost from undermining her reputation or violating her professional code of ethics—these marketing costs increase in the additional price.[17] Hence, we can define the cost of marketing as the welfare loss $k(a)$ from exploitation.

---

[15] In the latter case, however, firms have an incentive to disclose or demonstrate to naive consumers how much they will enjoy the product. This disclosure allows firms to sell a more expensive product to naive consumers without having to distort initial quality.

[16] The Postbank, a subsidiary of Deutsche Bank, recently heavily advertised an account with an up to 2.75% interest rate in Germany. The fine print, however, revealed that the guaranteed interest rate is only 0.75% but can be as high as 1.15%, plus an extra bonus of up to 1.0% that, however, is only paid for half a year, and a maximal "victory bonus *(Gewinnbonus)*" of 0.6% that, however, is only paid if numbers ending on the digits 81-99 are drawn in a public lottery—the so called Aktion Mensch—, the probability of which is is 32 times lower than winning the lottery in which one has to guess all six numbers correctly. The confusing details can be found in the following press article: http://www.spiegel.de/wirtschaft/service/postbank-gewinnsparen-verspricht-bis-zu-2-75-prozent-zinsen-a-882677.html .

[17] This assumption is similar in spirit to that made by Inderst and Ottaviani (2012).

2.4 Rational Benchmarks

To highlight the distinct predictions of our model, we briefly describe the effects of seller information about consumers in analogous classical environments with rational consumers. We argue that in all relevant cases, either the first-best obtains with or without information about consumers or information about consumers strictly increases social welfare.

Our framework allows for firms to receive perfect information about who is naive, a scenario we interpret as naivete-based first-degree price discrimination.[18] The review by Stole (2007) demonstrates that in a classical setting, first-degree (perfect) price discrimination maximizes social welfare for any given industry structure. As a particularly close classical analogue of our setup, consider a situation in which competitive firms offer a base product, and each firm can then sell an add-on to its consumers fully excluding competitors. First-degree price discrimination in the add-on market—whereby firms can price each unit at the consumer's value, and know the consumer's demand curve when pricing the base product—allows firms to extract the full social surplus from the interaction ex post, but competition leads them to hand this back to consumers ex ante, maximizing social welfare and leaving all of it in consumers' hands.

More generally, however, in our model firms acquire only imperfect information about consumer naivete, and we interpret this as naivete-based third-degree price discrimination. As possible analogues of our setup, we consider classical reasons that consumers might differ in their profitability: differences in their marginal cost and differences in their demand (i.e., their willingness to pay). Specifically, because it leaves the margin unchanged, the assumption that naive consumers generate ex-post profits of a may seem equivalent to the assumption that they are cheaper to serve by a. Similarly, the assumption that naive consumers pay a higher price than sophisticated consumers by the amount a may seem equivalent to the assumption that their demand curve is higher than

---

[18] When perfect information about naivete is coupled with perfect information about consumers' value for the product—as is possible in the setting of Section 4—firms know everything about consumers, creating a situation that clearly corresponds to first-degree price discrimination. More generally, Stole (2007) defines first-degree price discrimination as a situation in which each firm extracts the full surplus under its residual demand curve, or, equivalently in a classical setting, obtains the maximally achievable profits given voluntary participation by consumers and competitors' offers. By this definition, any situation in which firms receive perfect information about consumer naivete—which allows them to maximize profits for both types of consumers given competing offers—corresponds to first-degree price discrimination, even if firms do not know individual consumers' values for the product.

that of sophisticated consumers by the same amount.

We begin with analyzing the effect of information when the marginal cost of serving consumers is uncertain, as for instance when ex-ante identical consumers demand different amounts of free customer service from a firm ex post. Suppose that the cost of serving a consumer is c_L with probability α and $c_H > c_L$ with probability $1 - \alpha$, and the demand curve of both types is $D(\cdot)$. Consumers observe the total price set by a firm, and we define the competitive-equilibrium price as the (total) price at which firms make expected profits of zero. Analogously to our informational question above, we compare welfare when firms believe that the probability of a consumer being cheap is α_0 to that when they get a signal that changes their beliefs to $\alpha_1 > \alpha_0$ or $\alpha_2 < \alpha_0$.

Proposition 2 (Information about Costs). *For any $\alpha_0, \alpha_1, \alpha_2$, information about the cost of serving consumers raises social welfare.*

Intuitively, because firms must price the product according to the average cost in the population, cheap consumers underparticipate in the market and expensive consumers overparticipate. Information regarding the cost of serving a consumer leads firms to price closer on average to the marginal cost of serving that consumer, lowering the average distortion.

We now turn to the effect of information when there is uncertainty regarding consumers' demand. Suppose that the inverse market demand curve is $P_L(Q)$ with probability $1 - \alpha$ and $P_H(Q)$ with probability α, where Q is the total quantity produced. We assume $P_H(Q) > P_L(Q)$ for all Q, and that both curves are differentiable and strictly decreasing.

In this case, it matters whether firms set prices or quantities. If firms set prices, then under competition seller information has no effect: whatever sellers know, marginal-cost pricing and efficiency obtains. To complete the picture, we consider the possibility that firms set quantities, as for instance when hotel chains have to decide on the quantity or quality of hotel rooms before demand conditions are fully known. We define the competitive-equilibrium quantity level as the level for which firms make expected profits of zero. Again, seller information of the same type as above is always beneficial:

Proposition 3 (Information about Preferences Under Quantity Setting). *For any $\alpha_0, \alpha_1, \alpha_2$, information about the inverse demand curve increases social welfare.*

The intuition is similar to that in the case of cost heterogeneity. Information about the inverse demand curve leads competitive firms to choose quantities that are on average closer to the efficient level for the realized demand, thereby lowering the average distortion.

3 Participation Distortions

We now turn to our main question, analyzing the effect of information about consumer naivete. Our model generates two kinds of distortions from deception: distortions from socially suboptimal participation decisions—that consumers respond to the wrong prices when making purchase decisions—and distortions from the exploitation itself—that firms offer inefficient terms that help them exploit consumers. To highlight the origin and consequences of these two forms of inefficiency, in this section and the next we consider special cases of our model that isolate them, beginning with participation distortions. Section 5 analyzes the additional effects from the interaction of the two distortions.

3.1 Analysis

To study participation distortions, we assume that the additional price is fixed at $\bar{a} > 0$, and $k(\bar{a}) = 0$. This outcome obtains in the limiting case of our general model in which $k(a) = 0$ for $a \leq \bar{a}$, and $k(a) = \kappa \cdot (a - \bar{a})$ for $a > \bar{a}$, where $\kappa > 1$.

Since in equilibrium firms earn an expected additional price of $\alpha\bar{a}$ from the share α of naive consumers, the zero-profit condition implies that the competitive-equilibrium anticipated price is $f(\alpha) = c - \alpha\bar{a}$, leading to demand $D(c - \alpha\bar{a})$. Because the anticipated price is below marginal cost, it induces overparticipation by consumers. The welfare loss from this overparticipation equals

$$DWL(\alpha) = \int_{c-\alpha\bar{a}}^{c} [D(c - \alpha\bar{a}) - D(f)]df.$$

As we have explained above, we ask how the ability of firms to sort a pool of consumers with a share α_0 of naive types into two pools with shares $\alpha_1 > \alpha_0$ and $\alpha_2 < \alpha_0$ of naive types affects social welfare. Since the size of the two pools, p_1 and p_2, must satisfy $p_1\alpha_1 + p_2\alpha_2 = \alpha_0$, the welfare effect of information hinges on the shape of $DWL(\alpha)$. More precisely, a small amount of information

$(\alpha_1, \alpha_2 \approx \alpha_0)$ lowers social welfare if and only if $DWL(\alpha)$ is convex at α_0, and any information lowers social welfare if $DWL(\alpha)$ is convex on $[0, 1]$. Hence, we look for conditions under which $DWL(\alpha)$ is convex. The derivative equals

$$DWL'(\alpha) = -\bar{a} \int_{c-\alpha\bar{a}}^{c} D'(c - \alpha\bar{a})df = -\bar{a}^2 \alpha D'(c - \alpha\bar{a}). \tag{5}$$

The second derivative is

$$DWL''(\alpha) = \underbrace{-\bar{a}^2 D'(c - \alpha\bar{a})}_{\text{misallocation effect}} + \underbrace{\bar{a}^3 \alpha D''(c - \alpha\bar{a})}_{\text{output effect}}. \tag{6}$$

As reflected in Equation 5, information about consumer naivete leads to a lower anticipated price and hence more purchases in the more naive pool 1, lowering social welfare, and to a higher anticipated price and hence fewer purchases in the more sophisticated pool 2, increasing social welfare. The net impact can be decomposed into two effects that also arise in classical third-degree price discrimination. The first term in Equation 6 captures the misallocation of products across the two pools of consumer. A consumer who exits the market in pool 2 values the product more highly than a consumer who enters in pool 1, unambiguously lowering welfare. The second term, in turn, captures the effect of the total number of consumers in the market. In particular, if the demand curve is not linear, the number of consumers who exit is typically not the same as the number of consumers who enter. An increase in the total number of consumers lowers social welfare, while a decrease increases social welfare.

The above analysis implies a simple condition for when any information about consumer naivete lowers social welfare:

Proposition 4 (Participation Distortions)**.** *If* $-D'(f) + (c - f)D''(f) > 0$ *for all* f *satisfying* $c - \bar{a} \leq f \leq c$, *then for any* $\alpha_0, \alpha_1, \alpha_2$, *information about consumer naivete lowers social welfare.*

Although it is ultimately an empirical question, the condition in Proposition 4 appears relatively weak. In particular, it holds unless the demand function is sufficiently concave, including for the constant elasticity demand functions often used in applied work and for linear demand. In addition, while we do not believe this case applies in many markets we consider (for instance, it very much

21

does not apply in our calibration below), the condition holds if the additional price is small, so that:

Corollary 1. *For any $D(\cdot)$, there is an A such that if $\bar{a} < A$, then for any $\alpha_0, \alpha_1, \alpha_2$, information about consumer naivete lowers social welfare.*

As a complement to our analysis of total welfare, we briefly state conditions under which naivete-based price discrimination lowers not only total social welfare, but also the average utility of both types of consumers. As a notable extreme case for contrasting the implications of information about consumer naivete to the implications of information about consumer preferences in a classical setting, we consider only perfect seller information. As we explain in more detail in Section 7, in a classical setting perfect price discrimination always maximizes social welfare, so it cannot lower the utility for all groups of market participants.

Perfect information raises the anticipated price at which the product is available to sophisticated consumers, always making this group worse off. In contrast, perfect information lowers the anticipated price at which the product is available to naive consumers from $c - \alpha_0 \bar{a}$ to $c - \bar{a}$, benefiting those who would buy at either price but—since the total price consumers pay is c—hurting those attracted to the market by the lower price. Perfect price discrimination lowers the average utility of naive consumers if and only if the increase in the participation distortion exceeds the benefit from the price cut:

$$(D(c - \bar{a}) - D(c - \alpha_0 \bar{a}))\alpha_0 a + \int_{c-\bar{a}}^{c-\alpha_0 \bar{a}} (D(c - \bar{a}) - D(f))df > D(c - \alpha_0 \bar{a})(1 - \alpha_0)\bar{a}.$$

3.2   A Calibration Exercise

To illustrate the potential importance of our results, in this section we provide back-of-the-envelope calculations on the welfare losses predicted by our model in the credit-card market. While a full-blown quantitative analysis is outside the scope of this paper, our calculations suggest that the participation distortion and the welfare effect of seller information can both be massive.

For our illustration, we assume that the demand curve $D(\cdot)$ is linear. Then, the participation

distortion equals

$$\frac{1}{2}\alpha\bar{a}\cdot(D(c-\alpha\bar{a})-D(c)) = \frac{1}{2}(-D'(c-\alpha\bar{a}))(\alpha\bar{a})^2 = \frac{1}{2}\underbrace{\left(\frac{-D'(c-\alpha\bar{a})c}{D(c-\alpha\bar{a})}\right)}_{\text{elasticity}}\underbrace{\left(\frac{\alpha\bar{a}}{c}\right)^2}_{\substack{\text{deception} \\ \text{ratio}}}\underbrace{cD(c-\alpha\bar{a})}_{\substack{\text{market-wide} \\ \text{revenue}}}. \quad (7)$$

Two inputs into the participation distortion are standard market measures that are (at least in principle) observable, the price elasticity of demand and the size of the market as measured by firm revenues. The only caveat is that in calculating the elasticity, the price responsiveness of demand must be measured at the anticipated price $(c-\alpha\bar{a})$—the price that consumers are basing purchase decisions on—which is different from the average price actually paid by consumers (c). In addition, the welfare cost depends on what we call the deception ratio, the ratio of the expected additional price $(\alpha\bar{a})$ to the marginal cost (c). Using the fact that in a competitive industry $\alpha\bar{a} = c - f$, the deception ratio equals $(c-f)/c$, so it depends on the price consumers expect to pay for the product.

As an example, we consider the consumer side of the US credit-card industry. For simplicity, we suppose that the interchange fee issuers receive from merchants is set equal to the network externality from additional consumer purchases.[19] This means that in a competitive credit-card market, the marginal revenue from a consumer must equal the marginal social cost of issuance. Hence, in calculating the participation distortion on the consumer side of the market, our model can be applied to that side in isolation.

In 2009, the total revenue of credit-card issuers was about 104 billion, and of this $31 billion came from interchange fees, suggesting that the size of the market—i.e., card issuers' revenue from consumers—was roughly $73 billion. Using the facts that there were approximately 115.2 million US households in the period 2008-2012 and 68% of US families owned a credit card in 2010, and assuming that these figures did not change drastically from year to year, we estimate that there were about 78.3 million US households that held a credit card in 2009.[20] Combining these numbers,

---

[19] The literature on two-sided markets makes it clear that this is not necessarily the case, and a fuller analysis of the credit-card industry would need to account for the difference. For an introduction to this literature, see Rochet and Tirole (2006) and Rysman (2009).

[20] The facts cited above come, in order of appearance, from the following sources: IBIS World Credit Card Report (http://www.ibisworld.com/industry/default.aspx?indid=1293 accessed January 27, 2014), Wang (2012), the US

issuers earned roughly \$932 per household from credit-card consumers. Assuming a competitive industry, therefore, the marginal (social) cost c of serving a household was also \$932.

We calculate the participation distortion under several assumptions for the price f that households expect to pay for a credit card, assuming that the elasticity of demand with respect to the anticipated price is 1. Consistent with the fact that many or most credit cards have no annual fees and the impression that many consumers seem to believe they get credit-card services for free, as a first guess we suppose that $f = 0$. Then, the welfare loss is half of the market size, or \$36.5 billion, or \$466 per credit-card-owning household (or \$317 per US household). If, instead, consumers believe not only that credit cards are free, but that credit cards offer valuable perks, the up-front fee may be even lower. For example, if $f = -\$100$, the participation distortion increases to roughly 63% of the market size, or \$587 per card-owning household. In contrast, if consumers expect to pay fees over and above the perks they anticipate benefiting from, the welfare costs are lower. If consumers anticipate net yearly payments of \$300 when signing up for a credit card—a number that strikes us as implausibly high—then the total participation distortion still amounts to roughly 21% of the market size, or \$196 per household.

By comparison, the Office of Fair Trading (OFT) in the UK estimates that over the period 2010-2013, *all* of its activities combined lead to direct consumer benefits of 16 pounds (\$25) per household per year, and total (direct and indirect) benefits of 78-227 pounds (\$122-355) per household per year.[21] Furthermore, since the policies likely lower firms' profits, these numbers could significantly overstate the social benefit of OFT activities. While far from conclusive, these estimates make clear that the participation distortion from the deception of naive consumers in the credit-card market can be enormous—and this is just a single market and just one of the distortions from deception we study in this paper.

Census (`http://quickfacts.census.gov/qfd/states/00000.html`), and Bricker, Kennickell, Moore and Sabelhaus (2012).

[21] The OFT Impact report (`http://www.oft.gov.uk/shared_oft/reports/Evaluating-OFTs-work/oft1493.pdf`), estimates a total direct impact of 422 million pounds per year. While the OFT acknowledges that indirect consumer benefits—such as those through deterrence—of its activities are hard to estimate, it suggests that the "deterrence effect of [the] competition enforcement work (abuse of dominance, cartels, and other anti-competitive agreements) could be between 12 and 40 times the direct effect" of these activities, which is estimated to be 136 million pounds. The per-household figures use the fact that there are about 26.4 million households in the UK (see the Office for National Statistics (`http://www.ons.gov.uk/ons/rel/family-demography/families-and-households/2013/stb-families.html`).

Our calibration also allows us to think about the possible change in total welfare when sellers obtain information about consumer naivete. Equation (7) implies that starting from a situation in which the share of naive consumers is α_0, perfect information about consumer naivete multiplies the participation distortion by $1/\alpha_0$.[22] Bricker et al. (2012) document that 55% of US families carry a balance (i.e., pay interest) on their credit cards. Supposing that most do not anticipate carrying a balance, one may guess that a share $\alpha_0 = 1/2$ of consumers is naive. In that case, perfect information about consumer naivete would double the above welfare loss, again a potentially huge economic effect.

4   Exploitation Distortions

In this section, we analyze the effect of seller information on exploitation distortions—the distortions arising from the costs $k(a_n)$ firms pay for introducing additional prices. To isolate these distortions, we assume that all consumers value the product at $v > c$. We begin with the effect of seller information on social welfare, and then consider the effects on the individual consumer types.

4.1   Total Social Welfare

Given that all consumers have $v > c$ and participate, the welfare loss in a competitive equilibrium (relative to first-best) is simply $DWL(\alpha) = k(a(\alpha))$. As in the case of participation distortions, the welfare effect of information about consumer naivete depends on the shape of $DWL(\alpha)$, with the effect being negative if this function is convex. Hence, we look for conditions under which $DWL(\alpha)$ is convex. The first derivative equals

$$DWL'(\alpha) = k'(a(\alpha))a'(\alpha) = \frac{k'(a(\alpha))}{k''(a(\alpha))} > 0,$$

where the second equality follows from totally differentiating the equilibrium condition $k'(a(\alpha)) = \alpha$. Hence,

$$DWL''(\alpha) = \frac{d}{da}\left[\frac{k'(a(\alpha))}{k''(a(\alpha))}\right]\frac{1}{k''(a(\alpha))}, \tag{8}$$

---

[22] The welfare loss is $(1/2)\alpha^2\bar{a}^2 D'(c - \alpha\bar{a})$. Hence, perfect information increases the welfare loss of affected consumers by a factor $1/\alpha_0^2$. Given that the share of affected consumers is α_0, overall the loss in social welfare increases by a factor of $1/\alpha_0$.

and we immediately obtain:

Proposition 5 (Social Welfare). *If $k'(a)/k''(a)$ is increasing in a over the interval $[a(0), a(1)]$, then for any $\alpha_0, \alpha_1, \alpha_2$, any seller information about consumer naivete lowers social welfare.*

In response to information, firms increase the additional price to the pool with more naive consumers—thereby increasing the exploitation distortion and lowering social welfare—and lower the additional price to the pool with more sophisticated consumers—thereby lowering the exploitation distortion and raising social welfare. Because an increase in a pre-existing distortion is more costly than an identical decrease is beneficial, the net effect is often negative. Qualifying this tendency is that the effect of information may have asymmetric effects on the additional prices chosen for the two pools, which could mitigate or exacerbate the adverse effect of seller information. In particular, if the additional price decreases for the sophisticated pool sufficiently more than it increases for the naive pool, then the welfare effect of information is positive. This would, however, require that the marginal exploitation cost increases much faster for increases in a than it decreases for decreases in a. Hence, under conditions on the cost function $k(\cdot)$ that seem weak, seller information lowers social welfare. The condition in Proposition 5 is satisfied for most functional forms used in the literature for cost functions, including for instance all convex power functions.[23]

As an example in the context of one of our applications, consider the credit-card market discussed in Section 2.3.1. In that setting,

$$\frac{k'(a)}{k''(a)} = \frac{x(1 - u'(xa))}{-x^2 u''(xa)},$$

where $u(\cdot)$ is the borrower's utility function over money. The condition that $k'(a)/k''(a)$ increases in a is then *weaker* than that the individual is prudent (i.e., that $u'''(\cdot) \geq 0$), which in turn is weaker than non-increasing absolute risk aversion—both of which are generally considered to be weak conditions on a utility function.[24]

---

[23] For power cost functions $k(a) = ka^r$, strict convexity requires $r > 1$. Hence, $k'(a)/k''(a) = a/(r-1)$ is strictly increasing in a.

[24] The derivative of $[x(1-u'(xa))]/[-x^2 u''(xa)]$ with respect to a is positive if $[-x^2 u''(xa)]^2 + x^3 u'''(xa)[1-u'(xa)] \geq 0$, and hence a sufficient—although not necessary—condition is that $u'''(b) \geq 0$ for all $b \geq 0$. Non-increasing absolute risk aversion, in turn, requires that $-u'''(b)u'(b) + [u''(b)]^2 \leq 0$, for which $u'''(b) \geq 0$ is a necessary, but not sufficient, condition.

But while the condition in Proposition 5 seems weak, constructing examples in which information increases welfare is easy. As a limiting extreme case, suppose that the cost function has a capacity constraint at $a(\alpha_0) > 0$ (i.e., $k(a) = \infty$ for any $a > a(\alpha_0)$), and $\bar{a} = 0$. Then, perfect information leaves the additional price faced by naive consumers unchanged at $a(\alpha_0)$, while it reduces the additional price faced by sophisticated consumers to zero, raising total welfare.

4.2 Welfare of Individual Consumer Types

We now identify conditions under which perfect information about naivete lowers the welfare of both naive and sophisticated consumers. As in the case of participation distortions above, perfect naivete-based price discrimination serves as a notable extreme case for comparison to a classical setting, where perfect price discrimination always maximizes social welfare.

Our first result concerns sophisticated consumers:

Proposition 6 (Sophisticated Consumers). *For any $\alpha_0 \in (0, 1)$ and any $k(\cdot)$, perfect information about consumer naivete hurts sophisticated consumers.*

Intuitively, in a competitive equilibrium the ex-post profits firms make from naive consumers are redistributed to all consumers ex ante, so that—similarly to the logic of Gabaix and Laibson (2006) and the literature following it—without price discrimination naive consumers in effect cross-subsidize sophisticated ones. Since perfect price discrimination eliminates the cross-subsidy, it hurts sophisticated consumers.

In contrast to the impact on sophisticated consumers, the impact of perfect price discrimination on naive consumers is in general ambiguous. Proposition 7 below identifies some special cases in which the effect is negative or positive. To state Part I, we define the function $k_+(a') = k(\bar{a} + a')$; this is the part of the cost function above $\bar{a}$.

Proposition 7 (Naive Consumers).

I. For any $k_+(\cdot)$, if $\bar{a}$ and α_0 are sufficiently small, then perfect information about naivete hurts naive consumers.

II. If $k''(a(\alpha))a(\alpha) < 1$ for all α, then for any $\alpha_0 \in (0, 1)$ perfect information about naivete hurts naive consumers.

III. If there is an $\underline{\alpha}$ such that $k''(a(\alpha))a(\alpha) > 1$ for all $\alpha > \underline{\alpha}$, then for any $\alpha_0 \in (\underline{\alpha}, 1)$ perfect information about naivete benefits naive consumers.

The above results can be understood in terms of two effects on naive consumers' welfare. On the one hand, since perfect information (as we have explained above) eliminates the cross-subsidy from naive to sophisticated consumers, it benefits naive consumers. On the other hand, perfect information leads firms to increase the additional price, and since this exploitation is costly, it hurts naive consumers.

Part I of the proposition says that if there are few naive consumers and $\overline{a}$ is small, then perfect information hurts naive consumers. In this case, with no information naive consumers benefit from firms' incentive to serve the sophisticated consumers in an efficient way: firms choose a low additional price, which means that both the cross-subsidy effect and the exploitation effect above are small. But when firms know which consumers are naive, they exploit these naive consumers aggressively, so that in this case the exploitation effect is relatively large.

Parts II and III of the proposition identify conditions under which perfect information hurts versus benefits naive consumers when their share is not small. Since the cross-subsidy effect benefits while the exploitation effect hurts naive consumers, naive consumers are hurt by perfect information if the latter effect dominates. The exploitation effect, in turn, is large if and only if the additional price is very responsive to the share of naive consumers. This responsiveness is given by $a'(\alpha) = 1/k''(a(\alpha))$, and the condition $k''(a(\alpha))a(\alpha) < 1$ says that the responsiveness is sufficiently large for the exploitation effect to outweigh the cross-subsidy effect. If this condition is violated, on the other hand, then the cross-subsidy effect dominates the exploitation effect, and naive consumers benefit from perfect information.

5 Interaction Between Participation and Exploitation Distortions

We now consider the interaction between exploitation distortions and participation distortions in our general model, and show that this introduces additional effects that add to the welfare loss from seller information about naivete.

In our general model, the welfare loss relative to first-best consists of two components. First, firms pay an inefficient exploitation cost of $k(a(\alpha))$ for all consumers in the market. Second, consumers with values between $f(\alpha)$ and c purchase, even though this is socially suboptimal. Hence, the total welfare loss equals

$$DWL(\alpha) = D(f(\alpha))k(a(\alpha)) + \int_{f(\alpha)}^{c} [D(f(\alpha)) - D(f)]df.$$

Using that $f'(\alpha) = -a(\alpha)$ and $a'(\alpha) = 1/k''(a(\alpha))$—both of which follow from $k'(a(\alpha)) = \alpha$—the first derivative equals

$$DWL'(\alpha) = D(f(\alpha))k'(a(\alpha))a'(\alpha) - D'(f(\alpha))a(\alpha)k(a(\alpha)) - \int_{f(\alpha)}^{c} D'(f(\alpha))a(\alpha)df$$

$$= D(f(\alpha))\frac{k'(a(\alpha))}{k''(a(\alpha))} - \alpha a(\alpha)^2 D'(f(\alpha)).$$

The second derivative is then

$$DWL''(\alpha) = \underbrace{D(f(\alpha))\frac{\mathrm{d}}{\mathrm{d}a}\frac{k'(a(\alpha))}{k''(a(\alpha))} \cdot \frac{1}{k''(a(\alpha))}}_{\text{exploitation distortion}} \underbrace{- a(\alpha)^2 D'(f(\alpha)) + \alpha a(\alpha)^3 D''(f(\alpha))}_{\text{participation distortion}}$$

$$\underbrace{-3a(\alpha)D'(f(\alpha))\frac{k'(a(\alpha))}{k''(a(\alpha))}}_{\text{interaction}}. \tag{9}$$

The welfare effect of information is composed of three terms, the first two of which are the general versions of the effects we have analyzed in the previous two sections. First, holding constant the set of consumers in the market, information regarding consumer naivete increases the distortionary exploitation of consumers more likely to be naive and decreases the distortionary exploitation of consumers more likely to be sophisticated. The net effect is negative if $k'(a)/k''(a)$ is increasing in a. Second, holding constant the additional price, information leads to the entry of consumers more likely to be naive and to the exit of consumers more likely to be sophisticated. This generates two effects we have discussed under participation distortions, and lowers welfare if $-D'(f(\alpha)) + \alpha a(\alpha)D''(f(\alpha))$ is positive for all α. The third effect is novel, and comes from the two-way interaction of participation and exploitation considerations. On one side, consumers' participation decisions increase the exploitative distortions, as new consumers who enter the market

29

face a higher distortionary additional price than do consumers who exit the market. And on the other side, firms' exploitation decisions increase the participation distortion because firms lower the anticipated price for consumers more likely to be naive—thereby drawing in further consumers with low values for the product—and increase the anticipated price for consumers more likely to be sophisticated—thereby inducing the exit of further consumers with relatively high values for the product. Both of these interaction effects exacerbate the adverse welfare effect of information.

6 The Naivification Effect

To identify an additional effect of naivete-based price discrimination, in this section we briefly consider the implications of assuming that the exploitation cost is different for naive and sophisticated consumers. We capture this possibility in a simple way: we suppose that the welfare cost for sophisticated consumers is $k(a)$, and the welfare cost for naive consumers is $k(a)+x$, where x is a constant welfare adjustment that is unanticipated by and borne by naive consumers. The exploitation cost could be higher for naive consumers ($x > 0$) if, in a scenario that is likely for instance in the credit-card market, having to pay an unexpected additional price distorts consumers' consumption choices given participation. And the exploitation cost could be lower for naive consumers ($x < 0$) if, as is plausible for instance in the case of mail-in rebates, naive consumers expect to but in the end do not take costly steps to avoid the additional price. Although x is likely to depend on a in most situations, to make our point most clearly we assume that it is constant.

Since x is borne by naive consumers and is unexpected, it does not affect firms' considerations, so it leaves the market equilibrium unchanged. Hence, the above modification merely changes the dead-weight loss in the market by $\alpha D(f(\alpha))x$. This implies that the second derivative in Equation (9) becomes

$$DWL''(\alpha) = \underbrace{D(f(\alpha))\frac{\mathrm{d}}{\mathrm{d}a}\frac{k'(a(\alpha))}{k''(a(\alpha))} \cdot \frac{1}{k''(a(\alpha))}}_{\text{exploitation distortion}} \underbrace{- a(\alpha)^2 D'(f(\alpha)) - a(\alpha)^2(\alpha a(\alpha) + \alpha x)D''(f(\alpha))}_{\text{participation distortion}}$$

$$\underbrace{-3a(\alpha)D'(f(\alpha))\frac{k'(a(\alpha))}{k''(a(\alpha))}}_{\text{interaction}} \underbrace{- D'(f(\alpha))[2a(\alpha) + \alpha a'(\alpha)]x}_{\text{naivification effect}}. \tag{10}$$

30

A difference in exploitation costs changes the implications of price discrimination in one main way. To understand this additional effect, notice that price discrimination—by increasing the size of the more naive pool and decreasing the size of the more sophisticated pool—increases the share of naive consumers in the population. This "naivification effect" impacts the welfare effect of price discrimination negatively if exploitation imposes a higher loss on naive consumers ($x > 0$) and positively if exploitation imposes a lower loss on naive consumers ($x < 0$). In addition, a difference in exploitation costs modifies the participation distortion in Equation (10). Recall that price discrimination can increase or decrease the total number of consumers in the population, and this is respectively harmful or beneficial since the marginal consumer values the product $\alpha a(\alpha)$ below c. With a further welfare loss of x on a naive consumer, any change in the number of consumers lowers social welfare by an additional αx.

7   Related Literature

7.1   Empirical Background

One of the key assumptions of our model is that naive consumers incur unexpected charges. This assumption is made in different forms in many papers in behavioral industrial organization, and is consistent with empirical findings from a number of industries. For instance, Stango and Zinman (2009) find that consumers incur many avoidable fees, and the Office of Fair Trading (2008) reports that most consumers who use overdraft protection do so unexpectedly. Evidence by Agarwal, Driscoll, Gabaix and Laibson (2008) indicates that many credit-card consumers seem to not know or forget about various fees issuers impose. Ausubel (1999) documents that consumers receiving credit-card solicitations overrespond to the introductory ("teaser") interest rate relative to the post-introductory rate, suggesting that they end up borrowing more than they intended or expected. Woodward and Hall (2012) find that borrowers underestimate broker compensation, and Gerardi, Goette and Meier (2009) document that 26% of borrowers who face a prepayment penalty are unaware of it. Evidence by Wilcox (2003), Barber, Odean and Zheng (2005), and Anagol and Kim (2012) indicates that investors underweight operating expenses when choosing mutual funds. Hall

(1997) reports that 97% of buyers do not know the price of a cartridge when buying their printer, and as revealed in a survey by UK's Office of Fair Trading, retailers believe that 75% of consumers do not have an idea about printing costs. Finally, regulators are worried about the "bill shock" many mobile-phone consumers face when they run up charges they did not anticipate (Federal Communication Commission 2010).

The other central assumption in our model is that firms acquire and use information about consumer naivete for designing offers. Some direct evidence supports this assumption. Gurun et al. (2013) document that lenders targeted less sophisticated populations with ads for expensive mortgages. Schoar and Ru (2014) find that the offers credit-card companies sent to less educated borrowers feature more back-loaded payments, including low introductory interest rates but high late fees, penalty interest rates and overdraft limit fees. Beyond the direct evidence, there are economic reasons to believe that price discrimination based on information about naivete is or will soon be going on much more broadly. First, since consumers who incur unexpected charges tend to be more profitable yet often cannot be screened through self-selection (because they have the same beliefs as a sophisticated consumer), firms have a strong incentive to obtain outside information about consumers' naivete. Second, researchers have documented several simple correlates of the tendency to make financial mistakes,[25] and while it is unlikely that firms can access and use exactly the same measures of consumer naivete in designing offers, it is also unlikely that they cannot use any measures. This is especially so given recent technological advances in collecting and processing information about consumers. As a simple example, the complexity of the words a person uses in an email message may well be correlated with naivete, and Google allows firms to condition offers on this measure. Given that information about naivete is highly profitable and seems to be obtainable based on simple and intuitive measures, it is likely that firms are or will soon be able to use such information in targeting offers. Indeed, researchers such as Bar-Gill and Warren (2008, pages 23-25) take it for granted that firms are already doing so.

---

[25] For instance, Agarwal, Driscoll, Gabaix and Laibson (2007) find an age pattern in the amount of financial mistakes individuals make, Calvet, Campbell and Sodini (2007) report that consumers with lower levels of education or income make more investing mistakes, and Stango and Zinman (2011) document that it is possible to predict, based on two simple hypothetical questions on the Survey of Consumer Finances, the consumers who buy the most overpriced loans.

7.2 Related Theory

In asking how outside information about consumers affects economic outcomes, our paper is related to the literature on first- and third-degree price discrimination, as well as to the literature on privacy. To our knowledge, however, no paper has considered the question we address in this paper: how information about whether a consumer is likely to be naive affects welfare. The existing literature overwhelmingly assumes that the consumer type about which firms may acquire information concerns preferences, not naivete. As we argue in Section 2.4, seller information about consumer preferences increases social welfare if the information firms acquire or the competition firms face is perfect. In contrast, our main results indicate that information (including perfect information) about naivete often lowers social welfare under perfect competition.

We now discuss other work based on the classical framework. Building on a large literature, Aguirre, Cowan and Vickers (2010) analyze monopolistic third-degree price discrimination and establish how the overall welfare effect depends on the interplay between the misallocation effect first introduced by Pigou (1920) and the output effect originally discussed by Robinson (1933).[26] While both effects are also central for understanding the participation distortion in our model, a monopolist facing classical consumers supplies an inefficiently low amount of the good, so that—in contrast to our model— the output effect is positive whenever the monopolist chooses to produce a larger quantity.

The interpretation of our model in which firms get information about the overall share of naive consumers in a given pool is closely related to models in which firms learn about aggregate demand. Vives (2001, Chapter 8.3) considers how firms' acquiring more information about consumer demand affects oligopolistic distortions under different forms of competition. Since quantity is lower than efficient, forces that increase production when demand is high—i.e., when consumers value the product most—increase efficiency. Under Cournot competition, information that demand is high leads firms to increase production, so that in this case information raises expected welfare. Under Bertrand competition, in contrast, information that demand is high leads firms to raise prices and

---

[26] Stole (2007) highlights that the same basic logic determines the welfare effects in a homogenous-good Cournot model, while additional effects are relevant in differentiated price-competition models.

thereby shrink the market, so that in this case information lowers welfare.

In a similar vein to the literature on price discrimination, the literature on privacy often finds that it is socially beneficial for firms to know more about consumers or employees. Stigler (1980) argues that the protection of personal information leads others to substitute other, less efficient, forms of information acquisition or screening, and Posner (1981) contends that privacy protection creates asymmetric information that impedes the functioning of markets. Varian (1996) reasons that it is in both a consumer's and a firm's best interest to know *which* product the consumer would like—this lowers search costs for the consumer—although the consumer would not like the firm to know *how much* she likes the product. Hoffmann, Inderst and Ottaviani (2013) consider a model in which each product has two dimensions, and firms—having obtained information about the consumer's preferences—reveal a consumer's utility in the better of her dimensions. Although such "selective disclosure" is biased, it still provides useful information to consumers, and hence raises social welfare unless a number of market frictions are present at the same time.[27,28]

In considering how firms respond to the presence of naive consumers, our paper also belongs to the growing literature on behavioral industrial organization.[29] Research in behavioral industrial organization considers two distinct kinds of naive consumers, those who understand contracts but mispredict their own future behavior, and those who misunderstand contract terms or product features.[30] Our reduced-form model can capture both forms of naivete. A common theme in

---

[27] Taylor (2004) and Calzolari and Pavan (2006) consider dynamic pricing games in which a consumer makes two purchases in sequence from two monopolists, and her first purchase reveals information about her willingness to pay for the second product. The authors ask whether disclosing information about the first purchase to the second monopolist benefits sellers or consumers. Unlike in our paper, the focus is on how the potential for disclosure affects the first interaction, not on how any information revealed affects welfare in the second interaction. Similarly, Acquisti and Varian (2005) consider a single monopolist selling two products in a row, and analyze optimal pricing strategies when consumers can use anonymizing technologies in the first purchase. Taylor (2004) and Acquisti and Varian (2005) also point out that if consumers do not anticipate how firms will use information on early purchases, firms take advantage of this naivete, to the detriment of consumers and possibly social welfare.

[28] Hermalin and Katz (2006) take an entirely different perspective on the privacy debate. They show that releasing information about an individual before contracting can create a kind of reclassification risk that risk-averse individuals dislike. Hence, it is optimal to forbid the release of private information, even if consumers have no taste for privacy per se. Hermalin and Katz's argument applies best to settings where insurance is a major consideration, which does not seem to be the case for most of the applications we consider.

[29] See Spiegler (2011) for an introduction to and overview of this literature.

[30] For models incorporating the former type of naivete, see for DellaVigna and Malmendier (2004) Eliaz and Spiegler (2006), Eliaz and Spiegler (2008), Grubb (2009), Heidhues and Kőszegi (2010), and Laibson and Yariv (2007). For models using the latter type, see among others Armstrong and Vickers (2012), Gabaix and Laibson (2006), Heidhues et al. (2012b), Heidhues, Kőszegi and Murooka (2012a), Piccione and Spiegler (2010), and Spiegler (2006).

much of the literature, first emphasized by Gabaix and Laibson (2006), is that in competitive environments the profits firms make from exploiting naive consumers are handed back to consumers ex ante, leading to a cross-subsidy from the more profitable naive to the less profitable sophisticated consumers.

In contrast to our paper, the existing behavioral industrial organization literature takes firms' information regarding consumer naivete as given, and asks how firms respond to this knowledge. Several authors, nevertheless, have asked how firms may screen consumers with different degrees of naivete (Eliaz and Spiegler 2006, Eliaz and Spiegler 2008, Heidhues and Kőszegi 2010). Most of this research focuses on identifying the contract forms firms use in the presence of naive consumers, and understanding whether and how a consumer's outcomes and welfare depend on her level of naivete. In contrast to this "second-degree naivete-based price discrimination" that relies on self-selection between consumers who have different beliefs when contracting, we consider another important aspect of the problem, where firms use outside information about consumers for price discrimination between consumers who have identical beliefs when contracting.

8   Conclusion

The main message of our paper is that—at least from the perspective of social welfare—the economic consequences of firms acquiring information about preferences and firms acquiring information about naivete are qualitatively different. One important implication of this message is that empirical research on the nature of information acquired and used by firms is crucial to determine the effects of first- and third-degree price discrimination.

Our analysis also calls for further theoretical research on a number of economically important realities in markets with naive consumers that our framework ignores. As an immediate example, all of our analysis focuses on a competitive market. While we have not solved a fully specified general model with imperfect competition, our basic qualitative points are likely to survive in such a setting. Since the level of competition affects a consumer's outside option, it changes the anticipated price a firm can charge, but does not affect the optimal additional price. As a result, imperfect competition affects the participation distortion, but not the exploitation distortion. Furthermore,

while the intensity of competition impacts the *level* of the participation distortion, it does not seem to impact the conclusion that the effect of seller information about naivete is often negative. As in the case of naivete-based price discrimination under perfect competition as well as classical third-degree price discrimination under imperfect competition, naivete-based price discrimination under imperfect competition distorts the allocation of products across the two pools of consumers, and this welfare-decreasing effect can be mitigated or exacerbated by the welfare impact of a change in the total number of consumers in the market.

As we have emphasized, we designed our framework to isolate the question of how seller information about naivete affects welfare. Sellers, however, have other concerns related to consumer heterogeneity as well. In most settings, for instance, the information sellers can observe about consumers pertains not just to naivete, but also to preferences. In addition, it is likely to be optimal for a seller not just to rely on outside information to discriminate between consumers, but also to take advantage of self-selection according to preferences or naivete. How third-degree naivete-based price discrimination interacts with other forms of discrimination is an important topic for future research.

Finally, our paper does not analyze potential policy responses to price discrimination. For instance, a commonly advocated solution to privacy concerns is to require firms to obtain a consumer's consent before using her private information. In on-going work, we investigate whether this policy helps in our framework, and find that it does not: because a naive consumer does not understand that the firm will use information to exploit her, she agrees to giving her information too easily. In addition, a regime with required consent may transfer even more money from naive to sophisticated consumers than does a regime without required consent.

References

Acquisti, Alessandro and Hal R. Varian, "Conditioning Prices on Purchase History," *Marketing Science*, 2005, *24* (3), 367–381.

Agarwal, Sumit, John C. Driscoll, Xavier Gabaix, and David Laibson, "The Age of Reason: Financial Decisions Over the Lifecycle," 2007. MIT Department of Economics Working Paper No. 07-11.

___ , ___ , ___ , and ___ , "Learning in the Credit Card Market," 2008. NBER Working Paper #13822.

Aguirre, Iñaki, Simon Cowan, and John Vickers, "Monopoly Price Discrimination and Demand Curvature," *American Economic Review*, September 2010, *100* (4), 1601–1615.

Anagol, Santosh and Hoikwang Kim, "The Impact of Shrouded Fees: Evidence from a Natural Experiment in the Indian Mutual Funds Market," *American Economic Review*, 2012, *102* (1), 576–593.

Andrade, Eduardo B. and Ganesh Iyer, "Planned Versus Actual Betting in Sequential Gambles," *Journal of Marketing Research*, 2008, forthcoming.

Armstrong, Mark and John Vickers, "Consumer Protection and Contingent Charges," *Journal of Economic Literature*, 2012, *50* (2), 477–493.

Ausubel, Lawrence M., "Adverse Selection in the Credit Card Market," 1999. Working Paper, University of Maryland.

Bar-Gill, Oren and Elizabeth Warren, "Making Credit Safer," *University of Pennsylvania Law Review*, 2008, *157* (1), 1–101.

Barber, Brad M., Terrance Odean, and Lu Zheng, "Out of Sight, Out of Mind: The Effects of Expenses on Mutual Fund Flows," *Journal of Business*, 2005, *78* (6), 2095–2119.

Barberis, Nicholas, "A Model of Casino Gambling," *Management Science*, 2012, *58* (1), 35–51.

Becker, Gary S. and Kevin M. Murphy, "A Theory of Rational Addiction," *Journal of Political Economy*, 1988, *96* (4), 675–700.

Bernheim, B. Douglas and Antonio Rangel, "Addiction and Cue-Triggered Decision Processes," *American Economic Review*, 2004, *94* (5), 1558–1590.

___ **and** ___ , "Saving and Cue-Triggered Decision Processes," 2004. Mimeo, Stanford University.

Bricker, Jesse, Arthur B. Kennickell, Kevin B. Moore, and John Sabelhaus, "Changes in U.S. Family Finances from 2007 to 2010: Evidence from the Survey of Consumer Finances," *Federal Reserve Bulletin*, February 2012, *98* (2), 1–80.

Calvet, Laurent E., John Y. Campbell, and Paolo Sodini, "Down or Out: Assessing the Welfare Costs of Household Investment Mistakes," *Journal of Political Economy*, 2007, *115* (5), 707–747.

Calzolari, Giacomo and Alessandro Pavan, "On the Optimality of Privacy in Sequential Contracting," *Journal of Economic Theory*, 2006, *130* (1), 168–204.

DellaVigna, Stefano and Ulrike Malmendier, "Contract Design and Self-Control: Theory and Evidence," *Quarterly Journal of Economics*, 2004, *119* (2), 353–402.

Eliaz, Kfir and Ran Spiegler, "Contracting with Diversely Naive Agents," *Review of Economic Studies*, 2006, *73* (3), 689–714.

___ **and** ___ , "Consumer Optimism and Price Discrimination," *Theoretical Economics*, 2008, *3* (4), 459–497.

Federal Communication Commission, "Comment Sought on Measures Designed to Assist U.S. Wireless Consumers to Avoid "Bill Shock"," Technical Report 2010. Public Notice May 11 2010. CG Docket No. 09-158.

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

Gerardi, Kris, Lorenz Goette, and Stephan Meier, "Financial Literacy and Mortgage Outcomes," 2009. Paper presented at the ASSA meetings.

Grubb, Michael D., "Selling to Overconfident Consumers," *American Economic Review*, 2009, *99* (5), 1770–1805.

Gruber, Jonathan and Botond Kőszegi, "Tax Incidence when Individuals are Time-Inconsistent: The Case of Cigarette Excise Taxes," *Journal of Public Economics*, 2004, *88* (9-10), 1959–1987.

Gurun, Umit G., Gregor Matvos, and Amit Seru, "Advertising Expensive Mortgages," 2013. NBER Working Paper #18910.

Hall, Robert E., "The Inkjet Aftermarket: An Economic Analysis," 1997. Unpublished Manuscript, Stanford University.

Heidhues, Paul and Botond Kőszegi, "Exploiting Naivete about Self-Control in the Credit Market," *American Economic Review*, 2010, *100* (5), 2279–2303.

___ , ___ , **and Takeshi Murooka**, "Exploitative Innovation," 2012. Working Paper in Preparation.

___ , ___ , **and** ___ , "Inferior Products and Profitable Deception," 2012. Working Paper.

Hermalin, Benjamin and Michael Katz, "Privacy, property rights and efficiency: The economics of privacy as secrecy," *Quantitative Marketing and Economics*, 2006, *4* (3), 209–239.

Hoffmann, Florian, Roman Inderst, and Marco Ottaviani, "Hypertargeting, Limited Attention, and Privacy: Implications for Marketing and Campaigning," 2013. Working Paper.

Inderst, Roman and Marco Ottaviani, "Financial Advice," *Journal of Economic Literature*, 2012, *50* (2), 494–512.

Laibson, David I. and Leeat Yariv, "Safety in Markets: An Impossibility Theorem for Dutch Books," 2007. Working Paper, Caltech.

O'Donoghue, Ted and Matthew Rabin, "Optimal Sin Taxes," *Journal of Public Economics*, 2006, *90* (10-11), 1825–1849.

Office of Fair Trading, "Personal Current Accounts in the UK," Technical Report 2008.

Piccione, Michele and Ran Spiegler, "Price Competition under Limited Comparability," 2010. Working paper, LSE and UCL.

Pigou, Arthur C., *The Economics of Welfare*, London, Macmillian, 1920.

Posner, Richard A., "The Economics of Privacy," *American Economic Review*, 1981, *71* (2), 405–409.

Robinson, Joan, *The Economics of Imperfect Competition*, London, Macmillian, 1933.

Rochet, Jean-Charles and Jean Tirole, "Two-sided markets: a progress report," *RAND Journal of Economics*, 2006, *37* (3), 645–667.

Rysman, Mark, "The Economics of Two-Sided Markets," *The Journal of Economic Perspectives*, 2009, *23* (3), 125–143.

Schoar, Antoinette and Hong Ru, "Do Credit Card Companies Screen for Behavioral Biases?," 2014. Working Paper.

Spiegler, Ran, "Competition over Agents with Boundedly Rational Expectations," *Theoretical Economics*, 2006, *1* (2), 207–231.

_____ , *Bounded Rationality and Industrial Organization*, Oxford University Press, 2011.

Stango, Victor and Jonathan Zinman, "What Do Consumers Really Pay on Their Checking and Credit Card Accounts? Explicit, Implicit, and Avoidable Costs," *American Economic Review Papers and Proceedings*, 2009, *99* (2), 424–429.

_____ **and** _____ , "Fuzzy Math, Disclosure Regulation, and Market Outcomes: Evidence from Truth-in-Lending Reform," *Review of Financial Studies*, 2011, *24* (2), 506–534.

Stigler, George J., "An Introduction to Privacy in Economics and Politics," *Journal of Legal Studies*, 1980, *9* (4), 623–644.

Stole, Lars A., "Price Discrimination and Competition," in Mark Armstrong and Robert Porter, eds., *Handbook of Industrial Organization*, Vol. 3 2007, chapter 34, pp. 2221–2299.

Taylor, Curtis R., "Consumer Privacy and the Market for Customer Information," *RAND Journal of Economics*, 2004, *35* (4), 631–650.

Varian, Hal R., "Economic Aspects of Personal Privacy," 1996. Working Paper, UC Berkeley.

Vives, Xavier, *Oligopoly Pricing: Old Ideas and New Tools*, MIT Press, 2001.

Wang, Zhu, "Debit Card Interchange Fee Regulation: Some Assessments and Considerations," *Economic Quarterly*, Third Quarter 2012, *98* (3), 159–183.

Wilcox, Ronald T., "Bargain Hunting or Star Gazing? Investors' Preferences for Stock Mutual Funds," *The Journal of Business*, 2003, *76* (4), 645–663.

Woodward, Susan E. and Robert E. Hall, "Diagnosing Consumer Confusion and Sub-Optimal Shopping Effort: Theory and Mortgage-Market Evidence," *American Economic Review*, 2012, *100* (2), 511–515.

A   Proofs

Proof of Proposition 1. We have already argued in the text that $k'(a(\alpha)) = \alpha$ for any contract that is accepted with positive probability in equilibrium. Because $k', k'' > 0$, therefore, $a(\alpha)$ is increasing. The zero-profit condition implies that the anticipated price $f(\alpha) = c + k(a(\alpha)) - \alpha a(\alpha)$. Furthermore, the equilibrium contract must maximize $\alpha a(\alpha) - k(a(\alpha))$; when α increases this expression increases when holding a fixed, and thus must also increase when $a(\alpha)$ is re-optimized following an increase in α. Thus, the anticipated price $f(\alpha)$ decreases in α. $\square$

Proof of Proposition 2. It is socially efficient for a consumer to obtain the good if and only if her willingness to pay is above the cost of serving her, i.e. above c_H for high-cost and above c_L for low-cost consumers. Both types of consumers, however, consume the good whenever their willingness to pay is above the average cost of serving consumers, i.e. above $\alpha c_L + (1 - \alpha)c_H$. For any α, the deadweight loss

$$
\begin{aligned}
DWL(\alpha) = \ &(1 - \alpha) \int_{\alpha c_L + (1-\alpha)c_H}^{c_H} \left[D(\alpha c_L + (1 - \alpha)c_H) - D(c) \right] dc \\
&+ \ \alpha \int_{c_L}^{\alpha c_L + (1-\alpha)c_H} \left[D(c) - D(\alpha c_L + (1 - \alpha)c_H) \right] dc.
\end{aligned}
$$

Hence,

$$
\begin{aligned}
DWL'(\alpha) = &-\int_{\alpha c_L+(1-\alpha)c_H}^{c_H} \left[D(\alpha c_L + (1-\alpha)c_H) - D(c)\right] dc \\
&+ (1-\alpha)\int_{\alpha c_L+(1-\alpha)c_H}^{c_H} D'(\alpha c_L + (1-\alpha)c_H)(c_L - c_H)dc \\
&+ \int_{c_L}^{\alpha c_L+(1-\alpha)c_H} \left[D(c) - D(\alpha c_L + (1-\alpha)c_H)\right] dc \\
&+ \alpha \int_{c_L}^{\alpha c_L+(1-\alpha)c_H} D'(\alpha c_L + (1-\alpha)c_H)(c_H - c_L)dc \\
= &\int_{c_L}^{c_H} \left[D(c) - D(\alpha c_L + (1-\alpha)c_H)\right] dc.
\end{aligned}
$$

Therefore,

$$
\begin{aligned}
DWL''(\alpha) &= \int_{c_L}^{c_H} D'(\alpha c_L + (1-\alpha)c_H)(c_H - c_L)dc \\
&= D'(\alpha c_L + (1-\alpha)c_H)(c_H - c_L)^2 < 0.
\end{aligned}
$$

Since the deadweight loss is strictly concave in the share of naive consumers, and for any imperfect targeting technology $\alpha = p_1\alpha_1 + p_2\alpha_2$ with $\alpha_1 \neq \alpha_2$, $DWL(\alpha) > p_1 DWL(\alpha_1) + p_2 DWL(\alpha_2)$. $\square$

Proof of Proposition 3. Denote by $Q(\alpha)$ the competitive-equilibrium quantity when the probability of high demand is α. By definition, we have $\alpha P_H(Q(\alpha)) + (1-\alpha)P_L(Q(\alpha)) = c$. Note that $Q(\cdot)$ is strictly increasing and differentiable.

The welfare loss relative to first-best equals

$$
DWL(\alpha) = (1-\alpha)\int_{P_L^{-1}(c)}^{Q(\alpha)} (c - P_L(q))dq + \alpha\int_{Q(\alpha)}^{P_H^{-1}(c)} (P_H(q) - c)dq.
$$

Hence,

$$
\begin{aligned}
DWL'(\alpha) = &-\int_{P_L^{-1}(c)}^{Q(\alpha)} (c - P_L(q))dq + \int_{Q(\alpha)}^{P_H^{-1}(c)} (P_H(q) - c)dq \\
&+ Q'(\alpha)[(1-\alpha)(c - P_L(Q(\alpha))) - \alpha(P_H(Q(\alpha)) - c)].
\end{aligned}
$$

By definition of $Q(\alpha)$, the second line above equals zero. Using this,

$$
DWL''(\alpha) = -Q'(\alpha)[(c - P_L(Q(\alpha))) + P_H(Q(\alpha)) - c] = -Q'(\alpha)[P_H(Q(\alpha)) - P_L(Q(\alpha))] < 0.
$$

41

Since the deadweight loss is strictly concave in α, any information about α lowers the expected dead-weight loss; i.e., increases social welfare. $\square$

Proof of Proposition 4. The proof is contained in text. $\square$

Proof of Proposition 5. The proof is contained in text. $\square$

Proof of Proposition 6. Suppose otherwise. Then there exists an equilibrium contract offer f', a' for a pool in which all consumers are sophisticated and an equilibrium contract offer $f(\alpha), a(\alpha)$ for a pool in which a share of α is naive that satisfy

$$v - f' \geq v - f(\alpha).$$

Using that $k'(a) = \alpha$ in equilibrium, a' is bounded and hence zero-profits imply $f' = c$. Using this and equation 1, the above inequality is equivalent to

$$v - c \geq v - c + \alpha a(\alpha) - k(a(\alpha)).$$

Hence the contract offer $f' = c, a = 0$ satisfies the participation constraint in the candidate equilibrium with contract offer $f(\alpha), a(\alpha)$. A firm that deviates and offers the contract $c - \epsilon, (2\epsilon/\alpha)$ in the candidate equilibrium with contract offer $f(\alpha), a(\alpha)$ hence attracts all consumers and earns profits

$$c - \epsilon + \alpha \frac{2\epsilon}{\alpha} - k\left(\frac{2\epsilon}{\alpha}\right) = c + \epsilon - k\left(\frac{2\epsilon}{\alpha}\right),$$

which is strictly positive for sufficiently small ϵ. This contradicts the No Profitable Deviation condition of a competitive equilibrium. $\square$

Proof of Proposition 7. Let $a'(\alpha) = a(\alpha) - \bar{a}$. Using equation 2, naive consumers benefit from perfect targeting if and only if

$$(1 - \alpha)a(\alpha) + k(a(\alpha)) \geq k(a(1)).$$

Let $K(\alpha) = (1-\alpha)a(\alpha) + k(a(\alpha)) = (1-\alpha)\bar{a} + (1-\alpha)a'(\alpha) + k_+(a'(\alpha))$. Note that $K(1) = k(a(1))$. Since $k'_+(a'(\alpha)) = \alpha$, $a'(\alpha) \to 0$ as $\alpha \to 0$, and therefore $K(\alpha) \to \bar{a}$ as $\alpha \to 0$. Let $k(a(1)) = \bar{k} > 0$.

Then for all $\bar{a} < \bar{k}$, there exists an $\bar{\alpha}(\bar{k})$ such that for all $\alpha_0 < \bar{\alpha}(\bar{k})$ naive consumers are hurt from perfect targeting. This proves Part I of the Proposition.

Using that $k'(a(\alpha)) = \alpha$ and that $da(\alpha)/d\alpha = 1/k''(a(\alpha))$, one has

$$\frac{dK(\alpha)}{d\alpha} = -a(\alpha) + \frac{1}{k''(a(\alpha))},$$

which implies that $dK(\alpha)/d\alpha > 0$ if and only if $a(\alpha)k''(a(\alpha)) < 1$. Hence if $a(\alpha)k''(a(\alpha)) < 1$ for all α, $K(\alpha) < K(1)$, and thus naive consumers are hurt by perfect targeting. And similarly, in case $a(\alpha)k''(a(\alpha)) > 1$ for all $\alpha \in (\underline{\alpha}, 1)$, one has $K(\alpha) > K(1)$ for all $\alpha \in (\underline{\alpha}, 1)$; in this case, thus, naive consumers benefit from a perfect targeting technology. $\qquad\square$