Rev Ind Organ (2015) 47:341–354
DOI 10.1007/s11151-015-9473-0

On the Welfare Costs of Naiveté in the US Credit-Card Market

Paul Heidhues[1] · **Botond Kőszegi**[2]

Published online: 21 August 2015

Abstract In the presence of naive consumers, a participation distortion arises in competitive markets because the additional profits from naive consumers lead competitive firms to lower transparent prices below cost. Using a simple calibration, we argue that the participation distortion in the US credit-card market may be massive. Our results call for a redirection of some of the large amount of empirical research on the quantification of the welfare losses from market power, to the quantification of welfare losses that are due to the firms' reactions to consumer misunderstandings.

Keywords Sophistication · Naiveté · Credit market · Consumer exploitation

1 Introduction

Researchers have noted that exploitative contracting can induce a variety of total-welfare-reducing distortions (see Sect. 4 for a discussion of the literature). In this paper, we focus on one requirement for a market outcome to maximize total surplus: efficient participation requires that the marginal benefit to consumers equals the marginal cost of producing the good or service. We describe as a "participation

Heidhues and Kőszegi gratefully acknowledge financial support from the European Research Council (Starting Grant #313341).

✉ Paul Heidhues
Paul.Heidhues@esmt.org

Botond Kőszegi
botondkoszegi@gmail.com

[1] European School of Management and Technology, Schlossplatz 1, 10178 Berlin, Germany

[2] Central European University, Nádor utca 9, Budapest 1051, Hungary

distortion" the failure of this condition to be met and investigate the participation distortion that arises from the presence of naive consumers in the marketplace. In a competitive market, systematic underestimation of the total price by naive consumers as well as the transfer from more-profitable naive to less-profitable sophisticated consumers can lead to inefficient overparticipation. Based on a rough calibration for the US credit card market, we argue that the resulting welfare loss may be large when compared to conventional welfare losses such as those from market power. Our results call for more careful empirical investigation of the participation distortion and suggest that ignoring the distortions due to such contracting will lead to misguided welfare results in the credit-card industry. More generally, the results suggest that it may be fruitful to redirect some of the empirical work on the quantification of the welfare losses from market power to the quantification of the welfare losses due to firms' reactions to consumer misunderstandings.

Considerable discussion has focused on the redistributive character of consumers' misunderstanding of contingent charges or other costs that are associated with product ownership. For the UK banking industry, for example, Armstrong and Vickers (2012) demonstrate that contingent charges are—presumably unexpectedly—incurred by a minority of the population, who nevertheless often incurs these charges repeatedly and end up paying large amounts of money for contingent charges such as unarranged overdrafts of their current accounts. As this part of the population is presumably poorer than average,[1] the resulting "reverse Robin Hood" exercise leads to a disadvantageous redistribution. While we fully agree that the redistribution issue that results from consumer misunderstandings is important, we highlight that there are also likely to be considerable total welfare costs associated with consumer misunderstandings.

2 Setup

In this section, we introduce a simple reduced-form model of the credit market, which captures that some (naive) consumers underappreciate the interest on outstanding debt and penalties they will incur when owning a credit-card. Formally, $N \geq 2$ risk-neutral profit-maximizing firms that sell a homogenous good or service simultaneously make offers that consist of an anticipated price $f_n \in \mathbb{R}$ and an additional price $a_n \in [0, \overline{a}]$ to a population of consumers. Each consumer is naive with probability α and sophisticated with probability $1 - \alpha$,[2] 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

[1] Grubb (2014) provides strong evidence for this presumption.

[2] In our model, we assumed that the population consists of naive and sophisticated consumers only. But as long as naive consumers ignore the additional price, the fees that they end up paying ex post may differ across consumers without affecting our basic analysis of the participation distortion.

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 $D(f) = \text{Prob}[\text{v} \geq \text{f}]$ for the product, where $D(\cdot)$ is strictly decreasing, twice continuously differentiable, and—unless explicitly stated otherwise—$D(c) > 0$. 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.

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 (1) $[f(\alpha), a(\alpha)]$ earns zero profits; and (2) there is no f', a' with $f' < f(\alpha)$ that makes positive expected profits.

Since naive consumers ignore the additional price and sophisticated consumers can costlessly avoid paying it, the attractiveness of a contract is independent of the additional price a_n. Thus, condition (ii) implies that for any $\alpha > 0$ in the equilibrium contract $a(\alpha) = \overline{a}$. For firms to earn zero profits, thus, $f(\alpha) = c - \alpha\overline{a}$. Hence, the equilibrium contract is $(c - \alpha\overline{a}, \overline{a})$, the utility of a sophisticated consumer who buys the product is

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

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

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

3 Participation Distortions

3.1 Basic Calibration Exercise

To illustrate the potential importance of the participation distortion, we provide back-of-the-envelope calculations on the welfare losses that are 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 can be massive.

For our illustration, we assume that the demand curve $D(\cdot)$ is linear. Then, since the anticipated price equals $f(\alpha) = c - \alpha\overline{a}$, it follows from Figure 1 that for any $D(c) > 0$ the deadweight loss from the participation distortion equals

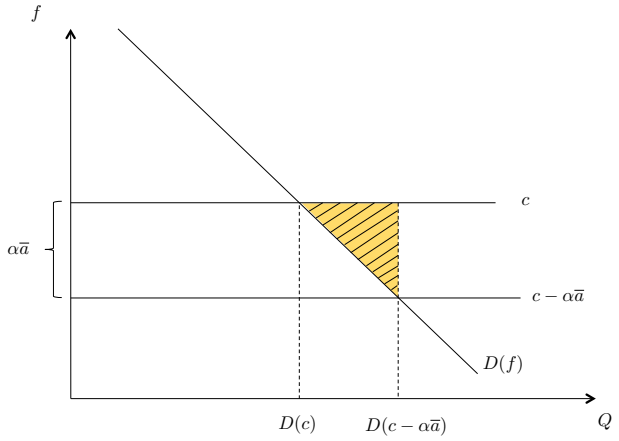

Fig. 1 Participation Distortion when $D(c) \geq 0$. The dead-weight loss from the participation distortion equals the shaded area between the marginal cost line and the inverse demand curve [i.e. is equal to $(1/2)\,\alpha\bar{a}\,[D(c-\alpha\bar{a})-D(c)]$]

$$\frac{1}{2}\alpha\bar{a}\cdot(D(c-\alpha\bar{a})-D(c)) = \frac{1}{2}(-D'(c-\alpha\bar{a}))(\alpha\bar{a})^2$$
$$= \frac{1}{2}\underbrace{\left(\frac{-D'(c-\alpha\bar{a})c}{D(c-\alpha\bar{a})}\right)}_{\text{elasticity}}\underbrace{\left(\frac{\alpha\bar{a}}{c}\right)^2}_{\text{misperception ratio}}\underbrace{cD(c-\alpha\bar{a})}_{\text{market-wide revenue}}. \qquad (3)$$

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 misperception 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 misperception 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 that issuers receive from merchants is set equal to the network externality from additional consumer purchases.[3] This

[3] 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 the two-sided-markets literature, see Rochet and Tirole (2006) and Rysman (2009). Much of the recent platform-market literature, however, implies that the interchange fee that merchant banks pay to those banks that issue credit cards to consumers is likely to be inefficiently high, leading to overparticipation of consumers even in models that abstract from consumer naiveté (see Bedre-Defolie and Calvano (2013)

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.[4] Combining these numbers, issuers earned roughly \$932 per household from credit-card consumers. If we assume a competitive industry and suppose that the fixed costs are negligible in the credit card issuing industry,[5] 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 (absolute value) 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 that 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).[6] If, instead, consumers believe not only that credit cards are free, but that credit cards offer valuable perks, the anticipated price may be even lower. For example, if $f = -\$100$, the participation distortion increases to roughly 61 % of the market size or \$566 per card-owning household.[7] In contrast, if consumers expect to pay fees over and above the perks that they anticipate benefiting from, the welfare costs

Footnote 3 continued
and the discussion of the literature therein). Although we have not investigated this formally, such baseline overparticipation would likely reinforce the welfare losses that we find.

[4] 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 Census (http://quickfacts.census.gov/qfd/states/00000.html), and Bricker et al. (2012).

[5] The assumption that fixed costs are negligible is consistent with the fact that there are many firms that issue credit cards in the US. A good estimation of credit card issuing firms' cost functions, however, would improve the welfare estimates below.

[6] To use the formula in Equ. 3, we need to verify that $D(c) \geq 0$ in our calibration. With linear inverse demand one has $f = A - BQ$, where Q is the quantity demanded and A, B demand parameters. Hence, $Q = (A/B) - (1/B)f$, and the (absolute value) elasticity in Equ. 3 is $\epsilon = c/BQ$, where Q is the observed market demand that corresponds to an anticipated price f. Hence, the intercept of the inverse demand is $A = f + (c/\epsilon)$. Since $D(c) \geq 0$ if and only if $A \geq c$, we thus need that $f \geq c(1 - (1/\epsilon))$. For an upfront price of $f \geq 0$ this holds whenever $\epsilon \geq 1$. Similarly, for $\epsilon = 1$, this holds if and only if $f \geq 0$. Thus, Equ. 3 applies to our baseline calibration.

[7] As is explained in Footnote 6, the inverse demand intercept $A < c$ in this case. As can be seen from Fig. 2, the dead weight loss from the participation distortion in case $A < c$ equals $(c - A)D(f) + (1/2)(A - f)D(f)$. Using that $A = f + (c/\epsilon)$, gives a participation distortion of $D(f)\{c - f - (c/2\epsilon)\}$.

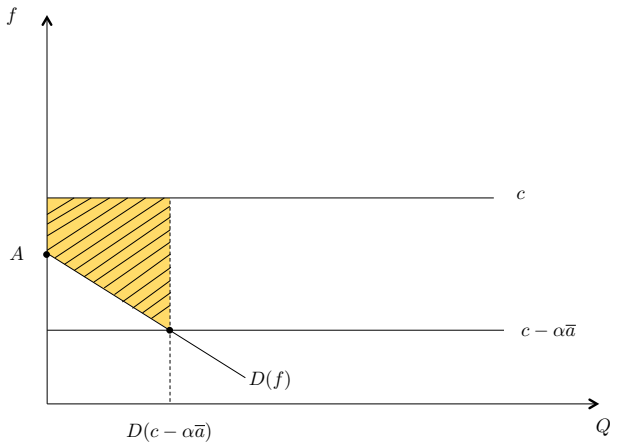

Fig. 2 Participation Distortion when $D(c) < 0$. The dead weight loss from the participation distortion equals the shaded area below marginal cost and above the inverse demand curve. It thus equals $(c - A)D(f) + (1/2)(A - f)D(f) = (c - a)D(c - \alpha\bar{a}) + (1/2)(A - c + \alpha\bar{a})D(c - \alpha\bar{a})$, where we use the fact that in a competitive market $f = c - \alpha\bar{a}$

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 23 % of the market size or $214 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—which include merger control, enforcement of competition law, enforcement of consumer protection law, and market studies—lead to direct consumer benefits of GBP 16 ($25) per household per year, and total (direct and indirect) benefits of GBP 78-227 ($122–355) per household per year.[8] 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 misperception of naive consumers in the credit-card market can be large—and this is just a single market and just one of the distortions from misperception.

[8] The OFT Impact report (http://www.oft.gov.uk/shared_oft/reports/Evaluating-OFTs-work/oft1493), estimates a total direct impact of GBP 422 million 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 GBP 136 million. 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)].

3.2 Sensitivity Analysis

3.2.1 Market Power

In our basic calibration, we supposed that firms have no market power in line with the fact that there are many firms offering credit cards in the US. We now ask how our results are affected if credit-card companies enjoy market power. Suppose that they have a degree of market power that induces them to charge an optimal markup of size m, so that $f = c + m - \alpha\bar{a}$ and total revenue is $c + m$. Obviously, as the participation distortion originates from an inefficiently low up-front price, a positive markup will reduce the welfare loss that we predict for an exogenously given number of firms. Nevertheless, it can still be very large even for sizable market power and a fixed number of firms.

Then, maintaining the assumption of linear demand, supposing that $D(c) \geq 0$, and using the same steps as in the derivation of Equ. 3, the welfare loss from participation distortion can be written as

$$\frac{1}{2}\underbrace{\left(\frac{-D'(c+m-\alpha\bar{a})(c+m)}{D(c+m-\alpha\bar{a})}\right)}_{\text{elasticity}}\underbrace{\left(\frac{\alpha\bar{a}-m}{c+m}\right)^2}_{\text{adjusted misperception ratio}}\underbrace{(c+m)D(c+m-\alpha\bar{a})}_{\text{market-wide revenue}}. \qquad (4)$$

If we assume, as in our baseline calibration above, that consumers expected to get a credit card for free ($f = 0$), the adjusted misperception ratio simplifies to $c/(c+m)$. Hence, when firms charge a x % markup over cost, our estimate above based on a competitive market needs to be divided by $1 + x/100$. Thus, if firms charge a 20 % markup, the welfare loss reduces from \$466 per credit-card-owning household to \$388, and even with a 50 % markup the welfare loss equals \$311 per credit-card-owning household.

3.2.2 Binding Price Floor

In related work with regard to the credit-card industry Heidhues et al. (2012, 2014, 2015), we argued that there is likely to be a price floor around zero in the credit-card industry. In our framework, this amounts to assuming that the anticipated price $f_n \geq 0$. Whenever the price floor is binding (i.e. $c - \alpha\bar{a} < 0$), this will in effect induce a positive markup m implicitly defined through $c - \alpha\bar{a} + m = 0$. Once the markup is determined, Equ. 4 can be used to determine the welfare loss from the participation distortion.

It is perhaps worth emphasizing that we calculate the short-run welfare loss, taking as given the number of firms. When a price floor leads to a positive markup due to the consumer exploitation, other welfare distortions—such as those of overentry or, as we analyze in detail in Heidhues et al. (2015), those due to the incentives to invent new ways to exploit consumers—are likely to arise.

3.2.3 Demand Elasticity

Perhaps the most questionable assumption is that on the (absolute value) elasticity of demand with respect to the anticipated price, which we set equal to 1. Under our assumption of a competitive industry and supposing that the credit-card annual fee is a fully visible price component, this amounts to assuming that (holding everything else constant) an increase in the annual fee from \$0 to \$90 would reduce the size of the credit-card market by 10 %. As demand becomes less elastic, the welfare loss decreases proportionally. For instance, under our baseline calibration with $f = 0$, the welfare loss per credit-card-owning household reduces to \$233 if the (absolute value) elasticity of demand with respect to the anticipated price is 1/2 instead of 1. If demand becomes more elastic, the welfare loss increases. For an elasticity of two with respect to the anticipated price, the welfare loss increases to \$699. Unfortunately, we are unaware of an empirical estimate of the elasticity necessary for our calibration. Crucially, one needs an estimate of the elasticity of demand with respect to the *anticipated price*. For example, a low elasticity of demand with respect to the interest rate may simply indicate that—consistent with the empirical findings of Ausubel (1991) and Shui and Ausubel (2004)—interest payments are fully or partially unanticipated.

In our baseline calibration we suppose for simplicity that demand is linear. It turns out that with linear demand and an elasticity of 1, the maximal willingness to pay for the credit card (A) is equal to the marginal cost, so it would be efficient to ban credit cards altogether. While our exact estimate of the welfare loss depends of course on the exact shape of the demand curve, the welfare loss can be very large even if it would be socially undesirable to ban credit cards altogether. For example, suppose that the linear demand curve of our baseline calibration is a good approximation over the range of up-front prices from \$0 to \$466, but at that point demand discontinuously increases to above marginal cost. Then, the total participation distortion is still \$350 per credit-card-owning household.

3.2.4 Adding an Exploitation Distortion

Another important caveat is that our model assumes that the exploitation of naive consumers does not generate any distortion beyond a participation distortion. Yet a number of additional distortions may arise: It may be costly for sophisticated consumers to avoid the additional price, firms may have direct costs of introducing an additional price, and firms may inefficiently adjust contracts to increase the additional price that they can collect. For instance, in Heidhues and Kőszegi (2010) we argue that credit-card companies use contracts that induce consumers to overborrow, thereby reducing the benefit that a consumer has when signing a credit-card contract.

To allow for such exploitation distortions, we need to alter slightly the model setup. We assume that when a firm chooses additional price $a_n \in \mathbb{R}_+$, it generates an "exploitation cost" $k(a_n)$, where $k(a) = 0$ for $a \leq \overline{a}$, $k'(\overline{a}) = 0$, $k''(a) > 0$ for $a \geq \overline{a}$, and $\lim_{a\to\infty} k'(a) > 1$. We suppose that the exploitation costs are incurred by all consumers that sign firm n's contract, and that consumers are aware that they will

incur these costs.[9] (Our results would be unchanged if firms and not consumers were paying the exploitation cost.) To fix ideas in our credit-card context, suppose the additional price is collected from unanticipated interest payments that are made by naive time-inconsistent borrowers. Firms that want to collect unanticipated interest payments beyond $\overline{a}$ need to induce consumers to borrow an inefficiently high amount. Furthermore, because firms cannot differentiate between sophisticated and naive consumers, the equilibrium contract induces both to overborrow. Then the exploitation cost $k(a_n)$ is the welfare loss that is associated with overborrowing.[10]

We next solve for the competitive-equilibrium contract $f(\alpha), a(\alpha)$. A consumer who accepts a contract, selects one among those that minimize her total anticipated cost $f + k(a)$. If we take as given that other firms offer the equilibrium contract, a given firm's maximization problem thus becomes

$$\begin{aligned} \max_{(f,a)} \quad & \alpha(f + a) + (1 - \alpha)f - c \\ \text{s.t.} \quad & f + k(a) \leq f(\alpha) + k(a(\alpha)). \end{aligned}$$

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 (2) of Definition 1. At the solution to this maximization problem, the participation constraint $f + k(a) \leq f(\alpha) + k(a(\alpha))$ must hold with equality. Substituting the constraint into the maximand and differentiating with respect to a yields $k'(a(\alpha)) = \alpha$. Using the zero-profit condition and the facts that sophisticated consumers pay only the anticipated price while naive consumers pay both prices, we get that the competitive-equilibrium contract has $a(\alpha) = (k')^{-1}(\alpha)$ and $f(\alpha) = c - \alpha a(\alpha)$, where $a(\alpha)$ is strictly increasing and $f(\alpha)$ is strictly decreasing. The utility of a sophisticated consumer with value v who buys the product is

$$U_s(\alpha) \equiv v - f(\alpha) = v - c + \alpha a(\alpha) - k(a(\alpha)), \tag{5}$$

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

$$U_n(\alpha) \equiv v - f(\alpha) - a(\alpha) = v - c - (1 - \alpha)a(\alpha) - k(a(\alpha)). \tag{6}$$

Figure 3 illustrates how explicitly incorporating the exploitation distortion affects the calibration of our welfare loss. If we ignore the exploitation distortion and set $\overline{a} = a(\alpha)$, the participation distortion is the triangle $\mathbb{A} + \mathbb{B}$ in the Figure. When we take the exploitation distortion into account, the equilibrium $a(\alpha)$ makes the contract less attractive, thereby reducing the number of consumers that end up buying it to those with valuation $v \geq f + k(a(\alpha))$ and the welfare loss from overparticipation to the triangle $\mathbb{B}$. At the same time, the cost $k(a(\alpha))$ is now borne

[9] For the case in which all consumers are naive, Grubb (2014), discusses how the results are affected if consumers only partially anticipate the exploitation costs.

[10] In Heidhues and Kőszegi (2014) we derive this reduced form formally from a credit-card contracting model with naive time-inconsistent borrowers, and establish that one can think of consumers as anticipating the exploitation cost that they bear.

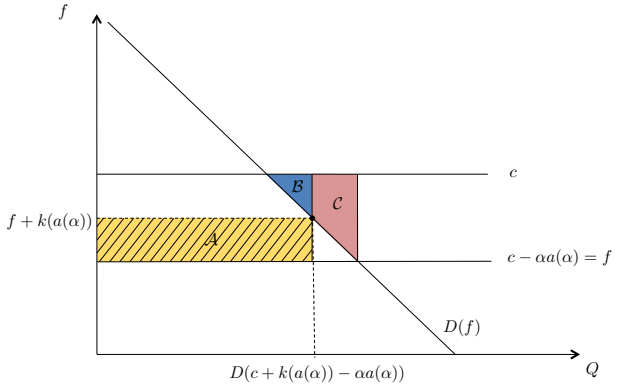

Fig. 3 Interplay between the exploitation and the participation distortion when $D(c) \geq 0$. The deadweight loss from the exploitation distortion is equal to the area $\mathbb{A}$ above, i.e., it equals $D(c + k(a(\alpha)) - \alpha a(\alpha))k(a(\alpha))$. The participation distortion equals the shaded triangle $\mathbb{B}$ as consumers will only buy if their valuation is above $f + k(a(\alpha))$. The explicit consideration of exploitation costs, hence, increases the welfare loss if the area $\mathbb{A} \geq \mathbb{C}$, and reduces it otherwise

by every consumer who signs a contract, adding the direct exploitation cost $\mathbb{A}$ in the Figure.

Using the same steps as in the derivation of Equ. 3 to determine the participation distortion, for the case of $D(c) \geq 0$ the total welfare loss can be written as

$$\frac{1}{2}\underbrace{\left(\frac{-D'(c+k(a(\alpha))-\alpha a(\alpha))c}{D(c+k(a(\alpha))-\alpha a(\alpha))}\right)}_{\text{elasticity}}\underbrace{\left(\frac{\alpha a(\alpha)-k(a(\alpha))}{c}\right)^2}_{\text{adjusted misperception ratio}}\underbrace{cD(c+k(a(\alpha))-\alpha a(\alpha))}_{\text{market-wide revenue}}$$
$$+\underbrace{k(a(\alpha))D(c+k(a(\alpha))-\alpha a(\alpha))}_{\text{direct exploitation distortion}}. \tag{7}$$

Taking our baseline calibration in which consumers believe they get a credit card for free ($f = 0$), and using that in a competitive equilibrium $f = c - \alpha a(\alpha)$, we again have $c = \alpha a(\alpha)$. Implicitly defining $\kappa \in [0, 1]$ through $k(a(\alpha)) \equiv \kappa \alpha a(\alpha)$,[11] we can rewrite Equ. 7 as

$$\left\{\frac{(1-\kappa)^2}{2}\left(\frac{-D'(c+k(a(\alpha))-\alpha a(\alpha))c}{D(c+k(a(\alpha))-\alpha a(\alpha))}\right)+\kappa\right\}cD(c+k(a(\alpha))-\alpha a(\alpha)).$$

In our baseline calibration in which the demand elasticity is set to one, this furthermore simplifies to

[11] Since the additional price maximizes $\alpha a(\alpha) - k(a(\alpha))$ in the competitive equilibrium contract, it is obvious that $\kappa \in [0, 1]$.

$$\frac{1+\kappa^2}{2}cD(c+k(a(\alpha))-\alpha a(\alpha)).$$

Incorporating the exploitation distortion in our baseline calibration thus—independently of the functional form and size of the exploitation cost—increases the resulting total welfare distortion.[12]

4 Related Literature

Our calibration is based on the premise 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 different consumer-retail-finance 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 et al. (2008) indicates that many credit-card consumers seem to not know or forget about various fees that issuers impose. Shui and Ausubel (2004) document that consumers that receive credit-card solicitations overrespond to the introductory ("teaser") interest rate relative to the post-introductory rate, which suggests that they end up borrowing more than they intended or expected. Woodward and Hall (2012) find that borrowers underestimate broker compensation, and Gerardi et al. (2009) document that 26 % of borrowers who face a prepayment penalty are unaware of it. Evidence by Wilcox (2003), Barber et al. (2005), and Anagol and Kim (2012) indicates that investors underweight operating expenses when choosing mutual funds.

Furthermore, there is evidence from a variety of industries that firms design contracts to exploit consumers' or workers' mistakes (DellaVigna and Malmendier 2004; Grubb 2009, 2015; Hoffman and Burks 2013). In such settings, recent research in behavioral contract theory emphasizes that consumer mispredictions allow firms to collect more money from consumers than the latter anticipate when entering the relationship with the firm (DellaVigna and Malmendier 2006, 2004; Gabaix and Laibson 2006; Laibson and Yariv 2007). But competition for naive consumers will return much of these unanticipated payments to consumers, so that it need not harm naive consumers (DellaVigna and Malmendier 2006; Gabaix and Laibson 2006; Laibson and Yariv 2007). The behavioral contracting literature, however, has identified a number of reasons why welfare may still suffer from exploitation even in competitive markets.

First, if firms cannot identify which consumers are naive and which are sophisticated, naive consumers often cross-subsidize more sophisticated ones (Gabaix and Laibson 2006; Armstrong and Vickers 2012), and this in itself may have undesirable redistributive consequences. In contrast, we focus on the total welfare cost of consumer exploitation.

[12] This finding extends to any (absolute value) demand elasticity that is less than one.

Second, firms may distort contracts in order to better exploit naive consumers (DellaVigna and Malmendier 2004; Eliaz and Spiegler 2006; Grubb 2009, 2015; Gottlieb and Smetters 2012; Spinnewijn 2012), thereby inducing distorted consumption choices by either sophisticated or naive consumers. For example, in Heidhues and Kőszegi (2010) firms offer credit with overly front-loaded repayment plans, and large penalties for delaying repayment. Naive time-inconsistent borrowers believe they will repay early, and thus that credit is cheap; but they end up repaying late and hence incurring large penalties that make credit expensive. Because credit seems cheap to naive consumers, they are willing to borrow too much, and this inefficient excessive overborrowing reduces welfare. The more naive consumers in the market place, the more firms focus on exploiting naive consumers, and the greater the contract distortion.[13] In contrast to these papers, our main calibration abstracts from the exploitation costs that society bears when contracts are inefficiently written so as to better exploit them, although we do illustrate how such distortions may affect our results.

Third, if there is a price floor on the basic price for a contract—e.g. because overly negative basic prices attract rational arbitrageurs who are willing to collect this money and at the same time can avoid paying the additional fees that are designed for naive consumers—firms cannot profitably hand back all of the unanticipated payments to consumers, which causes naive consumers to pay more than the cost of the service. These positive profits can induce firms to engage in inefficient activities, such as exploitative innovation (Heidhues et al. 2015) or excessive entry into the industry. While we briefly discuss the impact of such a price floor in our sensitivity analysis, we ignore reductions in total welfare due to exploitative innovation or over-entry in our calibration.

Fourth, if sophisticated consumers cannot avoid paying the extra costs that naive consumers underestimate—think for example of management fees in the mutual fund industry—sophisticated consumers will self-separate into buying an efficient contract—such as low-fee exchange traded fund—while naive consumers select an inefficient exploitative one, and firms may earn positive profits from selling to naive consumers (Heidhues et al. 2014). While the basic setup of our paper covers a variety of competitive markets other than the credit market, it is inappropriate for markets such as the managed mutual fund market in which all consumers must pay the same fee.

Fifth, and most closely related to our paper, as naive consumers underestimate the cost of the contract, they may purchase even if their valuation for the product is below its cost, leading to a participation distortion (Grubb 2014; Heidhues et al. 2014; Heidhues and Kőszegi 2014). Similarly, if naive consumers' misestimation of the price leads them to misestimate the product's quality as well, naive consumers may inefficiently buy low-quality products (Michel 2014). Or if naive consumers biases in interpreting past success induces them to believe falsely that firms produce a valuable service even though they do not, a completely useless industry may thrive by selling to naive agents (Spiegler 2006). Our work contributes by illustrating that

[13] See Heidhues and Kőszegi (2014) for additional examples in which exploitative contracting leads to distorted consumption choices.

the equilibrium participation distortion from naive consumers who underestimate their total payments can be large.

5 Conclusion

We considered the US credit-card market and argued that the participation distortion from the exploitation of consumer naiveté may lead to a very large welfare cost. There are a number of issues our calibration abstracted from. For example, we took the number of firms as given. The ability to generate revenue from exploiting consumers ex post will likely induce excessive entry; this is an issue that a full empirical investigation of the welfare cost would want to take account of. We also abstracted from estimating the welfare loss from consumers who sign distorted contracts—the direct exploitative contracting costs. A full analysis of the welfare costs of exploitative contracting would need to take these into account. How these can be estimated is an interesting question for future research.

Acknowledgments We thank Özlem Bedre-Defolie for helpful comments, and the editors Michael Grubb and Victor Tremblay for their helpful feedback. We are grateful to Johannes Johnen for his research assistance.

References

Agarwal, S., Driscoll, J. C., Gabaix, X., & Laibson, D. (2008). Learning in the credit card market. NBER working paper #13822.

Anagol, S., & Kim, H. (2012). The impact of shrouded fees: Evidence from a natural experiment in the Indian mutual funds market. *American Economic Review*, *102*(1), 576–593.

Armstrong, M., & Vickers, J. (2012). Consumer protection and contingent charges. *Journal of Economic Literature*, *50*(2), 477–493.

Ausubel, L. M. (1991). The failure of competition in the credit card market. *American Economic Review*, *81*(1), 50–81.

Ausubel, L. M. (1999). Adverse selection in the credit card market. Working paper, University of Maryland.

Barber, B. M., Odean, T., & Zheng, L. (2005). Out of sight, out of mind: The effects of expenses on mutual fund flows. *Journal of Business*, *78*(6), 2095–2119.

Bedre-Defolie, Ö., & Calvano, E. (2013). Pricing payment cards. *American Economic Journal: Microeconomics, 5*(3), 206–231.

Bricker, J., Kennickell, A. B., Moore, K. B., & Sabelhaus, J. (2012). Changes in U.S. family finances from 2007 to 2010: Evidence from the survey of consumer finances. *Federal Reserve Bulletin*, *98*(2), 1–80.

DellaVigna, S., & Malmendier, U. (2004). Contract design and self-control: Theory and evidence. *Quarterly Journal of Economics*, *119*(2), 353–402.

DellaVigna, S., & Malmendier, U. (2006). Paying not to go to the gym. *American Economic Review*, *96*(3), 694–719.

Eliaz, K., & Spiegler, R. (2006). Contracting with diversely naive agents. *Review of Economic Studies*, *73*(3), 689–714.

Gabaix, X., & Laibson, D. (2006). Shrouded attributes, consumer myopia, and information suppression in competitive markets. *Quarterly Journal of Economics*, *121*(2), 505–540.

Gerardi, K., Goette, L., & Meier, S. (2009). Financial literacy and mortgage outcomes. Paper presented at the ASSA meetings.

Gottlieb, D., & Smetters, K. (2012). Narrow framing and life insurance. NBER working paper #18601.

Grubb, M. D. (2009). Selling to overconfident consumers. *American Economic Review*, *99*(5), 1770–1805.
Grubb, M. D. (2014). Overconfident consumers in the marketplace. Boston College working paper.
Grubb, M. D. (2015). Consumer inattention and bill-shock regulation. *Review of Economic Studies, 82*(1), 219–257.
Heidhues, P., & Kőszegi, B. (2010). Exploiting naivete about self-control in the credit market. *American Economic Review*, *100*(5), 2279–2303.
Heidhues, P., & őszegi, B. K. (2014). Using information about consumer naivete to price discriminate. Working paper.
Heidhues, P., Kőszegi, B., & Murooka, T. (2012). Deception and consumer protection in competitive markets. In *The pros and cons of consumer protection*. Swedish Competition Authority, chapter 3. http://www.konkurrensverket.se/globalassets/english/research/readthebook23mb.pdf.
Heidhues, P., Kőszegi, B., & Murooka, T. (2014). Inferior products and profitable deception. Working paper, http://www.personal.ceu.hu/staff/Botond_Koszegi/inferior_products.pdf.
Heidhues, P., Kőszegi, B., & Murooka, T. (2015). Exploitative innovation. *American Economic Journal: Microeconomics*. (*forthcoming*).
Hoffman, M., & Burks, S. (2013). Training contracts, worker overconfidence, and the provision of firm-sponsored general training. Working paper.
Laibson, D. I., & Yariv, L. (2007). Safety in Markets: An impossibility theorem for dutch books. Working paper, Caltech.
Michel, C. (2014). Contractual structures and consumer misperceptions—the case of warranties. Working paper.
Office of Fair Trading. (2008). Personal current accounts in the UK, Technical report.
Rochet, J. C., & Tirole, J. (2006). Two-sided markets: A progress report. *RAND Journal of Economics*, *37*(3), 645–667.
Rysman, M. (2009). The economics of two-sided markets. *The Journal of Economic Perspectives*, *23*(3), 125–143.
Shui, H., & Ausubel, L. M. (2004). Time inconsistency in the credit card market. http://ssrn.com/abstract=586622.
Spiegler, R. (2006). The market for quacks. *Review of Economic Studies*, *73*(4), 1113–1131.
Spinnewijn, J. (2012). Unemployed but optimistic: Optimal insurance design with biased beliefs. Working paper.
Stango, V., & Zinman, J. (2009). What do consumers really pay on their checking and credit card accounts? Explicit, implicit, and avoidable costs. *American Economic Review Papers and Proceedings*, *99*(2), 424–429.
Wang, Z. (2012). Debit card interchange fee regulation: Some assessments and considerations. *Economic Quarterly, Third Quarter*, *98*(3), 159–183.
Wilcox, R. T. (2003). Bargain hunting or star gazing? Investors' preferences for stock mutual funds. *The Journal of Business*, *76*(4), 645–663.
Woodward, S. E., & Hall, R. E. (2012). Diagnosing consumer confusion and sub-optimal shopping effort: Theory and mortgage-market evidence. *American Economic Review*, *100*(2), 511–515.