THE
QUARTERLY JOURNAL
OF ECONOMICS

Vol. CXIX May 2004 Issue 2

CONTRACT DESIGN AND SELF-CONTROL: THEORY AND EVIDENCE*

STEFANO DELLAVIGNA AND ULRIKE MALMENDIER

How do rational firms respond to consumer biases? In this paper we analyze the profit-maximizing contract design of firms if consumers have time-inconsistent preferences and are partially naive about it. We consider markets for two types of goods: goods with immediate costs and delayed benefits (investment goods) such as health club attendance, and goods with immediate benefits and delayed costs (leisure goods) such as credit card-financed consumption. We establish three features of the profit-maximizing contract design with partially naive time-inconsistent consumers. First, firms price investment goods below marginal cost. Second, firms price leisure goods above marginal cost. Third, for all types of goods firms introduce switching costs and charge back-loaded fees. The contractual design targets consumer misperception of future consumption and underestimation of the renewal probability. The predictions of the theory match the empirical contract design in the credit card, gambling, health club, life insurance, mail order, mobile phone, and vacation time-sharing industries. We also show that time inconsistency has adverse effects on consumer welfare only if consumers are naive.

* We thank four exceptional referees, George Baker, Daniel Benjamin, Drew Fudenberg, Luis Garicano, Jerry Green, Oliver Hart, Caroline Hoxby, Markus Möbius, Daniele Paserman, Ben Polak, Andrei Shleifer, and in particular Philippe Aghion, Edward Glaeser, Lawrence Katz, and David Laibson. We received helpful comments from participants of seminars at the Econometric Society Summer Meeting 2001, AEA Meeting 2002, the Behavioral Public Finance Conference 2003, the University of California at Berkeley, Humboldt University (Berlin), Bonn University, Boston University, Harvard University, INSEAD, University of Maryland (College Park), the Massachusetts Institute of Technology, and Northwestern, Stanford, Yale, and Zurich universities. Nageeb Ali, Saurabh Bhargava, Madhav Chandrasekher, Tricia Glynn, Ming Mai, Boris Nenchev, and Christine Yee provided excellent research assistance. For financial support, DellaVigna thanks Bank of Italy and Harvard University, and Malmendier thanks Harvard University and the German Academic Exchange Service (DAAD).

The Quarterly Journal of Economics, May 2004

I. Introduction

A growing body of laboratory and field evidence documents deviations from standard preferences and biases in decision-making. If the deviations are systematic and persistent, profit-maximizing firms should respond to them and tailor their contracts and pricing schemes in response.

In this paper we consider the market interaction between profit-maximizing firms and consumers with time-inconsistent preferences and naive beliefs. We derive the optimal contract design and compare it with observed features of contracts in several industries. We also consider the implications of the market interaction for consumer welfare. We see this as a step in the direction of integrating the findings from behavioral economics into industrial organization and contract theory.

We assume that consumers have a quasi-hyperbolic discount function [Strotz 1956; Phelps and Pollak 1968; Laibson 1997; O'Donoghue and Rabin 1999a], with a higher discount rate between the present and the next period than between any of the subsequent periods. This discount function implies time inconsistency, since the discount rate between two periods depends on the time of evaluation. We consider both agents who are sophisticated about their time inconsistency, as well as agents who are (partially) naive about it [O'Donoghue and Rabin 2001]. The latter assumption is motivated by the experimental evidence on overconfidence about positive personal attributes [Larwood and Whittaker 1977; Svenson 1981] and is consistent with field evidence on 401(k) investment [Madrian and Shea 2001], task completion [Ariely and Wertenbroch 2002], and health club attendance [DellaVigna and Malmendier 2003]. We also consider the standard case of time-consistent discounting. Throughout the paper we maintain the assumption that consumers have homogeneous time preferences and beliefs, except for a brief discussion of heterogeneity in subsection IV.F.

We consider firms that produce investment and leisure goods. Investment goods have current costs and future benefits relative to the best alternative activity—for example, health club attendance involves current effort cost and delivers future health benefits. Leisure goods have current benefits and future costs relative to the best alternative activity—for example, credit card borrowing increases current consumption at the expense of future consumption.

In Section II we derive the profit-maximizing two-part tariff in a two-period model. In a market for investment goods, a monopolistic firm prices below marginal cost if consumers have time-inconsistent preferences, regardless of the degree of sophistication. Sophisticated individuals demand commitment devices to increase their consumption of investment goods. The firm supplies these devices in the form of low per-usage prices. Naive users overestimate their future self-control and therefore their usage of the investment goods. The firm offers a contract with a discount on the per-usage price and a higher flat fee, since the individuals overestimate the value of the discount. The result of below-marginal-cost pricing extends to the case of perfect competition.

In markets for leisure goods the opposite result holds. Individuals with time-inconsistent preferences demand commitment devices to limit their usage (if sophisticated) or underestimate their usage (if naive). In both cases, competitive or monopolistic firms price usage above marginal cost.

The profit-maximizing contractual design has different welfare effects for sophisticated and naive users. If the agents are sophisticated, market interaction has favorable welfare effects, both under monopoly and under perfect competition. Firms offer a perfect commitment device, which enables the agents to achieve the efficient consumption level. In the market equilibrium, therefore, the lack of self-control has no effect on consumer welfare for sophisticated agents.

If the agents, instead, are (partially) naive, firms design contracts that exploit the consumers' misperception of their future behavior, with two adverse effects on consumer welfare. First, the market outcomes are inefficient, since firms do not maximize the actual consumer-firm surplus, but only the fictitious surplus. Second, under monopoly, naiveté induces a redistribution of surplus from the consumers to the firm. Under perfect competition, there still is an efficiency loss, but there is no redistributional effect, since the consumers are the residual claimants of the profits from their naiveté. Perfect competition therefore tempers the adverse effects of naiveté on consumer welfare.

In Section III we compare the predicted contract design with the empirical features of contracts in several industries producing investment and leisure goods. In a typical health club contract, for example, users pay flat fees but no price per visit, despite marginal costs per attendance estimated between $3 and

$6 [DellaVigna and Malmendier 2003]. This deviation from marginal-cost pricing does not appear to be explained by price discrimination. Under flat-rate contracts, frequent health club users pay the same overall fee as infrequent users, who presumably have a lower willingness to pay. Price discrimination would predict that frequent users should pay more.

As a second example, consider credit card contracts. Credit card companies price usage of the credit line above marginal cost, as indicated by the 20 percent premium on the resale of credit card debt [Ausubel 1991]. This pricing structure is puzzling since users with no outstanding balance pay a negative price for the credit card services.

In Section IV we extend the model to a three-period setting with dynamic competition and switching costs. Firms choose the optimal contract design, including the transaction cost of cancellation, in a class of generalized two-part tariffs. If consumers are time-consistent or sophisticated time-inconsistent, firms have no incentive to charge renewal fees or to create cancellation costs. If, instead, consumers are naive, firms charge back-loaded fees and design contracts with automatic renewal and endogenous transaction costs of switching. Firms choose these features in response to the naive consumers' underestimation of the renewal probability. The contract offered is discontinuous in the degree of naiveté. Even a minimal amount of naiveté induces firms to charge back-loaded fees and design contracts with automatic renewal.

In Section V we provide evidence on these contractual features. The typical credit card offer grants a low interest rate on outstanding balances for an initial "teaser" period, after which the interest rate rises to a high level. The back-loaded fee structure is consistent with evidence on consumers' underestimation of borrowing after the teaser period [Ausubel 1999]. We discuss alternative explanations based on ex ante asymmetric information or ex post exploitation. As a second example, the typical health club membership is extended automatically from month to month. In order to quit, consumers have to mail a cancellation letter or cancel in person, even though firms could make cancellation as easy as sending an email or giving a phone call.

We also present stylized evidence on the contractual design in the gambling, life insurance, mail order, mobile phone, and vacation time-sharing industries. Our findings suggest a new, unified explanation for aspects of the pricing in these diverse industries, based on the intertemporal features of the goods and on switching costs. Firms appear to respond in their contract design to time inconsistency and partial naiveté among consumers.

Contract design, therefore, provides evidence on the prevalence of nonstandard preferences and behavioral biases among consumers.[1] The design of contracts is a key test for the relevance of deviations. Firms would not respond to consumer deviations that are not systematic or limited to small stakes.

This paper contributes to the literature on the market interaction between rational and nonrational agents [Akerlof and Yellen 1985; De Long et al. 1990].[2] Differently from the previous literature, we focus on the asymmetric interaction between rational firms and consumers with time-inconsistent preferences.

Several results obtained in this paper for time-inconsistent agents are likely to apply to other deviations from standard preferences, such as limited memory or cognitive abilities. In particular, these deviations need not reduce consumer welfare as long as consumers are fully aware of them. Market interaction may lead to a technology to overcome the limitations. Even a small amount of naiveté, however, implies welfare losses from the market interaction and can generate large deviations in contract design.

The remainder of the paper is organized as follows. In Section II we present a simple model of market interaction between consumers and profit-maximizing firms, including welfare effects. In Section III we compare the predictions of the model with stylized features of contract design in several industries. In Section IV we extend the model to repeated contracts with switching costs, and in Section V we compare those predictions with observed contract design. Section VI concludes.

II. Two-Period Model

In this section we set up a simple two-period model of consumer behavior and firm pricing. We allow consumers to be time-inconsistent and overconfident (with the standard model as the benchmark) and analyze the profit-maximizing contract design.

1. Angeletos et al. [2001], DellaVigna and Malmendier [2003], DellaVigna and Paserman [2000], Gruber and Koszegi [2001], Gruber and Mullainathan [2002], and Fang and Silverman [2001] use field data to test for time consistency.

2. Bénabou and Tirole [2003], Gabaix and Laibson [2004], O'Donoghue and Rabin [1999b], and Russell and Thaler [1985] also consider the interaction of standard and nonstandard agents.

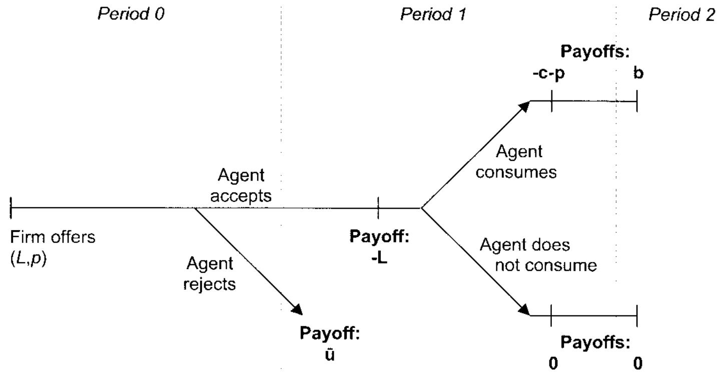

FIGURE I
Timing of Simple Model

II.A. *The Setting*

The consumer faces a monopolistic firm. In period 0 the firm proposes a two-part tariff (L,p) to the consumer (Figure I). If the consumer rejects, the firm makes zero profits, and the consumer attains the reservation utility $\bar{u}$ at $t = 1$. If the consumer accepts, she pays the lump-sum fee L in period 1 to the firm. Upon accepting, the consumer learns her cost type c (see below) and then chooses her consumption in period 1. The agent can choose to consume (C) or not to consume (NC). If she chooses C, she pays p to the firm and incurs cost c in $t = 1$; she then receives b in $t = 2$. If she chooses NC, she attains payoff 0 in $t = 1$ and in $t = 2$.

Consumer. Consumption C provides immediate payoff $-c$ at $t = 1$ and positive delayed payoff $b > 0$ at $t = 2$. Compared with the alternative activity NC, therefore, consumption C is costly at present and provides benefits in the future. We call goods like C with a net positive future payoff relative to the alternative activity *investment goods.* Examples include activities beneficial to health (visits to a doctor, health club attendance, sports), to education (schooling and vocational training), or to income (signing up for a retirement plan and switching to a cheaper telephone provider).

At $t = 0$, the agent knows the distribution F from which the costs c are drawn. We assume that F has a strictly positive density f over $\mathbb{R}$. At the end of period 0 the agent learns c. The benefits b are deterministic and known from the outset.[3]

Intertemporal preferences. We assume that the agent has quasi-hyperbolic preferences [Phelps and Pollak 1968; Laibson 1997; O'Donoghue and Rabin 1999a]. The discount function for time s, when evaluated at period t, equals 1 for $s = t$ and equals $\beta\delta^{s-t}$ for $s = t + 1, t + 2, \ldots$ with $\beta \leq 1$. The present value of a flow of future utilities $(u_s)_{s \geq t}$ as of time t is

$$u_t + \beta \sum_{s=t+1}^{\infty} \delta^{s-t} u_s. \tag{1}$$

We can interpret β as the parameter of short-run discounting and δ as the parameter of long-run discounting. The standard *time-consistent exponential* model corresponds to the case where β is equal to 1. If β is smaller than 1, the individual exhibits time-varying discounting. The discount factor between the present period and the next period is $\beta\delta$, while the discount factor between any two adjacent periods in the future is simply δ. The difference between the short-run and the long-run discount factors generates *time inconsistency.*

We allow for consumers who overestimate their time consistency. A *partially naive hyperbolic* agent with parameters $(\beta,\hat{\beta},\delta)$ [O'Donoghue and Rabin 2001] expects (erroneously) to have the discount function $1, \hat{\beta}\delta, \hat{\beta}\delta^2, \ldots$ with $\beta \leq \hat{\beta} \leq 1$ in all future periods. The individual therefore correctly anticipates that she will have hyperbolic preferences in the future, but she overestimates the future parameter of short-run discounting if $\beta < \hat{\beta}$. The difference between the perceived and actual future short-run discount factor $\hat{\beta} - \beta$ reflects the *overconfidence* about future self-control.

Three special cases deserve mention. An *exponential* agent has time-consistent preferences ($\beta = 1$) and is aware of it ($\hat{\beta} = 1$). A *sophisticated* agent has time-inconsistent preferences ($\beta < 1$) and is aware of it ($\hat{\beta} = \beta$). A fully *naive* agent has time-inconsistent preferences ($\beta < 1$), but is completely unaware of it ($\hat{\beta} = 1$). She believes that she will behave like a time-consistent agent in the future.

3. The benefits b can be interpreted as the expected net present value of the (possibly stochastic) future payoffs.

Firm. The production costs consist of a setup cost $K \geq 0$ that the firm incurs at $t = 1$ whenever a consumer signs the contract, and a per-unit cost $a \geq 0$, incurred at $t = 1$ whenever an agent consumes C. The monopolistic firm has all the bargaining power and offers a nonrenegotiable contract to the consumer at $t = 0$. The firm chooses a two-part pricing scheme (L,p) with a lump-sum fee L and a per-usage price p, both due at $t = 1$.

At time $t = 0$ the firm maximizes the discounted net present value of the expected future profits. Given that the firm can borrow and lend on the credit market, the discount factor is determined by the market interest rate r as $1/(1 + r)$. We assume that this discount factor equals the long-run discount factor[4] for individuals; i.e., $1/(1 + r) = \delta$. Finally, we suppose that the firm has complete knowledge of the individuals' preferences and that, as of $t = 0$, it knows the cost distribution F.

II.B. Consumer Behavior

At $t = 0$, the consumer evaluates consumption C as follows. She discounts by $\beta\delta$ the cost of consumption c and the per-usage fee p due at $t = 1$, and by $\beta\delta^2$ the benefits b accruing at $t = 2$. She therefore assigns discounted utility $\beta\delta(\delta b - p - c)$ to C and utility 0 to NC. Thus, she would like her future self to choose C at $t = 1$, upon learning the type c, whenever $c \leq \delta b - p$.

A *time-inconsistent* agent, though, will choose C less often than her previous self wishes. At the moment of deciding between C and NC, the net payoff from C equals $\beta\delta b - p - c$. Therefore, at $t = 1$ she chooses C if $c \leq \beta\delta b - p$, i.e., with probability $F(\beta\delta b - p)$. The parameter of short-run impatience, β, determines the difference between desired and actual consumption probability $F(\delta b - p) - F(\beta\delta b - p)$. This difference is zero for individuals with time-consistent preferences ($\beta = 1$). The smaller is β, the larger is this difference, and the more serious are the self-control problems.

A *partially naive hyperbolic* individual is not fully aware of her time inconsistency. Therefore, as of $t = 0$, she overestimates the probability that her future self will consume C at $t = 1$. She expects that she will consume if $\hat{\beta}\delta b - p - c \geq 0$, i.e., with probability $F(\hat{\beta}\delta b - p)$. The difference between forecasted and

4. In an economy populated mostly by hyperbolic agents, the market interest rate satisfies $1/(1 + r) = \delta$ if either (i) the richest agents have exponential preferences with discount factor δ, or (ii) the richest agents are sophisticated hyperbolic individuals who can afford commitment devices.

actual consumption probability, $F(\hat{\beta}\delta b - p) - F(\beta\delta b - p) \geq 0$, is a measure of the consumer's *overconfidence.* Time-consistent ($\beta = \hat{\beta} = 1$) and sophisticated agents ($\beta = \hat{\beta} < 1$) have rational expectations about their future time preferences and display no overconfidence.

Thus, a consumer who signs the contract (L,p) expects at time 0 to attain the net benefit $\beta\delta[-L + \int_{-\infty}^{\hat{\beta}\delta b - p} (\delta b - p - c)\, dF(c)]$. (Notice the $\hat{\beta}$ in the integral.)

II.C. Firm Behavior

If the consumer signs the contract, profits accrue to the monopolistic firm from the difference between the lump-sum fee L and the setup cost K and from the per-usage net gain $p - a$. The firm earns this latter part of the profit whenever the user chooses C, i.e., with probability $F(\beta\delta b - p)$. Therefore, as of $t = 0$, the expected per-usage net gain amounts to $\delta F(\beta\delta b - p)(p - a)$. The firm maximizes profits subject to the participation constraint that the discounted perceived utility equal the reservation utility $\beta\delta\bar{u}$. The maximization problem is thus[5]

$$\max_{L,p} \delta\{L - K + F(\beta\delta b - p)(p - a)\} \tag{2}$$

subject to

$$\beta\delta\left[-L + \int_{-\infty}^{\hat{\beta}\delta b - p} (\delta b - p - c)\, dF(c)\right] = \beta\delta\bar{u}. \tag{3}$$

Substituting for L in (2) yields

$$\max_p \Pi(p) = \max_p \delta\left[\left(\int_{-\infty}^{\beta\delta b - p} (\delta b - a - c)\, dF(c) - K - \bar{u}\right) + \int_{\beta\delta b - p}^{\hat{\beta}\delta b - p} (\delta b - p - c)\, dF(c)\right], \tag{4}$$

where the optimal L is determined by (3). The first term of (4) is the actual *social* surplus generated by the market interaction of the two parties (from the perspective of time 0). The integrand

5. We are assuming the existence of a two-part tariff (L,p) that satisfies the individual rationality constraint and that provides nonnegative profits to the firm. Otherwise, there is no market for the investment good.

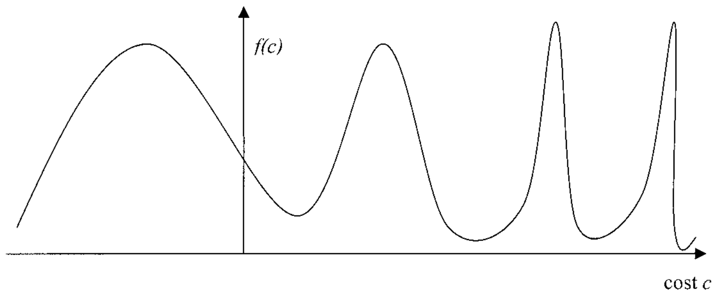

FIGURE IIa
Violation of Asymptotically Bounded Peaks (ABP) Condition

consists of the net benefit to the user, $\delta b - c$, minus the cost for the provider, a. The per-usage price p does not appear in the integrand since it is a mere transfer between the two parties. The second term of (4) is the fictitious *consumer* surplus that reflects the overconfidence of the consumer. At time 0 the consumer erroneously expects to attain this additional surplus. This term is null for time-consistent and sophisticated users ($\hat{\beta} = \beta$) and increases with naiveté as measured by $\hat{\beta} - \beta$.

In order to guarantee existence of a profit-maximizing contract, we introduce a technical assumption that we maintain throughout the paper.

ASSUMPTION ABP (Asymptotically bounded peaks). There is a pair $(M,z) \in \mathbb{R}^2$ such that $f(y'') \leq Mf(y')$ for all y', y'' with $z < |y'| < |y''|$ and $y' \cdot y'' > 0$.

Assumption ABP rules out the anomalous case of unbounded peaks on the tails of $f(c)$ (Figure IIa). It is satisfied by all standard distribution functions.

PROPOSITION 1 (Two-period model, investment goods, monopoly). Under monopoly, for $b > 0$, a profit-maximizing contract (L^*,p^*) exists. The per-usage price p^* equals marginal cost ($p^* = a$) for $\beta = 1$ and is set below marginal cost ($p^* < a$) for $\beta < 1$. The lump-sum fee L^* is set to satisfy the individual rationality constraint (3).

Proof of Proposition 1. We show that there exists a finite $\underline{M}$ such that the profit-maximizing price p must lie in $[\underline{M},a]$. The derivative of the profit function with respect to p is

$$\text{(5)}\quad \frac{\partial \Pi(p)}{\partial p} = \delta f(\beta\delta b - p)\left[(a-p) - (1-\hat{\beta})\delta b\,\frac{f(\hat{\beta}\delta b - p)}{f(\beta\delta b - p)} - \frac{F(\hat{\beta}\delta b - p) - F(\beta\delta b - p)}{f(\beta\delta b - p)}\right].$$

Given $f > 0$, $\partial\Pi(p)/\partial p < 0$ for $p > a$. Therefore, any contract with $p > a$ (and L satisfying (3)) generates lower profits than the contract with $p = a$. We now use Assumption ABP to construct the lower bound $\underline{M}$. Consider a pair (M,z) satisfying Assumption ABP. For $p < \beta\delta b - z$, the second term in brackets in (5) is bounded below by $-(1 - \hat{\beta})\delta bM$, and the third term is bounded below by $-(\hat{\beta} - \beta)\delta bM$. Therefore, the expression in brackets is positive for $p < a + \min\,(-(1 - \beta)\delta bM,\ \beta\delta b - z - a)$. The right-hand side of the inequality provides the lower bound $\underline{M}$. Any contract with p smaller than $\underline{M}$ generates lower profits than the contract with p equal to $\underline{M}$. The existence of a solution for p^* follows from continuity of the profit function on the compact set $[\underline{M},a]$. By continuity, any profit-maximizing price p^* must satisfy $\partial\Pi(p^*)/\partial p = 0$. As for the value of p^*, $\beta < 1$ implies that $\partial\Pi(p)/\partial p < 0$ at $p = a$ and therefore $p^* < a$. Finally, $\beta = \hat{\beta} = 1$ implies that $\partial\Pi(p)/\partial p = 0$ is satisfied only for $p = a$. QED

The first-order condition of (4) with respect to the per-usage price p can be rearranged to yield

(6)

$$p^* - a = -(1-\hat{\beta})\,\delta b\,\frac{f(\hat{\beta}\delta b - p^*)}{f(\beta\delta b - p^*)} - \frac{F(\hat{\beta}\delta b - p^*) - F(\beta\delta b - p^*)}{f(\beta\delta b - p^*)}.$$

For a consumer with standard *time-consistent* preferences ($\beta = \hat{\beta} = 1$), the right-hand side of (6) is zero, and the per-usage price p^* is set equal to the marginal cost a of the firm. The best policy for a firm facing an individual with perfect self-control and rational expectations is to align the incentives of the user with the cost for the firm. This guarantees that the agent undertakes C only if the investment generates positive surplus; that is, if $c \leq \delta b - a$.

For time-inconsistent users ($\beta < 1$), any optimal[6] per-usage price p^* lies below a. Below-marginal-cost pricing occurs for two distinct reasons. The first is a commitment rationale. An individ-

6. Program (2)–(3) may have multiple solutions, all of which satisfy the first-order condition (6) and yield the same profit level. For simplicity, we are going to assume uniqueness of the solution in what follows.

ual who is at least partially aware of the time inconsistency looks for ways to increase the probability of future investment. Choosing a contract with low p is one such way. The first term in (6) expresses this rationale. The firm lowers p^* below a to the extent that the user is conscious about her future *time inconsistency,* as measured by $1 - \hat{\beta}$. The term $(1 - \hat{\beta})\delta b$ is the additional utility that the time 0 self places on consumption C, relative to the (anticipated) utility of the time 1 self. It represents the value of commitment for an additional unit of consumption.

The second reason for pricing below marginal cost is consumer overconfidence. The firm knows that a (partially) naive user overestimates future consumption. Therefore, it offers a contract with a discount on p and an increase in L relative to the contract for a time-consistent agent. While the user would be indifferent between the two contracts if both were offered, the actual welfare is lower for the contract with a discount on p. The user will take advantage of the discount less often than she anticipates, and the firm will make higher profits. The firm sets p so as to increase this fictitious surplus, the second term in the profit function (4). To a first approximation, this fictitious surplus depends on $F(\hat{\beta}\delta b - p) - F(\beta\delta b - p)$, the overestimation of future consumption. Figure IIb illustrates this for a distribution F with a peak in the right tail. The shaded area corresponds to $F(\hat{\beta}\delta b - p) - F(\beta\delta b - p)$. The higher is the density f, the larger is this shaded area, and the higher are the profits from the overestimation of consumption. By choosing p so that the shaded area includes the peak, therefore, the firm maximizes the fictitious surplus. Condition ABP rules out peaks in the extreme right

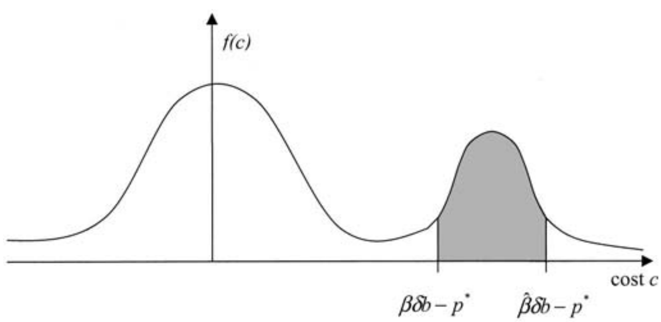

FIGURE IIb
Optimal Level of p: Pricing of Overconfidence

tail of the distribution (Figure IIa) which might lead the firm to set an arbitrarily negative p.

Summing up, in equilibrium both sophisticated and naive hyperbolic agents pay a lower per-usage price than time-consistent consumers. To the extent that the individual is aware of the time inconsistency, she prefers a low per-usage price as a commitment device that increases the usage in the future. To the extent that the consumer is overconfident about the strength of her future willpower, the firm exploits the misperception of the consumption probability by tilting the pricing toward the lump-sum fee L.

II.D. Leisure and Neutral Goods

The above model captures the case of activities with immediate benefits and delayed costs as well. Recall that C and NC are alternative ways to employ a given amount of resources (time and money). In contrast to the investment good case, we now consider the case in which C provides a higher payoff than NC at time 1 and a lower payoff at time 2. We call commodities with this front-loaded intertemporal profile *leisure goods*. Examples include goods that are harmful to future health (sweets, addictive goods), and goods that tempt the user to forgo more productive activities (gambling, cellular phone calls). With some abuse of notation, we employ the same variables as in the investment case (Figure I): the net payoff of C relative to NC is $-c$ in period 1 and b in period 2. Unlike in the investment case, however, $-c$ is the current benefit of C and $b < 0$ is the future cost. We otherwise maintain the same assumptions: the payoff $-c$ at time 1 (now, an immediate benefit) is stochastic, with F as the distribution of c, and the payoff b (now, a delayed cost) at time 2 is deterministic.

At time $t = 0$ the user desires to choose C in the future with probability $F(\delta b - p)$. At $t = 1$, however, she ends up choosing C with probability $F(\beta\delta b - p)$. Notice that for leisure goods time inconsistency leads to *overconsumption*: $F(\beta\delta b - p) > F(\delta b - p)$. A partially naive user is not fully aware of the future overconsumption. She expects to consume C with probability $F(\hat{\beta}\delta b - p) < F(\beta\delta b - p)$. She thus anticipates buying *fewer* leisure goods than she actually does. The difference between forecasted and actual consumption probability, $F(\hat{\beta}\delta b - p) - F(\beta\delta b - p) < 0$, is a measure of the *overconfidence*.

Similarly, we can deal with goods with no systematic intertemporal trade-off. We call *neutral goods* commodities C with future payoff b equal to 0. Although formally this category is a knife-edge case between the case $b > 0$ (investment) and the case $b < 0$ (leisure), we can think of it as an approximation for goods that do not challenge self-control. For neutral goods desired, perceived and actual consumption coincide.

COROLLARY 2 (Two-period model, leisure and neutral goods, monopoly). For $b \leq 0$, the profit-maximizing contract (L^*,p^*) exists. The per-usage price p^* equals marginal cost ($p^* = a$) for $\beta = 1$ or $b = 0$, and is set above marginal cost ($p^* > a$) for $\beta < 1$ and $b < 0$.

Proof of Corollary 2. The proof that the solution for p^* is interior follows along the lines of the proof of Proposition 1. Expression (6) implies that $p^* = a$ for $\beta = 1$ or $b = 0$, and $p^* > a$ for $\beta < 1$ and $b < 0$. QED

For a time-consistent user ($\beta = 1$), marginal-cost pricing aligns incentives for the consumer and costs of the firms. For a time-inconsistent user ($\beta < 1$), above-marginal-cost pricing occurs for both a commitment and an overconfidence reason. To the extent that the agent is sophisticated, the high per-unit cost is a commitment device designed to solve the overconsumption problem. To the extent that the agent is naive, above-marginal-cost pricing is aimed at exploiting the underestimation of the probability of a purchase. The pricing of neutral goods equals the pricing for time-consistent agents: given the absence of temptation, marginal-cost pricing is optimal.

II.E. Robustness

Perfect competition. We now consider the effect of competition on the profit-maximizing contract. In program (2)–(3) the degree of competition affects $\bar{u}$, the utility arising from the best alternative activity for the agent at time $t = 1$. Competing firms offer alternative contracts for the provision of the good C, and therefore raise $\bar{u}$. However, the solution for p^* in program (2)–(3) does not depend on $\bar{u}$. In expression (4) the reservation utility $\bar{u}$ is an additive constant. As a consequence, the results of below-marginal cost pricing for investment goods and above-marginal-cost pricing for leisure goods do not depend on the assumption of monopoly.

The equilibrium level of L, on the other hand, does depend on $\bar{u}$ and therefore on the market structure. Under perfect competition, $\bar{u}$ is determined so as to equate expected profits to 0. For a time-consistent individual ($\beta = 1$), zero profits and $p^* = a$ imply that L^* equals the setup cost K. For an individual with time-inconsistent preferences ($\beta < 1$), zero profits and $p^* \gtrless a$ imply that L^* is higher than K for investment goods and lower for leisure goods. The following Remark summarizes these results.

REMARK 1 (Two-period model, perfect competition). Under perfect competition, the profit-maximizing contract (L^*,p^*) exists. For $\beta = 1$ or $b = 0$, the per-usage price p^* equals marginal cost ($p^* = a$) and the lump-sum fee L^* equals the setup cost ($L^* = K$). For $\beta < 1$ and $b > 0$, the per-usage price is set below marginal cost ($p^* < a$), and the fee L^* is set above the setup cost ($L^* > K$). For $\beta < 1$ and $b < 0$, the price is above marginal cost ($p^* > a$), and the fee L^* is set below setup cost ($L^* < K$).

Certainty of costs. We can extend our results to the case where c is certain. Consider the investment good case and assume that the (monopolistic) firm and the consumer know c, and that the social surplus is positive, i.e., $c < \delta b - a$. If $\beta\delta b - a \leq c < \delta b - a$, the individual would not consume under marginal-cost pricing despite positive social surplus from consumption. In the optimum, the firm prices below marginal cost. If, instead, $c < \beta\delta b - a$, desired and actual consumption coincide for $p^* = a$, and the firm chooses any $p \leq \beta\delta b - c$. (See Appendix 1 for a proof.)

Timing. In the model the lump-sum fee L is paid at $t = 1$, one period after the contract is signed. If, alternatively, L is due at $t = 0$, a borrowing effect adds to the effects outlined above. Time-inconsistent consumers dislike the immediate payment, and firms offer lower ex ante fees and higher pricing per usage.

II.F. Welfare and Profits

In this subsection we analyze the effect of time inconsistency and overconfidence on the equilibrium level of consumption, consumer welfare, and firm profits. We also discuss briefly the generalization of these effects beyond the specific cases of time inconsistency and naiveté.

Consumption. Time-consistent agents consume if $c \leq \delta b - p$. Time-inconsistent agents have an additional impatience parameter $\beta < 1$ and consume if $c \leq \beta\delta b - p$. If the two agents face the same price p, therefore, the time-inconsistent agent consumes less of the investment good ($b > 0$) and more of the leisure good ($b < 0$) than the time-consistent agent.

This comparison, however, neglects firm behavior. In equilibrium firms best-respond to consumer preferences and adjust p^*. In a market with time-consistent consumers, firms offer marginal-cost pricing, and agents consume if $c \leq \delta b - p^* = \delta b - a$; that is, whenever the social surplus of consumption is positive. In a market with sophisticated time-inconsistent agents, firms set the price at its first-best level $p^*_{FB} \equiv a - (1 - \beta)\delta b$ (see expression (6)), and agents consume if $c \leq \beta\delta b - p^*_{FB} = \delta b - a$. Equilibrium consumption is therefore equal in the two markets, despite the time inconsistency of consumers in the second market. Thanks to the commitment device of below-marginal-cost pricing, time-inconsistent consumers are able to constrain their future selves to the consumption policy that is optimal for the self at time 0. They therefore attain the same efficient outcomes as if they did not suffer from limited self-control.

This result depends crucially on consumer sophistication. If agents are partially naive ($\beta < \hat{\beta}$), the equilibrium price p^* differs from the perfect commitment device (see (6)). Firms set prices so as to extract maximal profits from the overconfidence, and agents do not consume the first-best amount.

Consumer welfare and firm profits. Define the consumer welfare of a $(\beta,\hat{\beta},\delta)$ consumer as $U_j^{\beta,\hat{\beta},\delta} = \delta[-L^* + \int_{-\infty}^{\beta\delta b - p^*} (\delta b - p^* - c)\, dF(c)]$, with $j \in \{M,PC\}$ for the cases of *M*onopoly and *P*erfect *C*ompetition. This is the welfare evaluated with long-term preferences ($\beta = 1$), which coincides, up to a multiplicative constant β, with the welfare of the time 0 self.[7] For partially naive agents this utility differs from the one that they (mistakenly) expect to experience (left-hand side of expression (3)). After substituting for the profit-maximizing L^*, one obtains

$$U_j^{\beta,\hat{\beta},\delta} = \delta\left[\int_{-\infty}^{\beta\delta b - p^*} (\delta b - a - c)\, dF(c) - K\right] - \Pi_j^{\beta,\hat{\beta},\delta}, \tag{7}$$

where $\Pi_j^{\beta,\hat{\beta},\delta}$ is the profit of the firm facing a consumer with $(\beta,\hat{\beta},\delta)$ preferences. The monopoly profits $\Pi_M^{\beta,\hat{\beta},\delta}$ are given by expression (4) evaluated at $p = p^*$, and the perfect competition

7. See O'Donoghue and Rabin [2001] for a discussion of this welfare measure.

profits satisfy $\Pi_{PC}^{\beta,\hat{\beta},\delta} = 0$. Finally, we define the joint consumer-firm surplus as the sum of firm profits and the consumer welfare $U_j^{\beta,\hat{\beta},\delta}$. Formally,

$$(8) \quad S_j^{\beta,\hat{\beta},\delta} \equiv \Pi_j^{\beta,\hat{\beta},\delta} + U_j^{\beta,\hat{\beta},\delta} = \delta\left[\int_{-\infty}^{\beta\delta b - p^*} (\delta b - a - c)\, dF(c) - K\right].$$

Given that p^* coincides under monopoly and perfect competition, $S_M^{\beta,\hat{\beta},\delta}$ and $S_{PC}^{\beta,\hat{\beta},\delta}$ coincide. We therefore drop the subscript on $S^{\beta,\hat{\beta},\delta}$. Finally, we consider investment and leisure goods and neglect neutral goods ($b = 0$) for which time inconsistency and naiveté have no welfare effect.

We consider first the welfare effects of time inconsistency for sophisticated agents ($\beta = \hat{\beta} < 1$). After substituting $p_{FB}^* = a - (1 - \beta)\delta b$ into expression (7), it becomes apparent that welfare $U_j^{\beta,\hat{\beta},\delta}$ does not depend on $1 - \beta$. In equilibrium, welfare is unaffected by the strength of the conflict of preferences between the time 0 self, which signs the contract, and the later self, which makes the consumption decision. Similarly, equilibrium profit $\Pi_M^{\beta,\hat{\beta},\delta}$ and surplus $S^{\beta,\hat{\beta},\delta}$ are independent of the degree of time inconsistency. In the market, limitations to self-control do not affect welfare or profits, as long as the agents are sophisticated. The market provides agents with a contract that allows the time 0 self to achieve the first-best for any degree of time inconsistency.

We can then consider the welfare effects of naiveté, as measured by $\hat{\beta} - \beta$ for fixed β. Higher levels of naiveté imply more overconfidence about future consumption and a higher fictitious surplus $\int_{\beta\delta b - p^*}^{\hat{\beta}\delta b - p^*} (\delta b - p^* - c)\, dF(c)$ in (4). Not surprisingly, a monopolistic firm can extract more profits when the naiveté is higher, as the following Proposition (proved in Appendix 1) states.

PROPOSITION 3 (Two-period model, welfare).

- (i) For sophisticated agents, consumer welfare $U_M^{\beta,\beta,\delta}$ and $U_{PC}^{\beta,\beta,\delta}$, firm profits $\Pi_M^{\beta,\beta,\delta}$, and the joint consumer-firm surplus $S^{\beta,\hat{\beta},\delta}$ are unaffected by the degree of time inconsistency $(1 - \beta)$.
- (ii) For partially naive agents, profits $\Pi_M^{\beta,\hat{\beta},\delta}$ are strictly increasing in the degree of naiveté $\hat{\beta} - \beta$ for fixed β and $\hat{\beta} < 1$.
- (iii) Partially naive agents have lower surplus and consumer welfare than sophisticated agents with the same β and δ:

$S^{\beta,\hat{\beta},\delta} \leq S^{\beta,\beta,\delta}$ and $U_j^{\beta,\hat{\beta},\delta} \leq U_j^{\beta,\beta,\delta}$ for $\beta < \hat{\beta} \leq 1, j \in \{M,PC\}$.

(iv) The loss in consumer welfare due to monopoly power, $U_M^{\beta,\hat{\beta},\delta} - U_{PC}^{\beta,\hat{\beta},\delta}$, becomes larger in absolute value as naiveté $\hat{\beta} - \beta$ increases for fixed β: $dU_M^{\beta,\hat{\beta},\delta}/d\hat{\beta} - dU_{PC}^{\beta,\hat{\beta},\delta}/d\hat{\beta} < 0$ for $\hat{\beta} < 1$.

Proof of Proposition 3. In Appendix 1. QED

Proposition 3(i) summarizes the results of changes in β for sophisticated agents. Proposition 3(ii) states the effects of naiveté on monopoly profits. Partially naive agents are like people with a "fictitious demand curve." A monopolistic firm can get them to pay for a fictitious product that they will never actually consume. Increases in naiveté increase the fictitious surplus and therefore the monopolistic profits.

This increase occurs through two effects, a *distributional* and an *efficiency* effect. First, the overestimation of future surplus changes the welfare distribution. For any given price p, a naive agent is willing to accept a higher fee L than a sophisticated agent, since she incorrectly forecasts a higher surplus. Therefore, the monopolistic firm can charge a higher fee L to a naive agent than to a sophisticated agent and appropriate a larger share of the actual surplus. Second, the overestimation affects efficiency. The monopolistic firm sets the price p not in order to maximize the actual surplus S, but rather to maximize the profits from the overestimation of the surplus. Since the firm deviates from the first-best pricing p^*_{FB}, the joint consumer-firm surplus S is lower under naiveté than under sophistication (Proposition 3(iii)). Because of the combination of the distributional and the efficiency effects, a partially naive consumer achieves lower welfare than a sophisticated consumer with the same discounting parameters β and δ.

While the qualitative effects of naiveté on consumer welfare do not depend on market power, the magnitudes do, as summarized in Proposition 3(iv). Under perfect competition, the efficiency effect is the same as under monopoly. However, unlike under monopoly, there is no distributional effect of naiveté. Under perfect competition, the individual, rather than the firm, is the residual claimant of the profits from naiveté. Therefore, the detrimental effect of increases in naiveté on consumer welfare, captured by $dU^{\beta,\hat{\beta},\delta}/d\hat{\beta}$, is higher under monopoly. Stated differently, the loss in consumer welfare due to monopoly power, $U_M^{\beta,\hat{\beta},\delta} - U_{PC}^{\beta,\hat{\beta},\delta}$, becomes larger in absolute value as naiveté increases.

Pareto increases in welfare. So far, we have shown that, for naive agents, the profit-maximizing two-part tariff does not maximize the discounted sum of firm profits and consumer surplus as of period 0, $\Pi_j^{\beta,\hat{\beta},\delta} + U_j^{\beta,\hat{\beta},\delta}$. An alternative welfare criterion proposed in the literature takes into account the welfare of all selves of the consumers and not just the welfare of the time 0 self. In Appendix 1 we consider the case of investment goods. We show that there exists no Pareto-improving two-part tariff $(\tilde{L},\tilde{p})$ for sophisticated consumers (the equivalent of Proposition 3(i)). For naive users, instead, as long as $p^* < p_{FB} = a - (1 - \beta)\delta b$, there exists a set of prices $(\tilde{L},\tilde{p})$ that, compared with (L^*,p^*), increases the (actual) welfare of both period 0 and period 1 selves, while holding constant the discounted profits for the firm (the equivalent of Proposition 3(ii) and (iii)). In the case $p^* > p_{FB}$, there is no Pareto improvement possible for naive agents. While the self at time 0 would like to decrease p to p_{FB}, the self at time 1 would like to increase p up to a. The case of leisure goods is symmetric.

Generalization. The above results suggest a clear distinction between nonstandard preferences (time inconsistency) and nonrational expectations (naiveté). The first type of deviation does not necessarily affect surplus, profits, and welfare. Firms offer the contract that maximizes the joint surplus, and this contract may neutralize the behavioral effect of the nonstandard feature. This result generalizes beyond the specific application of time inconsistency. For example, if consumers have limited computational abilities and are aware of it, firms may offer simple contracts, or devices that help consumers perform the computations. In equilibrium the limited computational power is likely to be nonbinding.

This conclusion changes if consumers have nonrational expectations. If consumers misperceive their objective function or their constraints, the firms do not offer the surplus-maximizing contract, but rather a contract that accentuates the effects of the nonrational expectations so as to increase profits. As an example, consider again consumers with limited computational power, but now assume that they are unaware of their limits. The firms are likely to offer computationally complex contracts that induce the consumers to make choices that are profitable for the firms, but not surplus-maximizing. Therefore, the computational limitation will affect equilibrium consumption, consumer welfare, and profits. Under monopoly, the firm keeps the additional profits from contract distortions, while under perfect competition consumers themselves receive the "returns" to naiveté. Competition therefore tempers the adverse effects of naiveté on consumer welfare.

II.G. Contract Regulation

In this subsection we examine potential interventions of a benevolent government. Assume that the government can regulate the contracts by choosing either the per-usage pricing p or the flat fee L. Consider the case of a government that attempts to maximize the joint consumer-firm surplus. If consumers are sophisticated ($\beta = \hat{\beta} \leq 1$), there is no scope for government intervention since the market pricing coincides with first-best pricing p^*_{FB}. If consumers are partially naive ($\beta < \hat{\beta} \leq 1$), instead, government intervention could be beneficial, at least in principle. The derivative of $S^{\beta,\hat{\beta},\delta}$ with respect to p equals $-(p - p^*_{FB}) f(\beta\delta b - p)$. Therefore, the government can increase consumer surplus to the extent that it brings the market price p^* closer to the first-best price p^*_{FB}, that is, the perfect commitment device that solves the self-control problem. Unfortunately, the government cannot easily gauge the direction of adjustment ($p^* - p^*_{FB}$). Whenever the market price p^* is set below (above) marginal cost to take advantage of naiveté, the first-best price p^*_{FB} is also below (above) marginal cost. In order to estimate the first-best price p^*_{FB}, the government needs information on both consumer preferences (the parameters β, δ, b) and the production function (the cost a). Simpler policies may harm consumer welfare. For example, consider the restriction to no initial fees ($L = 0$) for investment goods ($b > 0$). This restriction induces above-marginal-cost pricing in the presence of setup costs, while the first-best pricing p^*_{FB} requires below-marginal-cost pricing.

If the government, instead, attempts to maximize consumer welfare with a choice of p, the welfare-enhancing intervention depends crucially on the market structure. Under perfect competition, the consumer welfare coincides with the joint surplus S, and the policy of setting $p = p^*_{FB}$ is still optimal. However, under monopoly, consumer welfare is

$$U_M^{\beta,\hat{\beta},\delta} = \beta\delta\left[\bar{u} - \int_{\beta\delta b - p}^{\hat{\beta}\delta b - p} (\delta b - p - c)\, dF(c)\right]. \tag{9}$$

The agent achieves the reservation utility ($\beta\delta\bar{u}$) minus the fictitious surplus. Since the fictitious surplus depends on the per-usage price p, the government can manipulate the overestimation by choosing an appropriate value of p. The price that maximizes expression (9), i.e., that minimizes the fictitious surplus, is unlikely to coincide with the first-best price p^*_{FB}.

To sum up, if consumers are sophisticated, the market interaction leads to the socially optimal outcome, and there is no scope for intervention of a benevolent government. If consumers do not have rational expectations, instead, government intervention can in principle be beneficial. The intervention, however, is subject to several limits: it requires extensive information that the government is unlikely to have; it depends on the market structure; it may not fully remedy the adverse effects of naiveté. A better policy for the government, in general, is to educate partially naive users and make them aware of their naiveté.

III. Evidence on Contracts

The model has two testable implications about contractual features for time-inconsistent agents. For investment activities, it predicts pricing below marginal cost and a lump-sum fee larger than the setup cost. For leisure goods, it predicts pricing above marginal cost and (under competition) a lump-sum fee smaller than the setup cost. We compare these predictions with empirical evidence on contract design in industries for investment and leisure goods. We select industries in which firms have access to individual consumption data and can therefore price per usage. While the predictions of the model so far do not distinguish between sophistication and naiveté, the naiveté explanation appears more plausible, particularly in the leisure good category. For each example we discuss leading alternative explanations.

III.A. Investment Goods

Health club industry. The revenue of the U. S. health club industry for the year 2000 totals $11.6 billion, and the memberships in this same period sum to 32.8 million. Most of the 16,983 clubs belong to regional companies.

Attendance in a health club is an activity with immediate effort costs and future health benefits. To document the contracts offered in the health club industry in an urban market, we conducted a telephone survey of all the clubs in Boston during the months April 2001–June 2002. We included all clubs listed in the Yellow Pages for the year 2000 in the metropolitan Boston area. Companies with several clubs in different locations were contacted only once. We eliminated 38 companies that had gone out of business, changed telephone number, merged with another company, or that served only professional athletes. The final sample included 64 companies, all of which provided the desired information. These companies operate 97 clubs in Boston. The survey, whose transcript is in Appendix 2, documents the menu of attendance options offered by the health clubs. For each club, we classify a contract as "frequent" if it is mentioned initially in answer to the question "which contracts are available at the health club," and as "infrequent" if it is only mentioned later in the phone conversation or in response to a specific inquiry.[8]

Table I shows the results weighted by company (columns (1) to (3)) and weighted by club (columns (4) to (6)). The health clubs in the Boston area offer three types of contracts. The *monthly contract* has an average initiation fee of $129 and an average monthly fee of $55. The *annual contract*[9] has an average initiation fee of $64 and an average fee for the year of $625. These numbers are somewhat higher when weighted by club. Neither contract charges per visit. Under the option of *payment per visit* (with single or ten-visit passes), users pay an average fee per visit of $11 ($12 if weighted by clubs).

The large majority of companies (columns (1) to (3)) offer all three contracts. However, contracts with no fee per visit are prevalent. The monthly contract is a frequent contract for a large fraction of the companies (40 out of 64), the annual contract for less than half of them (27 out of 64), and the pay-per-visit contract for only 2 companies out of 67. The results are similar in the sample of all the clubs (columns (4) to (6)).

For the clubs in DellaVigna and Malmendier [2003], we can estimate the marginal cost of a visit as the total variable cost for a month divided by the total number of visits in the month. The resulting figure of $5 includes the cost of providing towels, personnel, and replacing broken machines, but it excludes congestion effects, which are sizable at peak hours. Given the variation

8. We inquired about three specific types of contracts, in case they had not been mentioned: the monthly, the annual, and the pay-per-visit contract (details below).

9. The category includes contracts with a commitment of two or three years and display the annual equivalent of the overall fee.

TABLE I
HEALTH CLUB INDUSTRY IN BOSTON AREA—MENU OF CONTRACTS†

	Sample: One club per company			Sample: All clubs		
	Monthly contract (1)	Annual contract (2)	Pay-per-visit (3)	Monthly contract (4)	Annual contract (5)	Pay-per-visit (6)
Average fee [in $]:						
per visit			10.98 (4.18)			12.21 (5.12)
per month	54.74 (27.73)			56.06 (26.10)		
per year		624.80 (410.00)			645.70 (350.30)	
initiation fee	128.94 (118.92)	64.04 (111.33)		153.47 (106.02)	65.12 (98.69)	
Menu of contracts:						
No. of health clubs offering contract	54	57	50	87	90	82
No. of health clubs—Frequent contract	40	27	2	67	39	2
No. of health clubs—Infrequent contract	14	30	48	20	51	80
Cancellation procedure:						
Automatic renewal	50	15		83	20	
—cancel in person	50	15		83	20	
—cancel by letter	29 (12 certified)	14 (6 certified)		54 (25 certified)	19 (7 certified)	
—cancel by phone	7	3		7	3	
Automatic expiration	2	35		2	63	
Information not available	0	7		0	7	
Number of observations	N = 64	N = 64	N = 64	N = 97	N = 97	N = 97

† Standard deviations are in parentheses. This table summarizes the features of contracts offered by health clubs in the Boston metropolitan area. Information from a survey conducted by the authors (more details in subsection III.A, transcript in Appendix 2). The sample of companies contacted that fulfilled the requirements is 64. Accounting for companies with multiple clubs, the survey covers 97 different clubs. The sample "One club per company" includes only one observation from each company. The sample "All clubs" includes one observation per club. A contract is a Frequent Contract if the staff in the health club mentions the contract at the beginning of the phone interview. A contract is an Infrequent Contract if the staff mentions the contract later in the conversation or in response to specific inquiries. Fee per visit is amount due at each visit. Initiation fee is amount due at sign-up for monthly and annual contract.

in personnel and towel provision, the marginal cost in the industry is likely to lie between $3 and $6, excluding congestion costs. Health clubs therefore price attendance at least $3 to $6 below marginal cost. Under the assumption of uniform distribution for the costs, equation (6) then implies that $(1 - \beta)\delta b$ should equal $3–$6. For benefits b in the range $20–$100, this implies a calibrated β in the range .7–.95, comparable to the estimates in the literature [Angeletos et al. 2001; Paserman 2003].

Additional support for the hypothesis of time inconsistency comes from the behavior of consumers in health clubs. DellaVigna and Malmendier [2003] show that consumers who pick monthly or annual contracts would on average have saved money paying per visit, a behavior consistent with demand for commitment or overestimation of attendance.

Two alternative explanations of this pricing scheme are price discrimination and transaction costs of charging per visit. Discriminatory pricing aims at extracting more surplus from users with higher willingness to pay. However, under flat-rate contracts, frequent health club users pay the same overall fee as infrequent users, who presumably have a lower willingness to pay. Price discrimination would predict that frequent users should pay more.[10] Transaction costs, which can explain many instances of flat rates, are small in the health club industry. Almost all clubs keep track of attendance using electronic cards and could charge users for attendance at minimal extra cost. It may be the case, though, that transaction costs are psychological: consumers dislike being charged per unit [Loewenstein and Prelec 1998]. Below, we give an example of an investment good with nonzero below-marginal-cost pricing.

The history of contract design in the health club industry is also of interest. In the 1950s many health clubs operated under a pay-per-usage system. The clubs used coupons that the users could redeem at the entrance. This pricing scheme was largely abandoned by the 1970s in favor of flat-rate contracts. Pay-per-visit contracts remain uncommon today even though the introduction of electronic cards has reduced the cost of charging per visit. One interpretation of this history is that over time the firms have learned to design contracts that maximize profits given

10. Barro and Romer [1987] and Oi [1971] give conditions under which ski or Disneyland rides within a day may be charged a flat rate. They do not explain, though, the presence of season passes, the equivalent of a health club membership.

users with limited self-control. The model in this paper suggests that the next step in the industry may be the introduction of contracts with negative per-usage pricing.

Vacation time-sharing. Vacation time-share companies such as Resort Condominiums International (RCI) and Hapimag offer members the opportunity to book one or more weeks of holiday in different resorts each year. The estimated sales of time-shares in year 2000 amount to $7.5 billion worldwide. Booking a holiday is an activity with current effort costs—planning of holiday time, location, and logistics—and future benefits, the holiday itself. A contract with low price per week of holiday and high upfront fees appeals to time-inconsistent consumers—either because they demand a commitment device (sophisticated consumers) or because they overestimate the actual number of holidays they will book (naive consumers). The typical contract in the time-sharing industry involves a large initial fee (on average, $11,000 in the United States) to become member and only a small fee for each week of holiday used. RCI charges an exchange fee of $140.[11] An alternative explanation is that firms use this pricing scheme to raise funds for their real estate investments. This explanation, though, makes the unusual assumption of credit-constrained firms.

III.B. Leisure Goods

Credit cards. An easily accessible credit line has the intertemporal features of a leisure good: it allows credit-constrained individuals to increase current consumption at the expense of future consumption. Credit cards are an easy and widespread way to obtain credit. The average credit card debt amounts to over $5,000 per U. S. household [Gross and Souleles 2002; Laibson, Repetto, and Tobacman forthcoming].

We expect naive individuals to underestimate the usage of the credit line. Credit card companies should respond by charging an interest rate above marginal cost together with a low initial fee or even offer a bonus. Table II reports representative credit card offers from major U. S. issuers.[12] Most firms charge an

11. Amended Annual Report 10-K/A by Cendant Corporation for year 2000.
12. The information comes from each issuer's Web site. We selected the largest six issuers ranked by outstanding balances in 1997, excluding First Chicago NBD because of its merger with BankOne/FirstUSA [Evans and Schmalensee 1999, p. 229]. We include information on Providian (twelfth largest issuer) and CapitalOne (eighth largest issuer) since they provide information on their menu of credit cards. We add Discover and American Express since they are the biggest issuers outside Visa and Mastercard.

TABLE II
CREDIT CARD INDUSTRY—REPRESENTATIVE CONTRACTS†

	Type of credit card offer (1)	Regular interst rate (APR) (2)	Annual fee in $ (3)	Benefits (4)	Introductory interest rate (APR) (5)	Length of introductory offer (6)
Citibank	Platinum Select Visa	Prime + 12.99%	0		2.90%*	9 months
MBNA	Platinum Plus Visa	12.99%	0		3.90%*	6 months
First USA	Platinum Visa	Prime + 6.50%	0		9.90%*	9 months
Chase Manhattan	Wal-Mart Mastercard	Prime + 3.98% to Prime + 11.98%	0		0%	6 months
Bank of America	Visa Gold	Prime + 7.99% to Prime + 12.99%	0		3.90%	6 months
Household Bank	GM Mastercard	Prime + 9.99%	0	5% toward GM	2.90%	6 months
Providian	Visa Platinum	Prime + 3.24%	0		0%	3 months
	Visa Gold Prestige	Prime + 10.24%	0		0%	2 months
	Visa Gold Preferred	Prime + 13.24%	0		0%	2 months
	Visa Classic	Prime + 17.24%	0-59-89		0%	2 months
Capital One	Platinum Visa	9.90%	0		N/A	N/A
	Gold Visa	14.90%	0		2.90%*	6 months
	Classic Visa	19.80%	49		N/A	N/A
Discover	Platinum Card	13.99%	0	1% Cashback	1.70%*	6 months
American Express	Blue Credit Card	9.99%	0		0%	6 months
	Optima Credit Card	Prime + 7.99%	0		7.90%	6 months
	(Gold) Charge Card	N/A	55-75		N/A	N/A

† Information about typical credit card offers from the issuers' Web sites. The displayed issuers are the largest six issuers ranked by outstandings as of 1997, excluding First Chicago NBD that merged with BankOne/FirstUSA [Evans and Schmalensee 1999, p. 229]. We also include information on Providian (twelfth largest issuer) and CapitalOne (eighth largest issuer) because of the availability of information on the menu of cards offered. Finally, we included Discover and American Express that are the biggest issuers outside the Visa and Mastercard circle. We include also, for comparison, a charge card (last row). The regular interest rate is the APR on the outstanding balance for customers who pay regularly. Interest rate may depend on the credit history. Introductory APR (column (5)) is the interest rate on the outstanding balance for the introductory period (column (6)), after which the relevant interest rate becomes the one in column (2). Values with an asterisk (*) are credit cards for which the introductory offer applies only for Balance Transfers.

interest rate on outstanding balances that exceeds the prime rate by as much as 10 percentage points (column (2)). This high interest rate could reflect above-marginal-cost pricing or high default costs. Ausubel [1991] provides evidence in favor of the first interpretation: credit card debt resells on the private market at a 20 percent premium. This implies that the interest rate on credit card debt, net of default and operating costs, exceeds the cost of capital by 20 percent. Ausubel suggests that overconfidence about future borrowing may explain the high rates of interest. Our model embeds this prediction in a general theory of leisure good pricing for individuals with time-inconsistent preferences and naiveté.

Despite sizable costs, net of the interchange fee, of keeping a credit card account,[13] the issuers typically require no annual fee (column (3)), and offer extra benefits such as car rental insurance and cash back (column (4)). As a result, credit card companies experience losses on users with no outstanding balances, who can use the card for transactions, enjoy the extras, and borrow for up to 30 days at no cost.[14]

Las Vegas pricing. Psychological and pharmacological evidence [Comings et al. 1996; Wray and Dickerson 1981] suggests that gambling is addictive for a substantial portion of the active gamblers. Addictive goods [Becker and Murphy 1988] involve immediate benefits and future costs, since current consumption lowers future utility. Las Vegas hotels, which integrate gambling with accommodation and dining, charge very competitive rates on hotel rooms and in their buffets. They make up the difference in profits from gambling. These deals are attractive to naive agents who expect to gamble moderately. An alternative explanation is that high-gamblers have a higher willingness to pay than low-gamblers. The cartel of Las Vegas hotels uses the two-part tariff to implement discriminatory pricing.

13. The operating expenses, net of the revenues from interchange fees, averaged 3.4 percent of outstanding balances in 1997 (calculation of the authors using data from Evans and Schmalensee [1999, p. 249]). These expenses do not include borrowing and default costs.

14. Differently from credit cards, charge cards require the balance to be paid in full within 30 days. These cards, therefore, do not have interest rate revenue. Our theory thus predicts that these cards would not be able to offer the same bonus that credit cards offer. As the last row in Table I shows, the American Express charge card, the most common in the United States, has a substantial annual fee of $55 or $75 (if Gold).

TABLE III
MOBILE PHONE INDUSTRY: MENU OF CONTRACTS†

	Revenues in year 2000 [$m] (1)	Monthly allowance (2)	Monthly fee in $ (3)	Average price per minute in ¢ (4)	Price of additional minutes in ¢ (5)	(5)/(4) (6)
AT&T	7,627[a]	450	59.99	13.3	35	2.63
		650	79.99	12.3	35	2.84
		900	99.99	11.1	25	2.25
		1,100	119.99	10.9	25	2.29
		1,500	149.99	10.0	25	2.5
		2,000	199.99	10.0	25	2.5
Sprint PCS	6,341	20	19.99	100.0	40	.40
		200	34.99	17.5	40	2.28
		350	39.99	11.4	40	3.50
		450	49.99	11.1	40	3.60
		1,000	74.99	7.5	40	5.33
Verizon	14,236	150	35	23.3	40	1.71
		400	55	13.7	35	2.55
		600	75	12.5	35	2.80
		900	100	11.1	25	2.25
		1500	150	10.0	25	2.50
		2000	200	10.0	20	2.00
		3000	300	10.0	20	2.00

† Information is from the Web site of the companies. Allowances and rates are for calls for plans with domestic long distance included. For Sprint, the allowance applies only to daytime, weekday calls. For Sprint, additional minutes for evening and weekend calls are not included in the computations. Annual Revenue is from 10K filing and refers to the cellular phone business exclusively. Monthly allowance is the total number of monthly minutes that the consumer can use without incurring any extra charge. Average price per minute is ratio of columns (3) and (2). Price of additional minutes is cost per minute of each call beyond the monthly allowance.

a. This is the revenue in year 1999, since the data from year 2000 were not available.

Cellular phones. As of year 2000, the mobile phone industry had $52.5 billion revenues and 109.5 million subscribers. Cellular phones are convenient communication tools, but also fashionable gadgets. They tempt users with limited self-control to spend time on the telephone rather than on other more productive activities. Naive users underestimate the number of future calls when they choose the monthly airtime. Cellular phone companies can extract profits from naiveté by setting high marginal prices for minutes beyond the monthly allowance. In the typical contract, the consumers choose a monthly airtime allowance within a menu. Table III shows the annual revenue in millions of dollars (column (1)), the menu of allowances (column (2)) and the monthly fee (column (3)) for three major cellular phone compa-

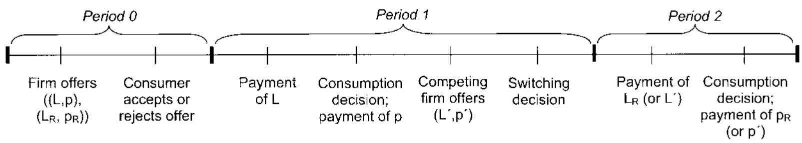

FIGURE III
Timing of Model with Renewal

nies.[15] The marginal price for minutes beyond the allowance (column (5)) is typically two to four times higher than the average price of a call within the limit (column (6)). The model with naive consumers does not explain the zero marginal price for minutes within the airtime allowance, which may be due to distaste of payment per usage. Models with heterogeneity in the value of time also yield a menu of contracts, but may not easily explain why within a contract the per-minute price increases with the number of minutes, but across contracts the per-minute price decreases with the number of minutes.

IV. SWITCHING COST MODEL

In many industries, including credit cards, mail order, newspapers, and utilities, firms offer contracts that are automatically renewed. Consumers have to incur a transaction cost to switch to a competitor or cancel. To capture this case, we extend the model of Section II to a three-period setting with switching costs. We allow for ex ante and ex post competition, as in the rational switching cost models [Farrell and Klemperer forthcoming].

IV.A. The Setting

Timing. Figure III illustrates the timing of the extended model. At the end of period 1 the agent can renew the contract for one more period. If she renews, she is entitled to choose either C or NC in period 2. Alternatively, she can pay an effort cost $k \geq 0$ at time 1 and switch to a competing firm offering a similar choice between C and NC at time 2.

15. The allowances and rates are for daytime, weekday minutes for plans with domestic long distance included; additional weekend and night-time minutes (where available) are not included.

Consumers. Consumption C provides an immediate payoff $-c_t$ at time t and a delayed payoff b at time $t + 1$, with $t = 1, 2$. We analyze the cases of *investment goods* ($b > 0$), *leisure goods* ($b < 0$), and *neutral goods* ($b = 0$). We assume that the costs c_1 and c_2 are i.i.d. and drawn from a known distribution F. In each period, consumers observe their realization of c_t before choosing C or NC. The payoff b is known ex ante.

Firms. Firms incur a setup cost $K \geq 0$ for enrollment at $t = 1$ and a per-unit cost $a \geq 0$ whenever an agent chooses C. If a consumer switches to a competing firm at $t = 1$, this firm also pays a setup cost K at $t = 2$. The contract offered at $t = 0$ is a generalized two-part tariff $((L,p), (L_R,p_R))$.[16] The new contractual elements are L_R, a renewal fee due at $t = 2$ if the agent *R*enews, and p_R, the per-usage price at $t = 2$ after *R*enewal. The contract offered by competing firms at $t = 1$ is a two-part tariff (L',p'). We assume that firms commit at $t = 0$ to the contracts offered at $t = 1$ and $t = 2$.

IV.B. Consumer Behavior

The new element of consumer behavior is the renewal decision at the end of period 1. Renewal occurs if

$$\text{(10)} \quad \beta\delta\left[-L_R + \int_{-\infty}^{\hat{\beta}\delta b - p_R} (\delta b - p_R - c)\, dF(c)\right] \geq -k + \beta\delta\left[-L' + \int_{-\infty}^{\hat{\beta}\delta b - p'} (\delta b - p' - c)\, dF(c)\right].$$

(Notice the β outside the brackets and the $\hat{\beta}$ in the integral.) Define as R the set of prices (L_R,p_R) such that renewal occurs: $R = \{(L_R,p_R) | \text{expression (10) holds}\}$. For given $\hat{\beta}$, this set is (weakly) decreasing in β. Less self-control (lower β) is associated with more renewal.

At time 0, when signing the contract, the agent expects to renew at $t = 1$ if

16. This contract is the most general contract under the restriction that the firms cannot condition the contract on observed consumer choices. The firms may want to condition on the choices under the case of naiveté.

$$(11)\quad \hat{\beta}\delta\left[-L_R + \int_{-\infty}^{\hat{\beta}\delta b - p_R} (\delta b - p_R - c)\, dF(c)\right] \geq -k + \hat{\beta}\delta\left[-L' + \int_{-\infty}^{\hat{\beta}\delta b - p'} (\delta b - p' - c)\, dF(c)\right].$$

(Notice the $\hat{\beta}$ also outside the brackets.) Define $\hat{R}$ as the set of values of the prices (L_R, p_R) such that the consumer expects to renew: $\hat{R} = \{(L_R, p_R) | \text{expression (11) holds}\}$. It is easy to see that $\hat{R} = R$ for $\hat{\beta} = \beta$ or $k = 0$, while $\hat{R} \subset R$ for $\hat{\beta} > \beta$ and $k > 0$. Whereas time-consistent and sophisticated individuals $(\beta = \hat{\beta})$ correctly anticipate the renewal rate, partially naive agents $(\beta < \hat{\beta})$ underestimate it. This is not surprising: partially naive individuals overestimate the probability of undertaking activities with current costs and delayed benefits, such as switching contracts.

IV.C. Firm Behavior

We solve the maximization problem starting from the last period. At $t = 1$, the competing firms determine the two-part tariff (L', p') as in the simple model of Section II. The profit-maximizing price p'^* is given by expression (6), and the lump-sum fee L'^* equals $K - F(\beta\delta b - p'^*)(p'^* - a)$ by the zero-profit condition.[17] We assume that the consumer is aware of this contract at $t = 0$.

At $t = 0$, firms choose the two-part tariff $((L^*, p^*), (L_R^*, p_R^*))$ that maximizes the (perceived) consumer surplus subject to a zero-profit condition:

$$(12)\quad \max_{L, L_R, p, p_R} \beta\delta\left[\begin{array}{l} -L + \int_{-\infty}^{\hat{\beta}\delta b - p} (\delta b - p - c)\, dF(c) \\ \quad + \mathbf{1}_{\{(L_R, p_R) \in \hat{R}\}}\delta\left(-L_R + \int_{-\infty}^{\hat{\beta}\delta b - p_R} (\delta b - p_R - c)\, dF(c)\right) \\ \quad + \mathbf{1}_{\{(L_R, p_R) \notin \hat{R}\}}\delta\left(-\frac{k}{\delta} - L'^* + \int_{-\infty}^{\hat{\beta}\delta b - p'^*} (\delta b - p'^* - c)\, dF(c)\right) \end{array}\right]$$

17. In equilibrium, the condition of nonnegative profits is binding since firms could otherwise reduce L and increase the perceived consumer surplus.

subject to

$$(13) \quad \delta\left[\begin{array}{l} L - K + \displaystyle\int_{-\infty}^{\beta\delta b - p} (p - a)\, dF(c) \\ + \mathbf{1}_{\{(L_R, p_R)\in R\}}\delta\left(L_R + \displaystyle\int_{-\infty}^{\beta\delta b - p_R} (p_R - a)\, dF(c)\right) \end{array}\right] = 0.$$

Equations (12) and (13) show that the contract design takes into account two distortions in the beliefs of partially naive agents. First, as in the simple model, consumers overestimate (underestimate) the consumption of investment (leisure) goods. Second, naive agents underestimate the renewal rate, as is apparent in the difference between R and $\hat{R}$. The following proposition summarizes the main features of the solution[18] for exponential, sophisticated, and partially naive agents.

PROPOSITION 4 (Model with renewal, competition). The optimal contract $((L^*, p^*), (L_R^*, p_R^*))$ has the following features:

(i) The per-usage prices p^* and p_R^* both equal p'^* and are implicitly defined by expression (6).
(ii) For $\hat{\beta} = \beta$ or $k = 0$, the renewal fee L_R^* is any L_R with $L_R \leq L'^* + k/\beta\delta$. For $\hat{\beta} > \beta$ and $k > 0$, the renewal fee L_R^* equals $L'^* + k/\beta\delta$. In both cases, the fee L^* equals $K - (1 + \delta)F(\beta\delta b - p^*)(p^* - a) - \delta L_R^*$.
(iii) The difference between the two lump-sum fees, $L_R^* - L^*$, is equal to $\delta K + k(1/\beta\delta + 1/\beta)$ for $\hat{\beta} > \beta$ and $k > 0$ and is (weakly) smaller than $\delta K + k(1/\beta\delta + 1/\beta)$ for $\hat{\beta} = \beta$ or $k = 0$.

Proof of Proposition 4. In Appendix 1. QED

Proposition 4(i) points out that the choice of the per-usage prices[19] p^* and p_R^* is driven only by static forces: the alignment of consumer incentives with the marginal cost of the firm (time-consistent agents), the provision of commitment devices (sophisticated agents), and response to overconfidence (naive agents). For users with time-inconsistent preferences we thus get the

18. As in subsection II.C, we assume the existence of a market for the good, that is, of a contract that satisfies (13) and guarantees perceived consumer surplus of at least $\beta\delta\bar{u}$.

19. As above, we assume uniqueness of the solutions for p^* and p_R^* in problem (12)–(13).

familiar result of below-marginal-cost pricing for investment goods, above-marginal-cost pricing for leisure goods and marginal-cost pricing for neutral goods (Proposition 1 and Corollary 2).

Proposition 4(ii) and (iii) present the novel results of the model with switching costs. For fully sophisticated consumers, L_R^* is set low enough so that the consumers renew (see inequality (10)), and no agent bears the effort cost k, which is a net waste from the point of view of the two parties. For partially naive agents ($\hat{\beta} > \beta$) and $k > 0$, instead, the firm chooses L_R^* to extract maximal revenue from the underestimation of renewal. The equilibrium fee $L_R^* = L'^* + k/\beta\delta$ is the highest renewal fee such that naive users expect to switch to a different firm (inequality (11)), but end up renewing (inequality (10) reversed). Consumers anticipate quitting, expect to pay k at $t = 1$ and L'^* at $t = 2$, and require ex ante a compensation of $\delta^2 L'^* + \delta k$ for switching. At $t = 1$, however, they renew and pay L_R^* to the firm at $t = 2$. The firm makes a net gain as of time 0 of $\delta^2(L_R^* - L'^* - k/\delta)$, which is positive since $L_R^* = L'^* + k/\beta\delta > L'^* + k/\delta$.

Proposition 4(iii) characterizes the relative magnitude of fees in periods 1 and 2. If consumers are partially naive and switching costs are positive, the fees are higher in the second period. For example, for the benchmark case of neutral goods ($b = 0$) with no setup costs, the firms set $L_R^* = k/\beta\delta$ and $L^* = -k/\beta$; that is, the firms charge a renewal fee and offer an initial bonus. The firms back-load the fees because consumers naively expect to switch and not to pay the later fee. The back-loading of the fee is higher, the higher the self-control problem $(1 - \beta)$. If consumers have rational expectations, the fees will be (weakly) less back-loaded than in the case of naive individuals.

Endogenous cancellation costs. So far, we derived the optimal contract for a given level of the switching cost k. We now allow the firms to choose the profit-maximizing cost k^*. We assume that the firm can increase the transaction cost k of cancellation from a minimum level of 0 up to a maximum level of $\bar{k}$. We interpret $k = 0$ as a case in which renewal does not happen by default. Rather, consumers have to be asked whether they are interested in contract renewal. We interpret a high level of k as a case of automatic renewal with costly switching procedures, such as the request of additional documents or of in-person cancellation. The following proposition (proved in Appendix 1) summarizes the result.

PROPOSITION 5 (Model with renewal, endogenous cancellation cost). For time-consistent and sophisticated agents ($\beta = \hat{\beta} \leq 1$), k^* is any $k \in [0, \bar{k}]$. For naive agents ($\beta < \hat{\beta} = 1$), $k^* = \bar{k}$.

If the agents are time-consistent or sophisticated, the optimal level of k is indeterminant because no cancellation takes place in equilibrium and the agents are aware of it. If the agents are partially naive, instead, the firms strictly prefer contracts with automatic renewal and high cancellation costs k. A higher k allows the firm to charge a higher fee L_R, and therefore to obtain more revenue from the underestimation of renewal.

Partial naiveté. Propositions 4 and 5 show that the profit-maximizing contract is discontinuous in the degree of naiveté. While for sophisticated agents any renewal fee L_R low enough to induce renewal is optimal, for partially naive agents the renewal fee is set to the level that maximizes the revenue from underestimation of switching. While the firm has no incentive to generate transaction costs k for sophisticated agents, it does so for partially naive agents. This difference holds even if the degree of naiveté is arbitrarily small, that is for $\hat{\beta}$ close to β.

This discontinuity in behavior illustrates a more general point on the impact of markets on the relevance of biases. In the absence of firms, as long as the agents themselves select the tasks they perform, small deviations from rational expectations generate unnoticeable differences in behavior and welfare. In a market setting this conclusion need not hold. The firms, aware of the consumer deviations from rational expectations, offer contracts that are explicitly designed to target the deviations, no matter how small. Consumers, therefore, face selected tasks that systematically magnify the effect of their biases.

IV.D. Robustness

Monopoly.[20] So far, we have assumed competition both ex ante (at $t = 0$) and ex post (at $t = 1$). We now consider the assumption of monopoly. The monopolist offers a two-period contract $((L,p), (L_R,p_R))$ to the agents. At $t = 1$, agents can cancel the contract and consume the alternative good NC, which yields 0 utility. Under this market structure, we obtain qualitatively the same results on the profit-maximizing pricing, with different

20. The proofs of the two results in this subsection are available from the authors upon request.

magnitudes for the lump-sum fees. For partially naive agents, the renewal fee L_R^* equals $\int_{-\infty}^{\hat{\beta}\delta b-p^*} (\delta b - p^* - c)\, dF(c) + k/\beta\delta$, and the initial fee L^* equals $\int_{-\infty}^{\hat{\beta}\delta b-p^*} (\delta b - p^* - c)\, dF(c) - k - \bar{u}$. Given $\bar{u} > 0$, the fees are back-loaded as in the competition case. For sophisticated and exponential agents, we obtain

$$L_R^* \leq \int_{-\infty}^{\hat{\beta}\delta b-p^*} (\delta b - p^* - c)\, dF(c) + k/\beta\delta$$

and

$$L^* = (1 + \delta) \int_{-\infty}^{\hat{\beta}\delta b-p^*} (\delta b - p^* - c)\, dF(c) - \delta L_R^* - \bar{u}.$$

Uncertainty. In this subsection we have maintained the assumption that the realization of costs c_t is i.i.d. across periods. Alternatively, we can assume $c_1 = c_2$, that is, the realization of costs in the first period carries over to the second period, with no ex post competition. This alternative assumption yields similar results: if consumers are naive, firms set the renewal fee L_R^* at least as high as $k/\beta\delta$ and have an incentive to increase k. In addition, in this alternative framework, the renewal fee is high enough that, for some realizations of costs, naive consumers may cancel. In this world, therefore, high cancellation costs have a real inefficiency effect.

IV.E. Welfare and Profits

Equilibrium consumption in the model with renewal occurs for the same realizations of c as in the two-period model. Since all types of agents renew the contract at the end of period 1, there is no effect of the switching option on actual consumption. As in the simple model, therefore, consumption is at the first-best level for sophisticated consumers and is suboptimal for naive consumers. The joint consumer-firm surplus in the model with switching costs equals $\delta[(1 + \delta) \int_{-\infty}^{\beta\delta b-p^*} (\delta b - a - c)\, dF(c) - K]$ (see Appendix 1). Under perfect competition, therefore, the per-period consumer welfare is essentially the same as in the two-period model. Since no consumer actually bears the switching cost k, there are no efficiency costs associated with naiveté, beyond the ones due to the pricing of p.

Under monopoly, on the other hand, switching costs matter for welfare if consumers are partially naive ($\beta < \hat{\beta}$): the monopolist can use switching costs to transfer surplus from consumer welfare into firm profits. The consumer welfare is given by $\delta[\bar{u} - (1 + \delta) \int_{\beta\delta b - p^*}^{\hat{\beta}\delta b - p^*} (\delta b - p^* - c)\, dF(c) - k(1 - 1/\beta)]$, which is decreasing in k for $\beta < 1$. In fact, if the firm could charge infinitely high switching cost, it could in principle extract an infinite amount of surplus from the consumer. Introducing competition into the market is enough to eliminate this extreme redistributional effect of naiveté.

IV.F. Heterogeneity

So far, we have assumed a homogeneous population of consumers. While a full analysis of the case with heterogeneous consumers is beyond the scope of this paper, we briefly discuss the impact of a simple form of heterogeneity.[21] Assume that a fraction λ of consumers has time preferences $(\beta_0,\hat{\beta},\delta)$ and the remaining fraction $(1 - \lambda)$ has time preferences $(\beta_1,\hat{\beta},\delta)$, with $\beta_0 > \beta_1$. The two groups of consumers, therefore, have different short-run time preferences β but share the same expectation about the future, $\hat{\beta}$; the second group of consumers is more naive. An example is the case of a mixed population of time-consistent consumers ($\beta_0 = \hat{\beta} = 1$) and fully naive time-inconsistent consumers ($\beta_1 < \hat{\beta} = 1$).

Under this form of heterogeneity, the two groups of consumers pool on the same contract at time 0, since they expect to attain the same outcomes at times 1 and 2. For both groups, the perceived utility is given by (12), up to a multiplicative constant. Unlike in problem (12)–(13), however, the firm profits are given by a convex combination of the profits for the two groups. The actual consumption of the two groups differs, and firms take it into account in their contract choice.

Equilibria similar to the ones in Proposition 4 are still equilibrium candidates. If β_0 equals $\hat{\beta}$ (sophistication), the firms may charge $L_R^* \leq L'^* + k/\beta_0\delta$, so that all consumers expect to renew, and actually renew. If β_0 is smaller than $\hat{\beta}$ (partial naiveté), the firms may charge $L_R^* = L'^* + k/\beta_0\delta$, so that consumers expect to switch, but do not do so in equilibrium. In addition, a new type of equilibrium emerges in which firms charge $L_R^* = L'^* + k/\beta_1\delta$. If firms charge this higher fee, the fraction λ of consumers with β_0 preferences actually switches to a different company, while the remaining $(1 - \lambda)$ expects to switch, but does not. This is the

21. Details are available from the authors upon request.

optimal strategy for the firms when the fraction of consumers of the first type λ is small enough: the increased returns to overconfidence on the $(1 - \lambda)$ fraction outweigh the inefficiency of cancellation for the λ fraction.

Heterogeneity affects the welfare conclusions of subsection IV.E. With a heterogeneous population of consumers, the switching costs have a real efficiency effect. In order to take advantage more fully of the naive consumers, firms may increase the renewal fees up to a level at which the less naive consumers decide to switch and pay the transaction costs. Naiveté therefore induces inefficient switching, in addition to inefficient consumption.

IV.G. Summary and Relation to Switching Cost Literature

We have shown that the results of the simple model for time-inconsistent agents replicate under dynamic competition and switching costs. Firms offer above-marginal-cost pricing for leisure goods and below-marginal-cost pricing for investment goods. The dynamic setting generates two novel predictions for the case of (partially) naive time-inconsistent consumers. First, firms back-load the lump-sum fees, possibly offering bonuses at sign-up. Second, firms prefer contracts with automatic renewal and endogenously generate cancellation costs k. The intuition is that naive agents underestimate the likelihood of renewal l, inducing firms to charge additional fees after renewal. The firms increase the switching costs because the higher are the switching costs, the higher are the renewal fees, and the more firms can exploit the underestimation of renewal.

The new predictions extend even to consumption goods with a flat intertemporal profile (neutral goods). Since switching requires immediate effort and yields future benefits in the form of savings, it is a form of investment good. Contracts with automatic renewal, therefore, bundle an investment good with the actual good being sold. Similar to the below-marginal cost pricing result in the simple model of Section II, consumers are offered comparative savings if they undertake this investment activity. Partially naive consumers, however, reap these savings less often than they expect to.

These results are robust to the degree of naiveté, the industry concentration, and the assumption about uncertainty. Even a small amount of nonrational expectations induces the firms to set renewal fees and create cancellation costs. Under monopoly, we obtain the same results on pricing. Finally, the results carry over if we assume that costs c remain constant after the first period rather than being i.i.d.

Previous literature on switching costs has shown that firms may back-load their fees and possibly offer an initial bonus even in the presence of standard preferences and rational expectations [Farrell and Klemperer forthcoming]. These models assume that firms cannot commit to future contracts. The consumers anticipate that the firms will take advantage of them ex post and demand an initial bonus. A back-loaded pattern of fees ensues. Under commitment, however, these models make no prediction of back-loaded fees.

Differently from this literature, our model predicts back-loaded fees (with naive agents) even when firms can commit to future contracts.[22] Moreover, our model predicts that, in the case of naive agents, the firms strictly prefer to introduce switching costs, a result that, to our knowledge, does not appear in the standard switching cost literature.

V. Evidence on Contracts with Switching Costs

We collect evidence from several industries on the design of renewable contracts with switching costs. This allows us to test for the prevalence of time inconsistency and naiveté.

Credit card industry. As Table II shows, the typical credit card contract features a low interest rate (column (5)) for an introductory period of typically six months (column (6)), followed by a high interest rate for the subsequent period (column (2)). The renewal after the introductory period is automatic. The heavily back-loaded structure of the charges is consistent with consumers underestimating renewal past the introductory period. Empirical evidence on consumer behavior confirms this interpretation. In a field experiment with randomized credit card offers, consumers appear to overrespond to introductory interest rates relative to postintroductory rates [Ausubel 1999].

Rational switching cost models could account for the initial bonus. However, in the credit card industry it is reasonable to assume that issuers can commit to future rates, in which case the rational switching cost models would not predict the back-loading

22. We obtain the standard results of the switching cost literature if we assume that firms cannot commit to future prices. In this case, back-loaded fees are the equilibrium prediction for all types of agents.

of the fees. In addition, rational switching cost models predict that the initial bonus should be no larger than the switching cost. In the case of credit cards, the bonus of a six-month zero-interest loan for a balance of $3,000 is approximately $150, assuming a 10 percent normal interest rate. The switching costs are likely to be one to two magnitudes smaller. The naive time-inconsistent model can generate a large bonus, given daily procrastination.

Alternatively, teaser rate offers may be due to asymmetric information about product quality. However, the services provided by the companies are standard and highly observable. Also, teaser rate offers induce "adverse" selection of rational consumers who use the credit card to borrow at the initial low rate, and switch to another card at the end of the introductory period. In a world without overestimation of switching, the prevalence of these offers would constitute a puzzle.

Health club industry. As column (4) in Table I shows, the monthly contract is a frequent contract in 67 out of the 97 health clubs in the Boston area. Overall, it is offered in 87 clubs. In 83 out of 87 clubs, this contract is automatically renewed from month to month, with an average monthly fee of $56. Most firms generate additional switching costs. While all the 83 clubs allow in-person cancellation, only 54 clubs allow cancellation by letter (certified letter for 25 clubs), and as few as 7 clubs accept cancellation by phone.

The prevalence of an automatically renewed contract is consistent with the presence of a large share of naive consumers who delay cancellation. DellaVigna and Malmendier [2003] find that the average number of full months between the last attendance and contract termination is 2.29 months, for an average expenditure in monthly fees of $185.

An alternative explanation is that automatic renewal may be the efficient default for a contract with a monthly duration. Health clubs, however, could easily devise a more efficient contract. They could automatically renew the membership of attenders, but stop charging users who have not attended the health club for more than, say, three months. Users who wanted to restart could do so by paying the membership fee again. Health clubs have the information and technology to implement such a contract, since they typically require consumers to swipe an electronic card at the entrance. Despite the potential gain in efficiency, to our knowledge no U. S. club offers this contract as an option.

Column (5) in Table I shows that the second most common contract is the annual contract, offered by 39 clubs as a frequent contract. Although in four-fifths of the cases this contract is terminated at the end of the year, the trend in the industry is to offer annual contracts with automatic renewal into a monthly contract after twelve months.[23]

Mail order industry. Sales of records through music clubs accounted for 14.7 percent of all albums sold in the United States in the mid-1990s. The two main companies, BMG and Columbia House, had $1.6 billion revenues in year 2000. Members of music and book clubs automatically receive the "selection of the month" and are charged for it, unless they return a card to decline.[24] These clubs offer free goods (e.g., 4 books, 11 compact discs) as an initial gift to new users, and market additional purchases at high prices: the CDs of the month for October and November 2001 at BMG sold at a $1.50 premium over the Amazon price. This back-loaded pricing scheme would be very costly for the firms if users enrolled, got the free goods, did the minimum required purchases, quit, and (possibly) reenrolled. The pricing may be optimal, though, if a large share of consumers remain enrolled and purchase several high-priced selections, as predicted for naive agents.

Newspapers. Major U. S. newspapers offer subscriptions with automatic renewal from week to week, with an option to quit at any time. As in the credit card example, the pricing is back-loaded, with a discount in the first eight to twelve weeks (50 percent off at the *New York Times* and the *Boston Globe;* 90 percent off at the *Washington Post*), after which the full subscription price applies. Standard switching cost models can account for this phenomenon if switching costs are as large as $50.[25]

Life insurance industry. Hendel and Lizzeri [2003] show that life insurance contracts, which are not automatically renewed, are substantially front-loaded. The authors suggest an interpretation based on the role of reclassification risk in dynamic contracts. An alternative interpretation is that naive consumers underestimate the likelihood of facing an income shock and drop-

23. Personal communication, Bill Howland, IHRSA Director of Public Relations.

24. Recently, these companies have allowed consumers to use their Web site to decline the periodic selection.

25. An interesting feature is that consumers get the low price for four additional weeks if they pay by credit card instead of by check. Credit card billing increases the transaction costs of cancellation relative to renewal.

ping out in the future, thereby losing any coverage. The appropriate firm response is to front-load the pricing.

One can extend the model of this section to encompass other examples of contract design directed toward naive users. In general, firms target the overestimation of the probability of undertaking a transaction with immediate effort costs. In vacation time-sharing and frequent flyer programs, members who miss the deadline to book a holiday resort[26] or to redeem miles lose their benefits. Naive users who sign up for these programs may overestimate the likelihood of using the services. Similarly, consumers who rent a video may delay returning it and pay a substantial fee.[27] Also, money-back guarantees, in addition to having a signaling function, may be targeted at naive users who overestimate the probability of returning the good in case of dissatisfaction. Finally, the diffusion of mail-in coupons may be due to the prevalence of naive users who expect to send the coupon but end up not doing so.[28]

VI. Conclusion

In this paper we have analyzed the contract design of rational, profit-maximizing firms that sell goods to consumers with time-inconsistent preferences and naive expectations. Firms deviate from marginal-cost pricing, charge back-loaded fees, and design contracts with automatic renewal and endogenous switching costs. These results are robust to the degree of naiveté, the market structure, the timing of payoffs, and the assumptions about uncertainty.

The predictions of the model match stylized features of the contract design in industries such as the credit card, gambling, health club, mail order, mobile phone, and vacation time-sharing industries. The observed contract design is consistent with the view that time-inconsistent preferences and naiveté are widespread features of consumer preferences.

We also discuss the welfare effect for both sophisticated and naive agents with time-inconsistent preferences. If the agents are

26. Vacation time-sharing companies require booking well in advance of the actual holiday, eleven to twelve months for RCI and five months for Hapimag.

27. Until February 2000, Blockbuster charged for each day of delay as much as for the entire previous five-day rental.

28. Cyberrebates.com operated by offering goods with a 100 percent mail-in rebate. Users who sent in the rebate received all the money back within three months. The company went bankrupt in May 2001.

sophisticated, the market interaction with the firms enables the individuals to achieve the efficient consumption level. If the agents are naive, instead, the firms design the pricing so as to take maximal advantage of the consumer overconfidence and underestimation of renewal. As a consequence, the interaction with the firms generates inefficient outcomes and, under monopoly, a redistribution of surplus from the agents to the firm.

While in this paper we consider the case of time inconsistency and naiveté, different models of nonstandard preferences or beliefs may yield similar conclusions. Models of self-control such as Gul and Pesendorfer [2001] are likely to induce below- and above-marginal-cost pricing. These rational-expectation models, however, would not predict renewal fees or endogenous cancellation costs. In fact, our results are at least as much about naiveté as they are about self-control. The model in this paper is a particular model of overconfident agents, where the overconfidence regards future self-control. The advantages of this model are the parsimony, the ability to predict if agents *under*- or *over*estimate consumption, and the support from other laboratory and field studies.

The most important message of this paper is probably the focus on firm response to nonstandard preferences and beliefs of consumers. We have argued that there is a rich set of implications for contract design and consumer welfare in the market. So far, the literature in behavioral economics has largely neglected the rational response of firms and organizations to biases of consumers. We hope that this paper will be a step in the nascent literature of behavioral industrial organization and behavioral contract theory.

Appendix 1: Mathematical Section

Case with certain costs. We consider the case of investment goods with positive social surplus, $\delta b - c - a > 0$. If the consumer signs the contract, firm profits are $\delta[-K + \mathbf{1}_{\{c<\beta\delta b-p\}}(p - a) + \mathbf{1}_{\{c<\hat{\beta}\delta b-p\}}(\delta b - p - c) - \bar{u}]$. (Notice the tie-breaking rule that no consumption takes place if $c = \beta\delta b - p$.) We distinguish cases I, II, and III. First, the firm can charge $p_I < \beta\delta b - c \leq \hat{\beta}\delta b - c$ and obtain profits $\Pi_I = \delta[-K - \bar{u} + \delta b - c - a]$. Second, the firm can charge $\beta\delta b - c \leq p_{II} < \hat{\beta}\delta b - c$ and get profits $\Pi_{II} = \delta[-K - \bar{u} + \delta b - c - p]$. Third, the firm can charge $\beta\delta b - c \leq \hat{\beta}\delta b - c \leq p_{III}$ and obtain profits $\Pi_{III} =$

$\delta[-K - \bar{u}]$. By assumption of positive surplus, $\Pi_I > \Pi_{III}$. For $\beta = \hat{\beta}$ region *II* does not exist, and the solution is any p such that $p < \beta\delta b - c$. For $\beta < \hat{\beta}$ the solution is $p^* = \beta\delta b - c$ for $\beta\delta b - a \leq c < \delta b - a$ and any p smaller than $\beta\delta b - c$ for $c < \beta\delta b - a$.

Proof of Proposition 3. After substituting for $p^* = a - (1 - \beta)\delta b$ in the expressions for $U_j^{\beta,\beta,\delta}$, for $\Pi_M^{\beta,\beta,\delta}$ and for $S^{\beta,\hat{\beta},\delta}$, these variables do not depend on β. This proves (i). To prove (ii), we use the envelope theorem to show that $\partial\Pi_M/\partial\hat{\beta} = \delta(1 - \hat{\beta})(\delta b)^2 f(\hat{\beta}\delta b - p^*)$. The latter expression is strictly positive for $b \neq 0$ and $\hat{\beta} < 1$. As for (iii), it is easy to see that $S^{\beta,\hat{\beta},\delta}$ is maximized for $p^* = a - (1 - \beta)\delta b$, which is the solution for sophisticated users. The inequality $S^{\beta,\hat{\beta},\delta} \leq S^{\beta,\beta,\delta}$ follows. Since surplus and consumer welfare coincide for the case of perfect competition, $U_{PC}^{\beta,\hat{\beta},\delta} \leq U_{PC}^{\beta,\beta,\delta}$ follows as well. Finally, the individual rationality constraint (3) implies that $U_M^{\beta,\hat{\beta},\delta} < \beta\delta\bar{u} = U_M^{\beta,\beta,\delta}$ for $1 \geq \hat{\beta} > \beta$. As for (iv), consider that $dU_j^{\beta,\hat{\beta},\delta}/d\hat{\beta} = \partial U_j^{\beta,\hat{\beta},\delta}/\partial\hat{\beta} + \partial U_j^{\beta,\hat{\beta},\delta}/\partial p \cdot \partial p_j^*/\partial\hat{\beta}$ for $j \in \{M,PC\}$. Using (7), we obtain $\partial U_{PC}^{\beta,\hat{\beta},\delta}/\partial p = -\{p - [a - (1 - \beta)\delta b]\} f(\beta\delta b - p)$; using (7) and the first-order condition (6), we also obtain $\partial U_M^{\beta,\hat{\beta},\delta}/\partial p = -\{p - [a - (1 - \beta)\delta b]\} f(\beta\delta b - p) = \partial U_{PC}^{\beta,\hat{\beta},\delta}/\partial p$. Given that $p_M^* = p_{PC}^*$, we can conclude that $dU_M^{\beta,\hat{\beta},\delta}/d\hat{\beta} - dU_{PC}^{\beta,\hat{\beta},\delta}/d\hat{\beta} = \partial U_M^{\beta,\hat{\beta},\delta}/\partial\hat{\beta} - \partial U_{PC}^{\beta,\hat{\beta},\delta}/\partial\hat{\beta} = -(\delta b)^2(1 - \hat{\beta}) f(\hat{\beta}\delta b - p^*) < 0$ for $\hat{\beta} < 1$ and $b \neq 0$. QED

Pareto increases in welfare. Consider the problem of maximizing (actual) consumer welfare at time 0 subject to the condition that the discounted firm profits be at least as large as the profits Π^* achieved in equilibrium:

$$(14) \qquad \max_{L,p} \beta\delta\left[-L + \int_{-\infty}^{\beta\delta b - p} (\delta b - p - c)\, dF(c)\right]$$

subject to

$$\delta\{L - K + F(\beta\delta b - p)(p - a)\} \geq \delta\Pi^*.$$

(Notice that $\hat{\beta}$ does not appear in this expression.) Assuming that the constraint is satisfied with equality, the problem simplifies to $\max_p \beta\delta[-K - \Pi^* + \int_{-\infty}^{\beta\delta b - p} (\delta b - a - c)\, dF(c)]$. The derivative of this expression with respect to p is $-\beta\delta[p - [a - (1 - \beta)\delta b]] f(\beta\delta b - p)$. Therefore, consumer welfare at time 0 increases as p gets closer to $p_{FB}^* = a - (1 - \beta)\delta b$, with L changing so that profits remain constant. Since for sophisticated consumers p^* already equals p_{FB}^*, no Pareto improvement is possible. For naive consumers, p^* is in general different from p^*_{FB}, and welfare of time 0 self can be improved. It remains to be shown that moving p closer to p^*_{FB} does not decrease welfare of the time 1 self; that is, $-L + \int_{-\infty}^{\beta\delta b-p} (\beta\delta b - p - c)\, dF(c)$. (Notice the β in the integrand.) After substituting for L from the constraint in (14), we obtain $\beta\delta[-K - \Pi^* + \int_{-\infty}^{\beta\delta b-p} (\beta\delta b - a - c)\, dF(c)]$. The derivative of this expression with respect to p is $-\beta\delta(p - a) f(\beta\delta b - p)$. The welfare of time 1 self therefore increases to the extent that p gets closer to a.

We consider two cases. In the first case the pricing for (partially) naive agents is such that $p^* < p_{FB}$. In this case, it is possible to increase p up to p_{FB}, change L so as to keep firm profits constant, and increase consumer surplus both from the point of view of time 0 self and time 1 self. In the second case, $a > p^* > p_{FB}$; instead, there exists no Pareto improvement. Any decrease of p toward p_{FB} (keeping firm profits constant) increases self 0 welfare but decreases self 1 welfare. Conversely, any increase in p toward a increases self 1 welfare but reduces self 0 welfare. Notice that so far, in designing a Pareto improvement, we have held the firm at a constant profit level and assumed that the consumer at time 0 gets all the surplus. Assigning more surplus to the firm in problem (14) only makes it more difficult to improve the welfare of the time 1 self. The case of leisure goods ($b < 0$) is parallel.

Proof of Proposition 4. We can rewrite program (12)–(13) as the following surplus maximization:

(15)

$$\max_{L_R,p,p_R} \beta\delta\left[\int_{-\infty}^{\beta\delta b-p} (p-a)\, dF(c) + \int_{-\infty}^{\hat\beta\delta b-p} (\delta b - p - c)\, dF(c) - K\right]$$

$$+ \beta\delta^2\left[\begin{array}{l} \mathbf{1}_{\{(L_R,p_R)\in\hat R\}}\left(-L_R + \displaystyle\int_{-\infty}^{\hat\beta\delta b-p_R} (\delta b - p_R - c)\, dF(c)\right) \\ \quad + \mathbf{1}_{\{(L_R,p_R)\in R\}}\left(L_R + \displaystyle\int_{-\infty}^{\beta\delta b-p_R} (p_R - a)\, dF(c)\right) \\ \quad + \mathbf{1}_{\{(L_R,p_R)\notin\hat R\}}\left(-\dfrac{k}{\delta} - L'^* + \displaystyle\int_{-\infty}^{\hat\beta\delta b-p'^*} (\delta b - p'^* - c)\, dF(c)\right) \end{array}\right],$$

where L^* is determined by the zero-profit condition in (13). The maximization program is separable in p and (L_R,p_R). The first line of program (15) is a positive affine transformation of the maximization program (4) in the two-period model. Therefore, the solution for p^* will coincide with the solution of program (2)–(3) and therefore also with p'^*. As for the last three lines of (15), we divide the maximization with respect to p_R and L_R in three constrained maximization programs, which we denote as Regions I, II, and III. Region I restricts the maximization to the subset of two-part tariffs (L_R,p_R) that belong to neither R, nor $\hat{R}$, that is, prices for which the agents do not renew, and expect not to renew. Region II restricts the maximization to the subset of prices (L_R,p_R) that belong to $\hat{R}$, but not to R, that is, prices for which the agents expect not to renew, but end up renewing. Finally, Region III restricts the maximization to the subset of prices (L_R,p_R) that belong to both R and $\hat{R}$, that is, prices for which the agents renew and expect to renew. These Regions include all possible cases since $\hat{R} \subset R$. We define the (perceived) benefit of switching (excluding the transaction cost of switching) as $s(p_R,L_R) \equiv -L' + \int_{-\infty}^{\hat{\beta}\delta b - p'} (\delta b - p' - c)\, dF(c) + L_R - \int_{-\infty}^{\hat{\beta}\delta b - p_R} (\delta b - p_R - c)\, dF(c)$. Using (10), the set of prices $(L_R,p_R) \in R$ can be written as $s(p_R,L_R) \leq k/\beta\delta$. Similarly, $(L_R,p_R) \in \hat{R}$ coincides with $s(p_R,L_R) \leq k/\hat{\beta}\delta$. We define the regions using this function s.

Region I ($k/\hat{\beta}\delta \leq k/\beta\delta < s(p_R,L_R)$). In this region the agent maximizes

$$\text{(16)}\quad \beta\delta^2\left[-\frac{k}{\delta} - L'^* + \int_{-\infty}^{\hat{\beta}\delta b - p'^*} (\delta b - p'^* - c)\, dF(c)\right]$$

$$= \beta\delta^2\left[-\frac{k}{\delta} - K + \int_{-\infty}^{\beta\delta b - p'^*} (p'^* - a)\, dF(c) + \int_{-\infty}^{\hat{\beta}\delta b - p'^*} (\delta b - p'^* - c)\, dF(c)\right],$$

which is independent of both L_R and p_R. Therefore, any prices L_R and p_R that are high enough to satisfy the constraint $k/\beta\delta < s(p_R,L_R)$ are optimal in this region.

Region II ($k/\hat{\beta}\delta < s(p_R,L_R) \leq k/\beta\delta$). Note that this region exists only for $\hat{\beta} > \beta$. The agent maximizes $\beta\delta^2[L_R + \int_{-\infty}^{\beta\delta b - p_R}$

$(p_R - a)\, dF(c) - k/\delta - L'^* + \int_{-\infty}^{\hat{\beta}\delta b - p'^*} (\delta b - p'^* - c)\, dF(c)]$ subject to the constraint $k/\hat{\beta}\delta < s(p_R,L_R) \le k/\beta\delta$. The constraint $s(p_R,L_R) \le k/\beta\delta$ is binding. If it were not, a firm could offer a slightly higher L_R, still satisfy the constraint $s(p_R,L_R) \le k/\beta\delta$, and increase the objective function. This also implies that the second constraint, $k/\hat{\beta}\delta < s(p_R,L_R)$, is automatically satisfied for $\hat{\beta} > \beta$. We can then solve for L_R from $s(p_R,L_R) = k/\beta\delta$ and reduce the maximization problem to

$$\text{(17)} \quad \max_{p_R} \beta\delta^2 \left[\int_{-\infty}^{\hat{\beta}\delta b - p_R} (\delta b - p_R - c)\, dF(c) + \int_{-\infty}^{\beta\delta b - p_R} (p_R - a)\, dF(c) - \frac{k}{\delta} + \frac{k}{\beta\delta} \right].$$

Once again, this function coincides, up to a positive affine transformation, with program (4) in the two-period model; the solution for p_R^* in Region II, therefore, coincides with $p^{*\prime}$. The solution for L_R^* is any L_R such that $s(p^{*\prime},L_R) = k/\beta\delta$.

Region III ($s(p_R,L_R) \le k/\hat{\beta}\delta \le k/\beta\delta$). The agent maximizes

$$\text{(18)} \quad \beta\delta^2 \left[\int_{-\infty}^{\hat{\beta}\delta b - p_R} (\delta b - p_R - c)\, dF(c) + \int_{-\infty}^{\beta\delta b - p_R} (p_R - a)\, dF(c) \right]$$

subject to the constraint $s(p_R,L_R) \le k/\hat{\beta}\delta$. The constraint is not binding, since for all p_R we can find an L_R^* low enough to satisfy the constraint without affecting the objective function. Once again, the function (18) coincides, up to a positive affine transformation, with program (4) in the two-period model; the solution for p_R^*, therefore, coincides with $p^{*\prime}$.

Given that the maximand is independent of p_R^* in Region I and that p_R^* and p'^* coincide in Regions II and III, it is easy to compare the value of the maximand at (p_R^*,L_R^*) in (16), (17), and (18) in the three regions, assuming that $K > 0$. For $k = 0$, Regions II and III yield the same value, higher than in Region I. For $k > 0$ and $\hat{\beta} = \beta$, Region III holds the highest payoff, since Region II does not exist. The optimum fee L_R^* satisfies $s(p_R^*,L_R^*) < k/\beta\delta$; that is,

$$(19)\quad L_R^* < L'^* + \frac{k}{\beta\delta} = K - F(\beta\delta b - p'^*)(p'^* - a) + \frac{k}{\beta\delta}.$$

Finally, for $k > 0$ and $1 \geq \hat{\beta} > \beta$, Region II is optimal, and the optimal fee L_R^* satisfies $s(p_R^*, L_R^*) = k/\beta\delta$; that is, $L_R^* = L'^* + k/\beta\delta$. The value of L^* is determined by the zero-profit condition for Regions II or III; that is,

$$(20)\qquad L^* = K - (1+\delta)F(\beta\delta b - p^*)(p^* - a) - \delta L_R^*.$$

This proves Proposition 4(ii). As for Proposition 4(iii), we can use (20) to obtain the difference between the two lump-sum fees for the case $k > 0$ and $\hat{\beta} = \beta$:

$$\begin{aligned} L_R^* - L^* &= -K + (1+\delta)F(\beta\delta b - p^*)(p^* - a) + (1+\delta)L_R^* \\ &\leq \delta K + (1+\delta)k/\beta\delta, \end{aligned}$$

where in the second row we used (19) to substitute for L_R^*. Similarly, we obtain $L_R^* - L^* = \delta K + (1+\delta)k/\beta\delta$ for the case $k > 0$ and $1 \geq \hat{\beta} > \beta$. This proves Proposition 4(iii). QED

Proof of Proposition 5. For sophisticated and time-consistent agents ($\beta = \hat{\beta} \leq 1$), the objective function is given by equation (18) with p_R^* substituting for p_R. This function is independent of k. This implies that sophisticated and time-consistent agents are indifferent between any choice of $k \in [0; \bar{k}]$. For naive agents ($\beta < \hat{\beta} = 1$), the objective function is given by equation (17), which is strictly increasing in k as long as $\beta < 1$. Therefore, the firms set k^* at the upper bound $\bar{k}$. QED

Welfare and profits in model with switching costs. Given the features of the equilibrium ($p^* = p_R^*$ and expression (20) for $L^* + \delta L_R^*$), consumer welfare under perfect competition is given by

$$\begin{aligned} U_{PC}^{\beta,\beta,\delta} &= \delta\left[-L^* + (1+\delta)\int_{-\infty}^{\beta\delta b - p^*}(\delta b - p^* - c)\,dF(c) - \delta L_R^*\right] \\ &= \delta\left[(1+\delta)\int_{-\infty}^{\beta\delta b - p^*}(\delta b - a - c)\,dF(c) - K\right]. \end{aligned}$$

This expression, which also indicates the joint consumer-firm surplus, is a multiple of expression (8) for welfare in the two-period model, except for the fact that setup cost K is paid only once. Therefore, the conclusions in Proposition 3 for the case of perfect competition also hold for the model with renewal. Consumer welfare under monopoly for the case $\beta < \hat{\beta}$ is given by

$$U_M^{\beta,\beta,\delta} = \delta\left[-L^* + (1+\delta)\int_{-\infty}^{\beta\delta b - p^*} (\delta b - p^* - c)\, dF(c) - \delta L_R^*\right]$$

$$= \delta\left[\bar{u} - (1+\delta)\int_{\beta\delta b - p^*}^{\hat{\beta}\delta b - p^*} (\delta b - p^* - c)\, dF(c) - k(1 - 1/\beta)\right].$$

Appendix 2: Survey Transcript

1. "Hi. My name is ***. I heard good things about your gym and I am considering joining. Could you please tell me which options you offer to attend your club?" If the respondent is not willing to answer, we insist: "I would really appreciate it if you could tell me. I would like to have some more information before I check out your gym."

For each option we ask questions 2, 3, and 4 if applicable: 2. "Is there an initiation fee?" 3. "Do I have to renew the contract at the end of the period or is it automatically renewed?" 4. (if expiration is not automatic) "If I want to quit the gym, how do I cancel?"

After these questions, we ask: 5. "Are there other ways to attend?"

If the person has not mentioned them yet, we ask: 6. "Do you also offer monthly contracts?" 7. "Do you also offer contracts for a longer time period, such as a year?" 8. "Could I just come and pay each time I use the gym?"

Finally, we inquire about the type of club: 9. "One last thing: What facilities do you offer? Do you also have racket courts?" We conclude the survey. 10. "Thank you very much. I appreciated your help."

Department of Economics, University of California Berkeley
Graduate School of Business, Stanford University

References

Akerlof, George A., and Janet L. Yellen, "Can Small Deviations from Rationality Make Significant Differences to Economic Equilibria?" *American Economic Review,* LXXV (1985), 708–720.

Angeletos, Marios, David I. Laibson, Andrea Repetto, Jeremy Tobacman, and Stephen Weinberg, "The Hyperbolic Buffer Stock Model: Calibration, Simulation, and Empirical Evaluation," *Journal of Economic Perspectives,* XV (2001), 47–68.

Ariely, Dan, and Klaus Wertenbroch, "Procrastination, Deadlines, and Performance: Self-Control by Precommitment," *Psychological Science,* XIII (2002), 219–224.

Ausubel, Lawrence M., "The Failure of Competition in the Credit Card Market," *American Economic Review,* LXXXI (1991), 50–81.

——, "Adverse Selection in the Credit Card Market," mimeo, University of Maryland, College Park, 1999.

Barro, Robert J., and Paul M. Romer, "Ski-Lift Pricing, with Applications to Labor and Other Markets," *American Economic Review,* LXXVII (1987), 875–890.

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

Bénabou, Roland, and Jean Tirole, "Intrinsic and Extrinsic Motivation," *Review of Economic Studies,* LXX (2003), 489–520.

Comings, David E., Richard J. Rosenthal, Henry R. Lesieur, et al., "A Study of the Dopamine D2 Receptor Gene in Pathological Gambling," *Pharmacogenetics,* VI (1996), 223–234.

DellaVigna, Stefano, and Ulrike Malmendier, "Self-Control in the Market: Evidence from the Health Club Industry," mimeo, 2003.

DellaVigna, Stefano, and Daniele M. Paserman, "Job Search and Hyperbolic Discounting," The Maurice Falk Institute for Economic Research in Israel, Jerusalem, Israel, Discussion Paper No. 00.15, 2000.

De Long, J. Bradford, Andrei Shleifer, Lawrence H. Summers, and Robert J. Waldmann, "Noise Trader Risk in Financial Markets," *Journal of Political Economy,* XCVIII (1990), 703–738.

Evans, David, and Richard Schmalensee, *Paying with Plastic: The Digital Revolution in Buying and Borrowing* (Cambridge, MA: MIT Press, 1999).

Fang, Hanming, and Daniel Silverman, "Time Inconsistency and Welfare Program Participation: Evidence from the NLSY," mimeo, 2001.

Farrell, J., and Paul Klemperer, "Co-ordination and Lock-in: Competition with Switching Costs and Network Effects," in M. Armstrong and R. Porter, eds., *Handbook of Industrial Organization,* vol. III, forthcoming.

Gabaix, Xavier, and David Laibson, "Competition and Consumer Confusion," mimeo, 2004.

Gross, David B., and Nicholas S. Souleles, "Do Liquidity Constraints and Interest Rates Matter for Consumer Behavior? Evidence from Credit Card Data," *Quarterly Journal of Economics,* CXVII (2002), 149–185.

Gruber, Jonathan, and Botond Koszegi, "Is Addiction 'Rational'? Theory and Evidence," *Quarterly Journal of Economics,* CXVI (2001), 1261–1303.

Gruber, Jonathan, and Sendhil Mullainathan, "Do Cigarette Taxes Make Smokers Happier?" NBER working paper No. 8872, 2002.

Gul, Faruk, and Wolfgang Pesendorfer, "Temptation and Self-Control," *Econometrica,* LXIX (2001), 1403–1436.

Hendel, Igal, and Alessandro Lizzeri, "The Role of Commitment in Dynamic Contracts: Evidence from Life Insurance," *Quarterly Journal of Economics,* CXVIII (2003), 299–327.

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

Laibson, David I., Andrea Repetto, and Jeremy Tobacman, "A Debt Puzzle," in Philippe Aghion, Roman Frydman, Joseph Stiglitz, and Michael Woodford, eds., *Knowledge, Information, and Expectations in Modern Economics: In Honor of Edmund S. Phelps,* forthcoming.

Larwood, Laurie, and William Whittaker, "Managerial Myopia: Self-Serving Biases in Organizational Planning," *Journal of Applied Psychology,* LXII (1977), 194–198.

Loewenstein, George, and Drazen Prelec, "The Red and the Black: Mental Accounting of Savings and Debt," *Marketing Science,* XVII (1998), 4–28.

Madrian, Brigitte C., and Dennis Shea, "The Power of Suggestion: Inertia in 401(k) Participation and Savings Behavior," *Quarterly Journal of Economics,* CXVI (2001), 1149–1187.

O'Donoghue, Ted D., and Matthew Rabin, "Doing It Now or Later," *American Economic Review,* LXXXIX (1999a), 103–124.
O'Donoghue, Ted D., and Matthew Rabin, "Incentives for Procrastinators," *Quarterly Journal of Economics,* CXIV (1999b), 769–816.
O'Donoghue, Ted D., and Matthew Rabin, "Choice and Procrastination," *Quarterly Journal of Economics,* CXVI (2001), 121–160.
Oi, Walter, "A Disneyland Dilemma: Two-Part Tariffs for a Mickey Mouse Monopoly," *Quarterly Journal of Economics,* LXXXV (1971), 77–96.
Paserman, M. Daniele, "Job Search and Hyperbolic Discounting: Structural Estimation and Policy Evaluation," mimeo, Hebrew University of Jerusalem, 2003.
Phelps, Edmund S., and Robert A. Pollak, "On Second-Best National Saving and Game-Equilibrium Growth," *Review of Economic Studies,* XXXV (1968), 85–199.
Russell, Thomas, and Richard Thaler, "The Relevance of Quasi Rationality in Competitive Markets," *American Economic Review,* LXXV (1985), 1071–1082.
Strotz, Robert H., "Myopia and Inconsistency in Dynamic Utility Maximization," *Review of Economic Studies,* XXIII (1956), 165–180.
Svenson, Ola, "Are We All Less Risky and More Skillful than Our Fellow Drivers?" *Acta Psychologica,* XLVII (1981), 143–148.
Wray, I., and M. Dickerson, "Cessation of High Frequency Gambling and 'Withdrawal' Symptoms," *British Journal of Addiction,* LXXVI (1981), 401–405.