Review of Economic Studies (2006) **73**, 689–714 0034-6527/06/00250689$02.00

Contracting with Diversely Naive Agents

KFIR ELIAZ
New York University

and

RAN SPIEGLER
School of Economics, Tel Aviv University

First version received August 2004; *final version accepted November* 2005 (*Eds.*)

In standard contract-theoretic models, the underlying assumption is that agent types differ in their preference or cost parameters, and the principal's objective is to design contracts in order to screen this type. We study a contract-theoretic model in which the heterogeneity among agent types is of a "cognitive" nature. In our model, the agent has dynamically inconsistent preferences. Agent types differ only in their degree of "sophistication", that is, their ability to forecast the change in their future tastes. We fully characterize the menu of contracts which the principal offers in order to screen the agent's sophistication. The menu does not exclude any type: it provides a perfect commitment device for relatively sophisticated types, and "exploitative" contracts which involve speculation with relatively naive types. More naive types are more heavily exploited and generate a greater profit for the principal. Our results allow us to interpret real-life contractual arrangements in a variety of industries.

1. INTRODUCTION

In standard contract-theoretic models, agent types differ in their preferences. Usually, the agent's type is a parameter that characterizes the agent's willingness to pay or the agent's cost of exerting effort. However, an agent's personal characteristics may also include cognitive features. One such feature is an agent's ability to forecast changes in his future tastes. In this paper, we study optimal contracting with dynamically inconsistent agents who differ in their forecasting abilities.

In recent years, economists have become increasingly interested in the implications of dynamically inconsistent preferences on economic behaviour in a variety of contexts. A number of authors (O'Donoghue and Rabin, 1999*a*; Gilpatric, 2003; Sarafidis, 2004; and especially DellaVigna and Malmendier, 2004; 2006) have begun exploring the question of how a rational entity would contract with dynamically inconsistent agents. We contribute to this line of research by studying the following question: What set of contracts would an ordinary profit-maximizing monopoly offer when facing dynamically inconsistent consumers who may differ in their forecasting abilities?

To study this question, we consider a two-period model of monopolistic contracting with dynamically inconsistent agents. In the model, a principal is the sole provider of some set of actions. In order to choose an action from this set, an agent must sign a contract with the principal, one period prior to his choice of action. The agent may also refuse to sign any contract, in which case he chooses some outside option. A contract specifies a monetary transfer for each of the actions. Other than that, we place no restriction on the space of contracts. The agent has quasi-linear utility over action–transfer pairs. However, the agent's utility from actions changes from period to period. At the time of signing the contract, his utility function is u, whereas at the time the action is chosen, it is v. The only assumption we impose on u and v is continuity. This allows us to accommodate a variety of sources for changing tastes.

As an illustration of this environment, imagine that a cable TV company offers you a basic package as well as an extended package containing all channels. In the absence of cable TV, you have a low willingness to pay for the extended package. However, once exposed to it, you become hooked and raise your willingness to pay for this high-end service. Alternatively, suppose that you visit a hotel-casino. Initially, you wish to gamble only a little for the sake of the experience. Once you start gambling, it becomes difficult for you to stop and you want to gamble beyond your initial plans.

In these situations, the principal and the agent are likely to disagree about the agent's propensity to change his tastes. Specifically, the cable TV company or the casino may assign a higher probability to the event that the agent's preferences will change. To incorporate this intuition, we assume that the principal *knows* that the agent's second-period utility is v. In contrast, the agent is only *partially aware of his changing tastes*. He believes that with probability θ, his second-period utility from actions will remain u, and that with probability $1-\theta$, it will be v. Drawing on the terminology of O'Donoghue and Rabin (1999*b*), we say that an agent with $\theta = 0$ is *fully sophisticated* and an agent with $\theta = 1$ is *fully naive*. We refer to agents with intermediate values of θ as *partially naive* . The principal does not observe θ, but he knows the distribution of this parameter on $[0, 1]$. Hence, the parameter θ plays the role of the agent's "private type". The principal's problem is to design a *menu of contracts* that maximizes his expected profits.

How does a partially naive, dynamically inconsistent agent evaluate a contract? Because he does not know if he will maximize u or v in the second period, he associates with each contract two actions and two corresponding transfers. One action–transfer pair maximizes the agent's net utility (the utility from the action minus the payment) according to u, while another maximizes his net utility according to v. From the principal's point of view, the former action is "imaginary", while the latter is the "real" action that is actually chosen in the second period. To compute his indirect utility from a contract, the agent calculates the expected u-value of each of these two actions and subtracts the expected sum of the corresponding transfers. Thus, the agent's dynamic inconsistency is reflected by the fact that he evaluates future actions according to his first-period utility function.

The principal's problem is therefore reduced to assigning two action–transfer pairs to each type: an imaginary action with its corresponding transfer and a real action with its corresponding transfer. Only the latter pair directly affects the principal's revenue. The former pair is used by the principal to induce naive types to sign "exploitative" contracts that extract more than their first-period willingness to pay. Exploitative contracts involve speculative trade: an agent who accepts these contracts believes that he extracts rents from the principal, and the principal believes the converse.

To better understand the workings of our model, consider the following example. An agent visiting a casino initially wants to play a slot machine only once and is willing to pay at most \$6 for the experience. However, once the agent tries a slot machine, he is willing to pay up to \$5 for each of 10 additional games. Suppose there are only two types: the fully sophisticated ($\theta = 0$) and the fully naive ($\theta = 1$). In order to maximize its revenues, the casino offers two options: (1) the "standard fare"—paying \$6 per game and (2) a "promotion"—getting the first game for free and 10 additional games for \$49.99.

A fully naive agent believes that his preferences would not change, such that he will play a single game in either option. Therefore, he will choose the second alternative, attracted by the free game it promises. This alternative constitutes an exploitative contract, in which playing a single game is the imaginary action and playing all 11 games is the real action. In contrast, the fully sophisticated agent opts for the first contract, which effectively commits him to play only one game. Note that in this example, the principal manages to get the first best from each type. In other words, there are no screening costs.

Section 4 characterizes the optimal menu of contracts with a continuum of types and a continuum of actions. The key features of this menu are the following.

Pooling of the sophisticated types (no exploitation near the bottom). The set of types is partitioned into two intervals. The relatively sophisticated types (low θ) all choose the same contract that essentially commits them to choose the action that maximizes the difference between the first-period willingness to pay, $u(\cdot)$, and the principal's costs, $c(\cdot)$. The relatively naive types (high θ) choose exploitative contracts. Whether there is finer discrimination among the exploited naive types depends on the more specific features of u and v.

Monotonicity of profits. The principal's actual profit from the agent increases with his naivete. At the same time, the transfer associated with the imaginary action decreases with the agent's naivete. The agent's first-period net utility decreases with his naivete.

No exclusion. The relatively sophisticated types, who end up being unexploited, exert no informational externality on the relatively naive types. It follows that as long as $\max_a[u(a) - c(a)] > 0$ (*i.e.* there is a surplus in the interaction with a fully sophisticated agent), the optimal menu does not exclude any type.

We expect these effects to persist in more complicated versions of our model. In Section 5, we provide simple examples, in which the optimal menu resembles real-life menus offered by credit agencies, casinos, and other kinds of firms. Section 6 establishes that in some natural environments, the optimal menu can be implemented with *three*-part tariffs but not with two-part tariffs. Thus, our model may be viewed as a justification for three-part tariffs.

It should be emphasized that from a formal point of view, our model relaxes two standard assumptions: time consistency and common prior. The reason we do so is that as argued above, dynamically inconsistent preferences provide a case in which assuming different priors seems particularly natural. This raises the question of what role is played by each of the two assumptions. We discuss this issue—as well as other key features of our model—in Section 7.

2. RELATED LITERATURE

Our paper follows up a small literature, which has begun exploring the problem of contracting with dynamically inconsistent agents. O'Donoghue and Rabin (1999*a*) study optimal incentive design for procrastinating agents, where the principal's objective is to complete tasks efficiently. Gilpatric (2003) extends this framework by enriching the agent's private information structure. Sarafidis (2004) studies a durable-good monopoly model with partially naive, dynamically inconsistent agents.

Within this literature, the most closely related work is that of DellaVigna and Malmendier (2004) (DM, henceforth). In this paper, a monopolistic firm offers a two-part tariff to an agent with (β, δ) preferences. The agent is partially naive in the sense of O'Donoghue and Rabin (2001): he believes that his future value of β is higher than it is. DM show that in the optimal two-part tariff, the per-usage price falls below the firm's marginal cost in the case of "investment goods", and lies above marginal cost in the case of "leisure goods". These predictions are qualitatively robust to competition. DM provide compelling evidence for these predictions in a variety of markets.

The main difference between the present study and that of DM is we focus on the problem of discriminating between diversely naive types, whereas in DM the firm knows the agent's type.

Moreover, DM restrict the firm's contract space to two-part tariffs, in line with their objective to explain systematic departures from marginal cost pricing. As it turns out, the per-usage pricing effects in DM are qualitatively independent of the agent's naivete. For a sophisticated agent, the per-usage pricing effect is a commitment device, whereas for a naive agent, it is an exploitation device. By comparison, we do not restrict the domain of feasible contracts. As a result, our characterization highlights the contractual arrangements that are specifically designed for the purpose of screening the agent's degree of naivete. Another distinguishing feature of our paper is that we do not commit to a particular kind of dynamic inconsistency.

Esteban, Miyagawa and Shum (2003) study a model of contracting with agents having self-control problems, taking a different approach. Specifically, they analyse a non-linear pricing problem, in which consumers' preferences are given by the functional form introduced by Gul and Pesendorfer (2001). Amador, Werning and Angeletos (2004) study the problem of designing the optimal savings/consumption path for an agent with a present bias, who anticipates a shock to his second-period utility from consumption. The authors take two approaches to modelling the agent's present bias: (i) a dynamically *inconsistent* approach using the (β, δ) model and (ii) a dynamically *consistent* approach using the Gul–Pesendorfer functional form.

The idea that a principal may wish to discriminate between consumer types according to their cognitive features appears for the first time (to our knowledge) in Rubinstein (1993). In this paper, consumers have bounded ability to categorize realizations of a random variable. Different consumer types have different categorization abilities, and the principal's optimal contract is designed to screen their type. Piccione and Rubinstein (2003) perform a similar exercise, when different consumer types differ in their ability to perceive temporal patterns.

The literature also contains models of competition when agents have elements of bounded rationality that are somehow related to the phenomena discussed in this paper. Laibson and Yariv (2004) analyse an intertemporal competitive economy, in which firms compete in "Dutch books" over agents with inconsistent preferences. Spiegler (2004) studies competition over agents who are incapable of fully perceiving multi-dimensional goods and therefore evaluate them according to a single, randomly selected dimension. Gabaix and Laibson (2005) analyse competition when agents have imperfect awareness of a particular contingency that may arise after the good is purchased.

3. THE MODEL

A principal faces a continuum of agents. The principal can provide each agent with the opportunity to choose an action $a \in [0,1]$. The cost of providing this action is $c(a)$, where $c(\cdot)$ is an increasing function. In order to have access to this set of actions, the agent must sign a contract with the principal one period beforehand. If the agent does not sign a contract with the principal, he is restricted to the default action $a = 0$. We refer to the period in which a contract is signed as period 1, and to the period in which the action is chosen as period 2. A contract is a function $t : [0,1] \to \mathbb{R}$, which specifies, for every second-period action, a (possibly negative) transfer from the agent to the principal. The principal is perfectly able to monitor the agent's second-period action.

Agents have quasi-linear preferences over action–transfer pairs. In period 1, the agents' utility from second-period actions is given by a continuous function $u : [0,1] \to \mathbb{R}$ with $u(0) = 0$. The principal and the agents have conflicting beliefs regarding the agents' preferences in period 2. The principal believes that in period 2, the agents' utility from actions will be given by a continuous utility function $v : [0,1] \to \mathbb{R}$. In contrast, an agent of type θ believes that with probability θ, his second-period utility from actions will remain u, and with probability $1-\theta$, it will change into v. The principal does not observe θ, but believes that θ is distributed according to a continuous *cdf* $F(\theta)$ with support $[0,1]$.

In period 1, an agent evaluates contracts according to the standard "multi-selves" approach. That is, he computes a probability distribution over his second-period actions, according to his beliefs, and evaluates this distribution according to his first-period utility function. Faced with the contract $t(\cdot)$, the agent knows that he will choose $a^{\upsilon} \equiv \arg\max_a[\upsilon(a) - t(a)]$ if his second-period utility is υ, and $a^u \equiv \arg\max_a[u(a) - t(a)]$ if his second-period utility is u. Therefore, the first-period indirect utility of a "partially naive" type $\theta \in [0,1]$ from a contract $t(\cdot)$ is

$$\theta[u(a^u) - t(a^u)] + (1-\theta)[u(a^{\upsilon}) - t(a^{\upsilon})].$$

The principal's objective is to maximize expected profits. By the revelation principle, a solution to this problem may be obtained via a direct revelation mechanism in which agents are asked to report their type, and each reported type ϕ is assigned a contract $t_\phi : [0,1] \to \mathbb{R}$. The principal's problem is then to find the optimal set of functions $\{t_\theta(a)\}_{\theta \in [0,1]}$.

3.1. *Non-common priors*

Formally, we analyse a principal–agent model with non-common priors. We have a particular interpretation in mind: a situation in which the agents have a systematic bias in forecasting their future tastes, whereas the principal has an unbiased forecast. This judgement is arbitrary, as far as analysing the optimal menu of contracts is concerned. However, it is crucial for its welfare implications. We believe that in many markets, it is reasonable to assume that the firm—with its army of marketing experts—has better knowledge of the agents' systematically changing tastes (*e.g.* their vulnerability to temptation, propensity to procrastinate, or sensitivity to reference points) than some of the agents themselves.

Such a description may bring to mind principal–agent models with an informed principal. However, this is not the case here because it is not common knowledge that the principal knows the state of nature. In order to justify our interpretation of the non-common-prior assumption, one should assume that either the agent is unaware of the principal's superior knowledge or believes (due to overconfidence) that this knowledge does not apply in his particular case, even if it does apply to the rest of the population. Under both justifications, the principal knows the true state of nature, whereas the agent holds an erroneous belief.

3.2. *Bounded bets*

It is important to note that although the principal and the agents disagree on the prior, they cannot make bets on the true state of nature because the state, that is, the agents' second-period utility function, is not verifiable. If the state were verifiable, the two parties would want to make infinite bets. The fact that only the agent's second-period action is verifiable imposes an endogenous upper bound on the bets.

3.3. *The source of dynamic inconsistency*

A key feature of our model is the assumption that the agents' preferences change between the time in which they sign the contract and the time in which they choose an action. Our formulation implicitly assumes that the mere passage of time causes this change. However, preferences may also change because of some action taken by the agent at period 1. For example, the principal may provide a good or service for a "free trial" period, thereby causing the agent to become "addicted", or experience an "endowment effect". We do not explicitly model these effects. However, it would be straightforward to complicate the model by assuming that the switch from u to υ depends on an action taken at period 1. This would not change the results, except for the specification of $t(a)$ for $a \notin \{a^u, a^\upsilon\}$ in some of the contracts offered by the principal.

4. CHARACTERIZING THE OPTIMAL MENU

In this section, we characterize the optimal menu of contracts. In Section 4.1, we obtain some general properties of the optimal menu. In Section 4.2, we adapt standard mechanism-design tools to derive an algorithm for computing the optimal menu. In Section 4.3, we carry out a welfare analysis.

4.1. *Qualitative features of the menu*

The principal's mechanism-design problem can be written as follows.

Observation 1. *The optimal menu of contracts* $\{t_\theta(a)\}_{\theta\in[0,1]}$ *is given by the solution to the following maximization problem:*

$$\max_{\{t_\theta(a)\}_{\theta\in[0,1]}} \int_0^1 [t_\theta(a_\theta^\upsilon) - c(a_\theta^\upsilon)]dF(\theta)$$

subject to the constraints,

$$\theta[u(a_\theta^u) - t_\theta(a_\theta^u)] + (1-\theta)[u(a_\theta^\upsilon) - t_\theta(a_\theta^\upsilon)] \geq 0 \tag{IR_θ}$$

$$\begin{aligned}\theta[u(a_\theta^u) - t_\theta(a_\theta^u)] &+ (1-\theta)[u(a_\theta^\upsilon) - t_\theta(a_\theta^\upsilon)]\\ &\geq \theta[u(a_\phi^u) - t_\phi(a_\phi^u)] + (1-\theta)[u(a_\phi^\upsilon) - t_\phi(a_\phi^\upsilon)]\end{aligned} \tag{$IC_{\theta,\phi}$}$$

for all $\phi \in [0,1]$, *where*

$$a_\theta^u \in \arg\max_{a\in[0,1]} [u(a) - t_\theta(a)] \tag{UR_θ}$$

$$a_\theta^\upsilon \in \arg\max_{a\in[0,1]} [\upsilon(a) - t_\theta(a)]. \tag{VR_θ}$$

The first and second constraints are the standard individual rationality and incentive compatibility constraints. Condition IR_θ says that an agent of type θ is at least as well off with his assigned contract as with the default option. Condition $IC_{\theta,\phi}$ says that an agent of type θ cannot be better off by pretending to be of type ϕ and signing the contract assigned to that type.

The novel conditions are UR_θ and VR_θ. These conditions represent the fact that an agent's indirect utility from a contract is determined by the actions he would choose in the states of the world he deems possible. If the agent's period 1 utility does not change in period 2 (an event to which the agent assigns a probability of θ), then he will choose the optimal action for him according to the utility function u. This is represented by UR_θ. If, on the other hand, the agent's utility changes into υ (an event to which the agent assigns a probability of $1-\theta$), then he will choose the optimal action for him according to the utility function υ. This is precisely the condition VR_θ. Note that there is no "second-period individual rationality" constraint. Once the agent has signed the contract, he is obliged to it, even if from his second-period perspective, he regrets having signed it.

Observation 1 implies that any contract t can be identified with a pair of actions a_θ^v and a_θ^u. The former action is consistent with v-maximization in the second period. The latter action is consistent with u-maximization in the second period. Since the principal believes that with probability 1 the agent behaves according to v in the second period, we refer to a_θ^u as the *imaginary action* and to a_θ^v as the *real action*. W.l.o.g., we may assume that $t(a) = \infty$ for every $a \notin \{a_\theta^v, a_\theta^u\}$.

The constraints IR_θ and $IC_{\theta,\phi}$ can be written more compactly by introducing the following notation. Let $U(\phi,\theta)$ denote the utility of a type θ agent who pretends to be of type ϕ, that is,

$$U(\phi,\theta) \equiv \theta[u(a_\phi^u) - t_\phi(a_\phi^u)] + (1-\theta)[u(a_\phi^v) - t_\phi(a_\phi^v)].$$

Then, IR_θ and $IC_{\theta,\phi}$ can be rewritten as $U(\theta,\theta) \geq 0$ and $U(\theta,\theta) \geq U(\phi,\theta)$ for all θ and ϕ.

Definition 1. A contract $t_\theta(\cdot)$ is exploitative if $t_\theta(a_\theta^v) > u(a_\theta^v)$.

Definition 1 formalizes the notion that an exploitative contract extracts more than the agent's willingness to pay, from his first-period perspective. An agent who fully anticipates the change in his future tastes would never accept an exploitative contract. Our first result shows that the type space can be partitioned into two intervals: a set of relatively sophisticated types who choose non-exploitative contracts and a set of relatively naive types who choose exploitative contracts.

Proposition 1. *There exists a type $\underline{\theta} \in [0,1]$, such that for every $\theta > \underline{\theta}$ the contract t_θ is exploitative, while for every $\theta < \underline{\theta}$ the contract t_θ is not exploitative. Moreover, without loss of generality (w.l.o.g), there is a unique non-exploitative contract: $t(a) = u(a)$ for $a \in \arg\max_a[u(a) - c(a)]$ and $t(a') = \infty$ for every $a' \neq a$.*

There are several aspects to this result. First, if the menu contains a non-exploitative contract, then this contract must be the first best against a fully sophisticated type. Only exploitative contracts can generate higher profits for the principal. The non-exploitative contract is a perfect commitment device.[1]

Second, there exists a cut-off $\underline{\theta}$, which partitions the type space into exploited and unexploited agents. Suppose that type θ accepts an exploitative contract. By the definition of exploitative contracts, $u(a_\theta^v) - t_\theta(a_\theta^v) < 0$. Therefore, by the IR_θ constraint, $u(a_\theta^u) - t_\theta(a_\theta^u) > 0$. In other words, in order for the agent to accept an exploitative contract, he must believe that with probability θ he will receive a *gift*. It follows that for any type $\phi > \theta$, $U(\theta,\phi) > U(\theta,\theta) \geq 0$. This is essentially a "single-crossing" argument. Because the non-exploitative contract yields zero indirect utility for all types, type ϕ must accept an exploitative contract, too.

The same type of argument leads to the following result:

Remark 1. If $\underline{\theta} \in (0,1)$, then $IR_{\underline{\theta}}$ is binding.

The non-exploitative contract yields zero indirect utility for *all* types. Therefore, a type θ who accepts this contract does not exert any informational externality on any type ϕ who accepts an exploitative contract because $IC_{\phi,\theta}$ is equivalent to IR_ϕ. This is a non-standard effect: in conventional models of price discrimination, "low" types exert an informational externality on

1. If we extended the model such that the switch from u to v was a result of some action taken by the agent at period 1, then the commitment device would not have to be a prohibitively high payment for $a \notin \arg\max_a[u(a) - c(a)]$. It could just as well be a decision not to take the action that causes the switch. Nothing else the results would change.

"high" types. That is, in those models, high types receive informational rents because they have an incentive to mimic the low types. In contrast, in our model, the high (naive) types do not require an informational rent to prevent them from choosing the non-exploitative contract intended for the low (sophisticated) types.

This no-externality effect has the following important implication:

Corollary 1. *As long as* $\max_a[u(a) - c(a)] > 0$, *the optimal menu does not exclude any type.*

Thus, as long as there is a positive surplus in the interaction with fully sophisticated types, the principal's imperfect information does not lead him to exclude sophisticated types.

We now proceed to characterize the exploitative contracts. As our discussion of Proposition 1 suggests, an agent who accepts an exploitative contract $t_\theta(\cdot)$ associates a "gift" with the imaginary action a^u_θ. This gift compensates him for the excessive payment for the real action a^υ_θ. Our next result shows that both the principal's profit from the real action and the gift associated with the imaginary action increase with the agent's naivete.

Proposition 2. *Suppose the optimal menu includes at least one exploitative contract. Then*

(i) *for every exploitative contract* t_θ, *we can set w.l.o.g.* $a^u_\theta \in \arg\max_a[u(a) - \upsilon(a)]$,
(ii) $t(a^u_\theta)$ *is non-increasing in* θ *in the range* $\theta > \underline{\theta}$, *and*
(iii) $t(a^\upsilon_\theta) - c(a^\upsilon_\theta)$ *is non-decreasing in* θ.

To understand the intuition for (i), recall that the imaginary action does not directly affect the principal's profit. The role of the imaginary action is to allow the principal to extract the highest possible amount from the agent. This can be achieved by choosing an imaginary action associated with the highest differential between the true utility of period 2 and the imaginary one. An agent who is assigned the action $\arg\max_a[u(a) - \upsilon(a)]$ in period 2 would be willing to pay a high amount in order to avoid it when he discovers that his period 2 utility is $\upsilon(\cdot)$ and not $u(\cdot)$.

Part (ii) is a result of an adapted single-crossing argument. Consider a pair of indifference curves for two types, ϕ and θ with $\phi > \theta$, drawn in the space of $u(a^\upsilon) - t(a^\upsilon)$ and $u(a^u) - t(a^u)$, where a^υ and a^u denote the real and imaginary actions. Note that because the indirect utility of an agent from a contract is linear in his type, the two indifference curves satisfy the single-crossing property. Incentive compatibility then implies that $u(a^u_\phi) - t(a^u_\phi) \geq u(a^\upsilon_\theta) - t(a^\upsilon_\theta)$. By part (i) of the proposition, $a^u_\phi = a^u_\theta$, hence $t(a^u_\phi) \leq t(a^u_\theta)$. Note that single crossing alone cannot deliver this result: the identity of the imaginary action for all exploitative contracts is necessary.

Finally, part (iii) follows from a standard single-crossing argument. Suppose $t(a^\upsilon_\theta) - c(a^\upsilon_\theta)$ has a single peak at $\theta < 1$. Then, the principal can increase his profit by omitting all contracts $t_\phi(\cdot)$ for $\phi > \theta$. To see why such a modification in the menu does not violate any of the constraints, recall that if $U(\theta, \theta) \geq 0$, then $U(\theta, \phi) > 0$. Second, the incentive compatibility constraints imply that if type θ prefers his contract $t_\theta(\cdot)$ to any contract $t_\varphi(\cdot)$ for $\varphi < \theta$, then any type higher than θ also prefers θ's contract to any $t_\varphi(\cdot)$. Finally, if $t_\theta(\cdot)$ satisfied VR_θ in the original menu, then it would satisfy this constraint for any type who chooses that contract.[2]

2. Without further details about the structure of u and υ, we cannot determine how $t_\theta(a^\upsilon_\theta)$ changes with θ. In some applications, a^υ_θ increases with θ. In these cases, part (iii) implies that $t_\theta(a^\upsilon_\theta)$ increases with θ.

4.2. *Computing the optimal menu*

We begin this sub-section by introducing some helpful notation. Proposition 2 implies that w.l.o.g. we may restrict attention to an optimal menu in which all exploitative contracts assign the same imaginary action $\arg\max_a[u(a)-v(a)]$. We shall denote this action by a^*. Also, denote $u(a^*)-v(a^*)=\Delta^*$. If type θ chooses a^* in period 2, he will earn a net surplus of $v(a^*)-t_\theta(a^*)$. Hence, to satisfy VR_θ, the agent's second-period net surplus must be at least as high as $v(a^*)-t_\theta(a^*)$. We denote the slack in the VR_θ constraint by δ_θ. For the final piece of notation, we rewrite the indirect utility of type θ from a contract $t_\theta(\cdot)$ as follows:

$$U(\theta,\theta)=\theta[u(a^u_\theta)-t_\theta(a^u_\theta)-u(a^v_\theta)+t_\theta(a^v_\theta)]+[u(a^v_\theta)-t_\theta(a^v_\theta t)]. \tag{1}$$

We then define

$$q(\theta)\equiv[u(a^u_\theta)-t_\theta(a^u_\theta)]-[u(a^v_\theta)-t_\theta(a^v_\theta)]. \tag{2}$$

The quantity $q(\theta)$ may be interpreted as the period 1 *consumer surplus* from the *speculative* trade between the principal and an agent of type θ. If the agent had the same prior as the principal, then the agent's indirect utility from his contract (evaluated in period 1) would be $u(a^v_\theta)-t_\theta(a^v_\theta)$. However, an agent of type θ believes that with probability θ, a state which the principal has not anticipated will occur. In this state, he expects to obtain a net surplus (as evaluated in period 1) of $u(a^u_\theta)-t_\theta(a^u_\theta)$. The difference between this net surplus and what the principal believes to be the agent's net surplus represents the agent's speculative surplus from the transaction.

Given (2), we may rewrite (1) as follows:

$$U(\theta,\theta)=[u(a^v_\theta)+\theta q(\theta)]-t_\theta(a^v_\theta). \tag{3}$$

Hence, type θ's utility when he truthfully reports his type is given by the difference between the gross surplus generated by his assigned contract and the transfer he pays to the principal. The gross surplus from a contract consists of the period 1 utility from the real action, $u(a^v_\theta)$, plus the speculative surplus, $q(\theta)$, weighted by the agent's degree of naivete, θ.

We have already noted that by the definition of exploitative contracts, $u(a^v_\theta)-t_\theta(a^v_\theta)<0$, and therefore by IR_θ, $u(a^u_\theta)-t_\theta(a^u_\theta)>0$. This implies the following result:

Observation 2. *$q(\theta)\geq 0$ for any type θ who chooses an exploitative contract.*
It follows that for any type θ who chooses an exploitative contract we can write

$$q(\theta)=[u(a^u_\theta)-v(a^u_\theta)]+[v(a^v_\theta)-u(a^v_\theta)]-\delta_\theta\geq 0. \tag{4}$$

Equation (4) allows us to simplify two of the constraints in the principal's maximization problem described in Observation 1.

Proposition 3.

(i) *Any contract that satisfies $IC_{\theta,\phi}$ for all ϕ implies that*

$$U(\theta,\theta)=\int_{\underline{\theta}}^{\theta}q(\theta)dx. \tag{5}$$

(ii) *The optimal menu of contracts satisfies that VR_θ binds for all $\theta\geq\underline{\theta}$.*

The representation of the incentive compatibility constraint given in (5) is obtained using standard tools of optimal mechanism design. By equating (5) with (3), we obtain the following expression for $t_\theta(a^v_\theta)$, the transfer that type θ pays to the principal:

$$t_\theta(a^v_\theta) = u(a^v_\theta) + \underline{\theta} q(\underline{\theta}) - \int_{\underline{\theta}}^{\theta} q(\theta) dx. \tag{6}$$

To understand the intuition for this expression, recall that both the IR and VR constraints are binding for the lowest exploited type $\underline{\theta}$. By solving these two equations, one obtains that the principal extracts $u(a^v_{\underline{\theta}}) + \underline{\theta} q(\underline{\theta})$ from this type. That is, the lowest exploited type pays his period 1 utility from the real action he chooses, plus the expected speculative surplus, where the expectation is taken with respect to that type's degree of naivete $\underline{\theta}$. To satisfy the incentive compatibility constraints, any type $\theta > \underline{\theta}$ must be left with some informational rent in order to induce him to choose the appropriate contract. By part (i) of Proposition 3, these rents are equal to $\int_{\underline{\theta}}^{\theta} q(\theta)$. It follows that the surplus the principal is able to extract from each $\theta \geq \underline{\theta}$ is given by (6).

Part (ii) of Proposition 3 implies that the *speculative surplus* in the transaction between the principal and type θ, first given by expression 2, has an alternative representation:

$$q(\theta) = [u(a^u_\theta) - v(a^u_\theta)] + [v(a^v_\theta) - u(a^v_\theta)]. \tag{7}$$

The interpretation of this alternative expression for the speculative surplus is as follows. The speculative surplus is the sum of the speculative gains that the two parties expect from their transaction. The speculation results from the parties' disagreement over the agent's second-period utility function and consequently over his second-period behaviour. The disagreement is that the agent believes he will choose a^u_θ and the principal believes that the agent will choose a^v_θ. If the agent is correct, then the agent's speculative gain is the difference between his first-period evaluation of a^u_θ and what the principal believes to be the agent's second-period evaluation of this action. If the principal is correct, then the principal's speculative gain is the difference between what he believes to be the agent's second-period evaluation of a^v_θ and the agent's first-period evaluation of this action.

The principal's objective is to find the value $\underline{\theta}$ and the profile of actions $(a^v_x)_{x \in [\underline{\theta},1]}$ that maximize

$$F(\underline{\theta}) \max_a [u(a) - c(a)] + [1 - F(\underline{\theta})] \cdot E[t_\theta(a^v_\theta) - c(a^v_\theta) | \theta \geq \underline{\theta}], \tag{8}$$

where $E[t_\theta(a^v_\theta) - c(a^v_\theta) | \theta \geq \underline{\theta}]$ is given by the expression

$$\frac{1}{1 - F(\underline{\theta})} \cdot \int_{\underline{\theta}}^{1} \{\psi(x) \cdot [\Delta^* + v(a^v_x) - u(a^v_x)] + u(a^v_x) - c(a^v_x)\} \cdot f(x) dx \tag{9}$$

and

$$\psi(x) = x - \frac{1 - F(x)}{f(x)}.$$

As is standard in the mechanism-design literature (see Krishna, 2002, p. 69), we impose the following assumption:

Condition 1. *The hazard rate $f(\cdot)/(1 - F(\cdot))$ is a continuously increasing function.*

This condition implies that $\psi(\cdot)$ is also increasing. We are now ready to give the "recipe" for solving the principal's optimization problem. The "recipe" is an application of textbook mechanism-design tools (see Krishna, 2002, p. 69). In the "textbook problem", the principal looks for *one* action–transfer pair for each type θ. In our original problem (see Observation 1), the principal looks for *two* action–transfer pairs: $(a_\theta^u, t_\theta(a_\theta^u))$ and $(a_\theta^v, t_\theta(a_\theta^v))$. However, we have already determined a_θ^u (see Proposition 2), and we established that VR_θ is binding (see Proposition 3); hence we can express $t_\theta(a_\theta^u)$ in terms of the other three variables. Therefore, we have reduced the problem to a "textbook problem", in which the principal only needs to look for one action–transfer pair $(a_\theta^v, t_\theta(a_\theta^v))$ for each exploited type θ.

The first step in the "recipe" yields the real action for each exploited type θ.

Proposition 4. *For every $\theta \geq \underline{\theta}$, a_θ^v is determined by the equation:*

$$a_\theta^v = \arg \max_{a \in [0,1]} \{\psi(\theta) \cdot [\Delta^* + v(a) - u(a)] + u(a) - c(a)\}. \tag{10}$$

This step is analogous to deriving the optimal quantity (or probability of attaining a good) in a standard mechanism-design exercise. It is achieved via point-by-point optimization, which we are able to do by imposing Condition 1 and by incorporating the necessary condition for incentive compatibility (captured by (5)) into the principal's objective function.

The next step in the "recipe" is the derivation of the cut-off type $\underline{\theta}$.

Proposition 5. *The cut-off $\underline{\theta}$ is determined by the equation:*

$$\max_a [u(a) - c(a)] = \max_a \{\psi(x) \cdot [\Delta^* + v(a) - u(a)] + u(a) - c(a)\}. \tag{11}$$

If the R.H.S. of the above equation is non-decreasing in a at its highest solution, then set $\underline{\theta}$ to be equal to that solution. Otherwise, set $\underline{\theta} = 1$.

The cut-off $\underline{\theta}$ is determined roughly as follows. In optimum, the benefit from marginally raising the cut-off is exactly offset by the loss. Suppose the current cut-off is at some ϕ. Then, if the principal raises the cut-off, he loses the surplus he could have extracted from ϕ, given that this is the lowest exploited type. This loss is equal to $\phi q(\phi) + u(a_\phi) - c(a_\phi)$. However, the benefit from raising the cut-off above ϕ is twofold. First, the principal's profit from type ϕ is $\max_a [u(a) - c(a)]$. This is the profit from every unexploited type. Second, by removing ϕ from the set of exploited types, the principal can raise the transfers received from all higher types. The increase in revenues from this raise turns out to be $\frac{1-F(\theta)}{f(\theta)} \cdot q(\theta)$. It follows that the principal should set $\underline{\theta} = \phi$ if

$$\phi \cdot q(\phi) + u(a_\phi) - c(a_\phi) = \max_a [u(a) - c(a)] + \frac{1 - F(\theta)}{f(\theta)} \cdot q(\phi),$$

which is the condition given in (11).

Finally, to solve for $t_{\underline{\theta}}(a_{\underline{\theta}}^v)$ and $t_{\underline{\theta}}(a_{\underline{\theta}}^u)$, we use the results that $IR_{\underline{\theta}}$ and $VR_{\underline{\theta}}$ are binding. To derive $t_\theta(a_\theta^v)$ and $t_\theta(a_\theta^u)$ for all $\theta > \underline{\theta}$, we equate (5) with (1) and use again the result that VR_θ is binding for all θ. We summarize these steps in the next proposition.

Proposition 6. *The transfers $t_\theta(a^*)$ and $t_\theta(a_\theta^v)$ in the optimal menu are given by the following set of equations. For $\theta = \underline{\theta}$:*

$$0 = \underline{\theta}[u(a^*) - t_{\underline{\theta}}(a^*)] + (1 - \underline{\theta})[u(a_{\underline{\theta}}^v) - t_{\underline{\theta}}(a_{\underline{\theta}}^v)]$$

$$v(a_{\underline{\theta}}^v) - t_{\underline{\theta}}(a_{\underline{\theta}}^v) = v(a^*) - t_{\underline{\theta}}(a^*).$$

For $\theta > \underline{\theta}$:

$$\theta[u(a^*) - t_\theta(a^*)] + (1-\theta)[u(a^v_\theta) - t_\theta(a^v_\theta)] = (\theta - \underline{\theta})\Delta^* + \int_{\underline{\theta}}^{\theta}[v(a^v_x) - u(a^v_x)]dx$$

$$v(a^*) - t_\theta(a^*) = v(a_\theta) - t_\theta(a_\theta).$$

By our characterization of $\underline{\theta}$, it must be the case that $\psi(\underline{\theta}) \geq 0$. Because $\psi(0) < 0$ and $\psi(\cdot)$ is a continuous function, we obtain the following corollary.

Corollary 2. $\underline{\theta} > 0$.

The meaning of this result is that the non-exploitative contract is always chosen by a positive measure of agents. It cannot be the case that all agents in the population are exploited. This is analogous to exclusion results in standard price discrimination models. Just as in standard models, the principal's imperfect information leads him to exclude low types, so in the present model, the principal's imperfect information leads him to avoid speculating with sophisticated types.

Our last result in this sub-section establishes the necessary and sufficient condition for the inclusion of exploitative contracts in the optimal menu. This condition depends only on the specification of u and v; it is independent of the distribution of types.

Proposition 7. *The optimal menu of contracts must contain at least one exploitative contract if and only if*

$$\max_a[v(a) - c(a)] + \max_a[u(a) - v(a)] > \max_a[u(a) - c(a)]. \tag{12}$$

The intuition for this result is as follows. The maximal possible profit from an exploitative contract is the profit that can be obtained when facing only fully naive agents. The L.H.S. of (12) represents this profit. Therefore, the principal offers an exploitative contract in the optimal menu if and only if this profit exceeds the profit from the optimal non-exploitative contract.

4.3. *Welfare analysis*

In this sub-section, we evaluate the agents' welfare under the principal's optimal menu of contracts. Recall that welfare judgements are conceptually problematic when agents have dynamically inconsistent preferences: should we apply the point of view of the first- or second-period selves? The non-common-prior assumption adds another ambiguity to the welfare analysis: whose prior should we apply in the evaluation of *ex ante* welfare? In line with our interpretation of the model, we adopt the principal's prior belief. Therefore, we evaluate the agent's welfare, assuming that he ends up choosing according to v.

In two-period settings such as ours, economists tend to prefer the point of view of the first-period self. Our reference to contracts with $t_\theta(a^v_\theta) > u(a^v_\theta)$ as exploitative implicitly reflects this practice. Let us therefore begin with the first-period self's point of view. The next result confirms the intuition that more naive agents are more heavily exploited by the principal.

Proposition 8. *In the optimal menu, $u(a^v_\theta) - t(a^v_\theta)$ is non-increasing in θ.*

By comparison, welfare analysis from the second-period self's point of view is equivocal. In fact, second-period welfare—measured by $v(a^v_\theta) - t(a^v_\theta)$—may actually rise with the agent's naivete. For instance, if u *decreases* with a and υ *increases* with a, then a sophisticated agent would choose a non-exploitative contract that commits him to a low action, whereas a more naive agent would choose an exploitative contract that induces a high real action. Hence, the naive agent may be left with a higher second-period surplus than the sophisticated agent. See the example in Section 5.1 for such a case. In contrast, whenever the optimal menu satisfies that $\upsilon(a_\theta^\upsilon) - u(a_\theta^\upsilon)$ weakly decreases with θ, Proposition 8 immediately implies that $\upsilon(a_\theta^\upsilon) - t(a_\theta^\upsilon)$ weakly decreases with θ.

5. EXAMPLES

The following set of examples serves a double role. First, it demonstrates how to apply the "recipe" of Propositions 4–6. Second, the optimal menu in each example resembles a real-life contractual arrangement. Of course, given the simplicity of our model, the examples are highly stylized and are not meant to serve as descriptive models of the concrete economic environments referred to. Throughout the sub-section, we assume $F(\theta) = \theta$, for simplicity.

5.1. *Credit card "teaser" rates*

Consider a liquidity-constrained consumer, whose only means of making purchases at period 2 is a credit card. Let $a \in [0, 1]$ denote the amount of period 2 consumption. Let u and υ represent the consumer's net value of period 2 consumption, from his period 1 and period 2 perspectives. Let $u(a) = a$ for $a \leq \frac{1}{2}$ and $u(a) = \frac{1}{4} + \frac{1}{2}a$ for $a > \frac{1}{2}$. Let $\upsilon(a) = a$ for every $a \in [0, 1]$. That is, the consumer's two selves agree on the net value of consumption up to $a = \frac{1}{2}$, but they diverge for $a > \frac{1}{2}$: the second-period self is tempted to increase the consumption level.

A credit card company is able to extend credit to the consumer. The credit is directly tied to consumption. An interest rate schedule is thus a function $t : [0, 1] \to \mathbb{R}$, such that $t(a)$ is the amount paid to the credit card company for a consumption level of a. The company's cost of providing credit is $c(a) = c \cdot a$, where $\frac{1}{2} < c < 1$. The company issues a menu of interest rate schedules.

The non-exploitative contract induces the action $\arg\max_a [u(a) - c(a)] = \frac{1}{2}$. Let $r > 1$. A contract that enforces this action is:

$$t^{\text{soph}}(a) = \begin{cases} a & \text{for } a \leq \frac{1}{2} \\ \frac{1}{2} + r \cdot \left(a - \frac{1}{2}\right) & \text{for } a > \frac{1}{2} \end{cases}.$$

Let us turn to the exploitative contracts. Because $\max_a [u(a) - \upsilon(a)] = 0$, we have

$$\psi(\theta) \cdot [\Delta^* + \upsilon(a) - u(a)] + u(a) - c(a) = \begin{cases} (1 - c)a & \text{if } a \leq \frac{1}{2} \\ (\theta - c)a + \frac{1}{2}(1 - \theta) & \text{if } a > \frac{1}{2} \end{cases}. \tag{13}$$

Hence,

$$\max_a \{\psi(\theta) \cdot [\Delta^* + \upsilon(a) - u(a)] + u(a) - c(a)\} = \begin{cases} \frac{1}{2}(1 - c) & \text{if } \theta < c \\ \frac{1}{2}\theta + \frac{1}{2} - c & \text{if } \theta \geq c \end{cases}.$$

Since $\max_a [u(a) - c(a)] = \frac{1}{2}(1 - c)$, it follows that $\underline{\theta} = c$ (note that at this value, the R.H.S. of (20) is non-decreasing as required). This implies that $a_\theta^\upsilon = 1$ for all exploited types $\theta \geq c$. Because every exploited type is assigned the same real action, each of these types must also be assigned the same real transfer. In addition, by Proposition 2(ii), we can also set $a_\theta^u = \frac{1}{2}$ for each $\theta \geq c$. We can therefore design an optimal menu with a single exploitative contract, which we denote by t^{naif}. To construct this contract, we first compute the imaginary and real transfers, $t^{\text{naif}}(\frac{1}{2})$ and $t^{\text{naif}}(1)$, as implied by $IR_{\underline{\theta}}$ and $VR_{\underline{\theta}}$:

$$c\left[\frac{1}{2}-t^{\text{naif}}\left(\frac{1}{2}\right)\right]+(1-c)\left[\frac{3}{4}-t^{\text{naif}}(1)\right]=0 \qquad (IR_{\theta=c})$$

$$1-t^{\text{naif}}(1)=\frac{1}{2}-t^{\text{naif}}\left(\frac{1}{2}\right). \qquad (VR_{\theta\geq c})$$

Solving these two equations yields $t^{\text{naif}}(\frac{1}{2})=\frac{1}{4}c+\frac{1}{4}$ and $t^{\text{naif}}(1)=\frac{1}{4}c+\frac{3}{4}$. These transfers can be induced by the following transfer function:

$$t^{\text{naif}}(a)=\begin{cases}\frac{1}{4}c+\frac{1}{2}a & \text{for } a\leq\frac{1}{2}\\ \frac{1}{4}c-\frac{1}{4}+a & \text{for } a>\frac{1}{2}\end{cases}.$$

Thus, the optimal menu consists of a pair of "three-part tariffs". Both tariffs charge a higher interest rate on high consumption levels ($a>\frac{1}{2}$). However, the non-exploitative contract charges a prohibitively high rate for $a>\frac{1}{2}$, while the exploitative contract's rates are not so high as to deter the consumer from choosing $a>\frac{1}{2}$. Relatively naive consumers ($\theta\geq\frac{1}{2}$) underestimate the likelihood that they would consume more than $\frac{1}{2}$ and hence prefer the contract with the lowest rate on small consumption levels. Sophisticated consumers, on the other hand, prefer the contract with the higher interest rate because it acts as a commitment device.

Our example suggests that naive consumers underestimate the extent of their borrowing and hence overrespond to "teaser rates" on small-size loans. DM formalized a similar idea using a (β,δ) model of time preferences. They showed that consumer's naivete regarding their future discount rate may explain why credit card companies charge an interest rate above marginal cost together with a low (possibly negative) initial fee.

There is some empirical evidence for these claims. Ausubel (1991, 1999) argues that consumers who are sensitive to changes in the interest rate are those "who do not intend to borrow on their accounts but find themselves doing so anyway". Ausubel presents evidence suggesting the presence of such consumers and argues that their naivete helps explain the high profits made by credit card issuers. Comparing two loans with the same post-introductory interest rate, Ausubel (1999) shows that the acceptance rate of an offer of 4.9% for 6 months was significantly higher than that of 6.9% for 9 months. However, the data on the actual account usage of these consumers reveal that most of the consumers who chose the former contract would have been better off with the latter. Ausubel concludes that "many consumers systematically underestimate the extent of their current and future credit card borrowing" (see Shui and Ausubel, 2004, for further evidence).

5.2. *Negative option offers*

Consider a situation in which an agent exhibits a "reference-point effect": his evaluation of available actions depends on whether he is *choosing* a contract or *cancelling* (or modifying) an existing contract (for experimental evidence, see Shafir, 1993). When choosing a contract, the agent cannot find a good enough reason for taking any action $a>0$; hence, $u(a)=0$ for all a. After signing a contract, the agent's reference point changes, and he strictly prefers "higher" actions, such that $v(a)=-(1-a)^2$. Assume zero costs.

The following scenario fits this specification. The agent considers insuring himself against a set of possible damages. The range of possible actions represents the amount of coverage offered by some insurance policy. In the absence of insurance, the agent believes that the contingencies specified in the policy are so unlikely that thinking about them is not worth his while. However, once the agent obtains some insurance plan, he starts viewing the contingencies as realistic. Consequently, the agent prefers greater coverage.

Let us characterize the optimal menu. The optimal non-exploitative contract is trivial because $\max_a u(a) = 0$. Therefore, the only non-trivial contracts in the optimal menu are exploitative. Turning to exploitative contracts, for $a \in [0,1]$, the expression $\psi(\theta)\cdot[\Delta^* + v(a) - u(a)] + u(a)$ is positive and monotonically increasing in a if and only if $\psi(\theta) > 0$. Therefore, for every exploited type θ, $a^v = 1$. In addition, for every exploited type θ, $a^u_\theta = \arg\max_a[u(a) - v(a)] = 0$. Because the real action is identical for all exploited types, the optimal menu contains a single exploitative contract, denoted by $t^{\text{naif}}(\cdot)$.

The remaining steps are straightforward. We use the fact that $IR_{\underline{\theta}}$ and $VR_{\underline{\theta}}$ are binding to calculate $t^{\text{naif}}(a^v)$ and $t^{\text{naif}}(a^u)$ as functions of $\underline{\theta}$. We then derive the cut-off by maximizing expected profits. The optimal cut-off turns out to be $\underline{\theta} = \frac{1}{2}$, and $t^{\text{naif}}(1) = \frac{1}{2}$ while $t^{\text{naif}}(0) = -\frac{1}{2}$. These transfers can be implemented by the following transfer function:

$$t(a) = \begin{cases} 2a - \frac{1}{2} & \text{if } a \le \frac{1}{3} \\ \frac{1}{2}a & \text{if } a > \frac{1}{3} \end{cases}.$$

All types $\theta < \frac{1}{2}$ choose to opt out, whereas all types $\theta > \frac{1}{2}$ choose the exploitative contract and end up insuring $a = 1$.

The exploitative contract has two distinctive features. First, it offers a gift of $\frac{1}{2}$ upon signing the contract, coupled with the possibility to cancel the transaction while keeping the gift. Second, it offers a quantity discount. Naive agents believe that with high probability, their reference point will not change and that consequently, they will cancel the deal and earn the gift. However, once they sign the contract and their reference point changes, the quantity discount impels them to choose full coverage. Sophisticated agents anticipate the change in their reference point and avoid signing the contract. A real-life example of such a contract is the practice of credit card companies to offer a cheque to clients who are willing to accept a free trial of a credit card protection plan. The client is entitled to cash the cheque even if he cancels the plan within the free trial period.[3]

The practice of offering a service or a product to a consumer, and requiring him to explicitly reject it in order to avoid a charge, is called a "negative option offer". Although it is not legally defined, the term is often applied to situations where some kind relationship already exists between the buyer and the seller. This marketing technique was first used by book clubs and naturally extended to CD and DVD/video clubs. Today, negative option offers are commonly used in other industries such as telephone and cable companies, who use these offers to enroll subscribers in optional services. For a detailed analysis of negative option offers, see Lamont (1995).

5.3. *Casino players' clubs*

Consider an agent who contemplates gambling at a casino. Initially, the agent prefers to gamble for small amounts. However, once he starts gambling, his satiation point increases. Let a represent the amount of second-period gambling, and suppose that $u(a) = \frac{1}{2} - \left|a - \frac{1}{2}\right|$ and $v(a) = \frac{1}{2}a$. We abstract from the inherent uncertainty involved in gambling activity, thus ignoring the possibility that the gambler's tastes may depend on the outcome of his initial gamble. Assume zero costs.

3. Note that there are other types of dynamic inconsistency, other than reference-point effects, which could explain this contract. For example, the agent may have self-control problems which cause him to postpone cancelling the deal.

By following the recipe, we can obtain the following optimal menu. All types below $\frac{5}{6}$ choose the non-exploitative contract, which may be written as follows:

$$t^{\text{soph}}(a) = \begin{cases} \frac{1}{2} & \text{if } a \leq \frac{1}{2} \\ a & \text{if } a > \frac{1}{2} \end{cases}.$$

This contract acts as a commitment device that induces the action $a = \frac{1}{2}$. All types above $\frac{5}{6}$ choose the exploitative contract, which may be written as follows:

$$t(a) = \begin{cases} \frac{3}{4}a & \text{if } a \leq \frac{1}{2} \\ \frac{1}{8} + \frac{1}{2}a & \text{if } a > \frac{1}{2} \end{cases}.$$

This contract induces the imaginary and real actions $a^u = \frac{1}{2}$ and $a^v = 1$.[4]

This menu resembles a contractual arrangement adopted by some real-life casinos. The gambler may bet according to the standard fare, consisting of a minimal bet and no discounts on larger amounts. Alternatively, he may join a *players' club*, which subsidizes both modest- and large-stakes gambling.[5] One way to interpret the players' club is as an exploitative contract. The discounts on modest-stakes gambles attract the naive customer, who falsely believes that he will not be tempted by the subsidies on large-stakes gambling. Sophisticated customers avoid joining the players' club.[6]

5.4. *Discussion*

Some features in the menus of the above examples can be rationalized with a standard model. A model of "customer poaching" can rationalize offers of free gifts as in our second example. A model in which consumers lack information about the quality of a particular good or service may help explain offers of free-trial periods. The casino players' clubs may be viewed as a scheme designed to dissuade existing clients from switching to competitors, much like "frequent flyer" clubs in the airline industry.[7]

Other features of the above menus cannot be obtained within a standard framework. The credit cards menu in our first example offers two contracts, one of which dominates another (by offering uniformly lower rates).[8] It seems impossible to rationalize choice of the dominated contract, except as a commitment device. Evidence for such anomalous choices of loans is provided in White and Munger (1971), who show that 29% of the consumers in their sample deliberately chose a high interest loan over a low interest loan.[9]

It is also difficult to explain with a standard model why a firm would tie a gift (*i.e.* $t(a) < 0$ for some a) to one contract but not to another. Consider a variant on the example analysed in Subsection 5.2, in which $\max_a u(a) > 0$. In this case, the non-exploitative contract would be non-trivial. However, it would not involve any gift to the consumer, whereas the exploitative contract

4. Given t^{naif}, the agent is indifferent among all actions $a \geq \frac{1}{2}$. Therefore, $a^v = 1$ is only weakly sustained as the real action. It is easy to get rid of indifferences, using a slightly more complicated three-part tariff.

5. For example, see http://www.venetian.com/casino/vpc.cfm and http://www.playersclub.com/Default.asp.

6. As pointed out by DM, the policy of subsidizing accommodation, adopted by Las Vegas hotel-casinos, may be interpreted as another means of exploiting naive gamblers. Our example ignores this aspect of "Las Vegas pricing".

7. However, note that at the Venetian players' club, betting for the modest amount of $500, already entitles the gambler to some benefits.

8. The Providian menu described by DM has a similar feature: the Visa Gold Preferred card charges a higher interest rate than Visa Gold Prestige (for other examples, see the menus in www.expertcredit.com).

9. Of course, in order to determine whether these consumers were deliberately choosing a commitment device, we need to examine how they would have repayed a low-interest loan. Shui and Ausubel (2004) address this concern.

would continue to contain a gift. If the objective of the gift is customer poaching, why not extend it to all customers? A standard explanation based on customer poaching would therefore have to add another source of heterogeneity among consumers.

All of our examples involved a unique exploitative contract. This is only because we have chosen linear utility functions to facilitate the analysis. Notice that from (10), it follows that if u and v are non-linear, there may be a continuum of exploitative contracts reflecting fine discrimination among the naive types.

6. TWO-PART AND THREE-PART TARIFFS

In Sub-section 4.2, we derived the optimal menu without restricting the domain of contracts. This allowed us to reduce a contract to the pair of transfers assigned to the imaginary and real actions by imposing prohibitively large fines on all other actions. Since real-life contracts typically lack this feature, the question is whether the optimal menu can be implemented by more prevalent pricing schemes.

Among the common forms of monopolistic pricing, the simplest one is the two-part tariff. This is a linear price schedule consisting of a fixed fee k and a per-unit price p such that $t(a) = k + pa$. We may therefore ask whether there is any loss of generality in restricting attention to contracts of this form. The answer is affirmative. For example, the exploitative contract of Sub-section 5.2 cannot be implemented by a two-part tariff. To see why, note that in order for $t(0) = -\frac{1}{2}$ and $t(1) = \frac{1}{2}$, the two-part tariff would have to be $t(a) = -\frac{1}{2} + a$. But then, the optimal second-period action for an agent who accepts this contract would be $a = \frac{1}{2}$ and not $a = 1$ (the latter being the real action dictated by the exploitative contract in the optimal menu).

A slightly more complicated pricing scheme, which is also used in practice, is the *three-part* tariff, consisting of a fixed fee k, a per-unit price p_1 up to some threshold a^*, and a per-unit price p_2 above it. Examples of three-part tariffs include price schedules used by cell phone providers, cable TV providers (where a is the amount of premier channels), and insurance companies (where a may be the term period in a term life insurance policy).

While the optimal menu in Sub-section 5.2 is not implementable with two-part tariffs, it can be implemented with three-part tariffs. This raises the question of whether we can extend this conclusion to a larger class of problems. In other words, is there a broad class of c, u, and v functions, for which the optimal menu is implementable by a menu of three-part tariffs but not by a menu of two-part tariffs?

Assume the cost function is convex and differentiable and let u and v satisfy the following conditions: (i) they are increasing and strictly concave, (ii) $u(0) = v(0)$, (iii) $u'(a) < v'(a) < \infty$ for all $a \geq 0$, and (iv) $\max_{a \in [0,1]}[u(a) - c(a)]$ has an interior solution.

Proposition 9. *If conditions* (i)–(iv) *hold, then the optimal menu can be implemented by a menu of three-part tariffs, but it cannot be implemented by a menu of two-part tariffs.*

Thus, our result provides a rationale for three-part tariffs. As we show in the proof, they are instrumental for designing both exploitative and non-exploitative contracts.

7. CONCLUDING REMARKS

7.1. *The role of dynamic inconsistency*

Our model relaxes two standard assumptions: common priors and dynamic consistency. As argued in Section 1, the reason we do so is that violations of these assumptions are likely to coexist in many contracting relations. However, it is important to understand the role played by each individual assumption. One merit of our model is that it allows us to isolate the effects of each assumption.

Consider an alternative model in which agents are time consistent, but they still hold biased beliefs about their future tastes. In this model, agents evaluate second-period actions according to a state-dependent utility function, which takes the form $u(\cdot)$ with probability θ and the form $\upsilon(\cdot)$ with probability $1-\theta$. Hence, the first-period expected utility of an agent of type θ, who pretends to be of type ϕ, is as follows:

$$U(\phi,\theta) = \theta[u(a_\phi^u) - t_\phi(a_\phi^u)] + (1-\theta)[\upsilon(a_\phi^\upsilon) - t_\phi(a_\phi^\upsilon)], \tag{14}$$

where $a_\phi^u \in \arg\max_{a\in[0,1]}[u(a) - t_\phi(a)]$ and $a_\phi^\upsilon \in \arg\max_{a\in A}[\upsilon(a) - t_\phi(a)]$. Compared with the expression for $U(\phi,\theta)$ in the model of Section 3, the agent's evaluation of the υ-optimal action at period 1 is made according to υ, rather than according to u.

Time consistency by itself does not preclude the use of exploitative contracts involving speculation. The motive to speculate stems from the non-common-prior assumption. However, the definition of an exploitative needs to be modified: a contract t_θ is exploitative if $t_\theta(a_\theta^\upsilon) > \upsilon(a_\theta^\upsilon)$.

The important difference between the two models is that under time consistency, the principal can no longer offer agents a commitment device that brings all types to their reservation utility. The reason for this is that agent's expected utility from a contract that induces $a^u = a^\upsilon$ is not independent of θ. This means that the principal may be unable to contract with sophisticated types without worrying about the informational externalities that they exert on naive types. Hence, the optimal menu may exclude some types, unlike the case of dynamically inconsistent agents.

Note that the dynamic element in this alternative model is superfluous from a formal point of view. Hence, this model may be interpreted as a static principal–agent problem, in which the two parties hold different priors about some state of nature. A general analysis of such a model lies beyond the scope of this paper.

7.2. *The role of non-common priors*

Let us now examine what happens if we retain dynamic inconsistency and restore common priors. Consider an alternative, more standard model, in which θ is the *objective probability* that the agent's second-period utility will be u. As in the model of Section 3, the agent has dynamically inconsistent preferences: his first-period self has utility u, and he believes that his second-period self's utility will be u with probability θ and υ with probability $1-\theta$. As in the model of Section 3, θ is the agent's type, and the principal cannot observe it. However, in contrast to the model of Section 3, the principal agrees that θ is the true probability of u.

In this case, it is straightforward to show that the principal's optimal menu consists of a single contract, with $t(a) = u(a)$ for $a \in \arg\max_a[u(a) - c(a)]$, and $t(\cdot) = \infty$ otherwise. That is, there is no discrimination between types and there is no exploitation. To see why, suppose that the principal knew θ. Then, his maximization problem would be:

$$\max_{a^u,a^\upsilon,t^u,t^\upsilon} \theta\cdot[t^u - c(a^u)] + (1-\theta)\cdot[t^\upsilon - c(a^\upsilon)]$$

subject to the constraints:

$$\begin{aligned} \theta[u(a^u) - t^u] + (1-\theta)[u(a^\upsilon) - t^\upsilon] &\geq 0 \\ u(a^u) - t^u &\geq u(a^\upsilon) - t^\upsilon \\ \upsilon(a^\upsilon) - t^\upsilon &\geq \upsilon(a^u) - t^u. \end{aligned}$$

The solution to this maximization problem is $a^u = a^v \in \arg\max_a[u(a) - c(a)]$ and $t^u = t^v = u(a^u)$. Given that the principal's first best contract is independent of θ, there will be *no* discrimination when the principal does not observe θ. Thus, the assumption of non-common priors is crucial for the effects analysed in this paper.

What would happen if we restored both common priors and dynamic consistency, while retaining the sequential contracting structure? In particular, modify the present common-prior model such that the agent's preferences are time consistent. In other words, the agent's first-period evaluation of a contract $t(\cdot)$ is given by (14). It is easy to show that for a large class of u and v, the optimal contract would involve no discrimination: the first best contract would be identical for all types. We refer the reader to a small literature (Armstrong, 1996; Courty and Li, 2000) that analyses more complicated models, in which discrimination does take place in optimum.

7.3. *Heterogeneity in stability of preferences*

As mentioned in Section 1, this paper focuses on price discrimination based *solely* on differences in the agents' degree of naivete. We have therefore completely abstracted from other sources of heterogeneity. In particular, we have assumed that all agents experience a change in their utility. However, it may be interesting to extend our model, such that agents will differ along two dimensions: their *subjective* probability that their tastes will change and the *objective* probability of this event. We shall now briefly discuss one such model.

Suppose that the principal believes that each agent is characterized by a pair (θ, p), where θ is the agent's prior belief that his preferences will not change, and p is the actual probability of this event. While there is asymmetric information with regard to the prior θ, neither the principal nor the agent observe p. Therefore, the agent's type continues to be represented by the scalar θ. The principal believes that the pair (p, θ) is distributed on $[0, 1]^2$ according to some joint cdf.[10]

In this extended model, the principal's objective function is given by:

$$E_{p,\theta}\{p \cdot [t_\theta(a^u_\theta) - c(a^u_\theta)] + (1 - p) \cdot [t_\theta(a^v_\theta) - c(a^v_\theta)]\}.$$

Note that a^u_θ enters the objective function. This is in contrast to the original model: see Observation 1. However, all the constraints given in Observation 1 remain the same as in the original model: IR and IC rely on the agent's own subjective belief θ while UR and VR pertain to the second period, when the agent already knows his preferences over actions.

The fact that a^u is no longer an imaginary action—there is a positive probability that it will actually be taken—has several implications. First, the principal has weaker incentives to engage in speculation: he can no longer afford to attract naive agents by charging low (and possibly negative) transfers for choosing a^u_θ. This would be reflected in a higher cut-off $\underline{\theta}$. Second, Proposition 2(i) ceases to hold, that is, the principal is no longer free to set a^u_θ without worrying about its direct impact on his profits. This means that for general u and v, we may not be able to solve the principal's problem using standard techniques. In a simple setting in which there are only two available actions, one would still be able to solve for the optimal menu. In this case, the only difference relative to the original model (in which $p = 1$ for all types) is a higher cut-off $\underline{\theta}$.

10. Previous works have discussed (in the context of (β, δ) preferences) the possibility that the population of agents will contain agents who share the same perception of their dynamic inconsistency, while differing in the actual bias that they exhibit (see DM and Gilpatric, 2003).

7.4. *Alternative notions of partial naivete*

We formalized the agent's "degree of naivete" as a biased prior belief. An alternative measure of naivete is introduced in O'Donoghue and Rabin (2001) for the (β, δ) model (this is the measure of naivete used in DM). According to this measure, a partially naive agent believes that his present bias is given by a parameter $\hat{\beta} \in (\beta, 1)$. Hence, the higher an agent's $\hat{\beta}$, the higher his degree of naivete.

Our approach has three distinguishing features. First, it allows us to disentangle the effect of the agent's naivete from the effect of his dynamic inconsistency. Second, it automatically implies that the agent's utility is linear in his type, and this facilitates the adaptation of standard contract-theoretic techniques. Third, it links our study to other models with non-common priors, which focus on issues such as overconfidence and overoptimism in contracting and bargaining models (Van den Steen, 2001; Benabou and Tirole, 2002; Yildiz, 2003; Fang and Moscarini, 2005).

Loewenstein, O'Donoghue and Rabin (2003) (henceforth, LOR) proposed an alternative definition. A partially naive agent believes that his second-period utility function will be a utility function w, which is in some sense "between" his current utility function u and his true future utility function v. The further away w is from v, the greater the agent's naivete. Thus, an agent who is partially naive in the LOR sense knows that his preferences will change, but is systematically wrong in estimating the magnitude of that change. In contrast, a partially naive agent in our model is not sure in the first period whether in the second period he will maximize u or v. He is uncertain as to whether his preferences will change, but he knows exactly what they could change into.

APPENDIX

Notation. To simplify the exposition, we shall use the following notation in some of the proofs. For any type θ who chooses a contract t_θ, let $D_\theta^u \equiv u(a_\theta^u) - t_\theta(a_\theta^u)$ and $D_\theta^v \equiv u(a_\theta^v) - t_\theta(a_\theta^v)$.

Proof of Proposition 1. Suppose the optimal menu includes both an exploitative contract and a non-exploitative contract. Let θ be a type that chooses the exploitative contract, and let ϕ be a type that chooses the non-exploitative contract. By $IC_{\theta,\phi}$,

$$\theta(D_\theta^u - D_\phi^u + D_\phi^v - D_\theta^v) \geq D_\phi^v - D_\theta^v. \tag{15}$$

Because t_θ is exploitative, but t_ϕ is not, $D_\theta^v < 0$ and $D_\phi^v \geq 0$. By (15), $D_\theta^u - D_\phi^u + D_\phi^v - D_\theta^v \geq 0$. Assume $\phi > \theta$. Then

$$\phi(D_\theta^u - D_\phi^u + D_\phi^v - D_\theta^v) > D_\phi^v - D_\theta^v.$$

But this violates $IC_{\phi,\theta}$, a contradiction. We conclude that $\theta > \phi$.

We next show that w.l.o.g. we may restrict attention to optimal menus that offer a single non-exploitative contract. Suppose the principal was facing a single agent with a known prior of θ. The optimal non-exploitative contract for this agent is a function $t : [0,1] \to \mathbb{R}$ that solves

$$\max_{a^v, t(a^v)} [t(a^v) - c(a^v)]$$

subject to the IR_θ, UR_θ, and VR_θ constraints, in addition to the constraint that $t(a^v) \leq u(a^v)$. A solution to this optimization problem is a function t^* satisfying $t^*(a) = u(a)$ if $a \in \arg\max_{a \in [0,1]}[u(a) - c(a)]$, and $t^*(a) = \infty$ for any other a.

Assume next that an agent's prior is not observed and the optimal menu contains a non-exploitative contract. We claim that the principal can achieve at least the same profits by offering only a single non-exploitative contract, t^*.

This is obvious if the menu contains no exploitative contracts. Suppose the optimal menu contains at least one exploitative contract. Let $\hat{t}$ denote the least profitable exploitative contract and let $\hat{\theta}$ denote the type who chooses this contract. Then

$$\hat{t}_{\hat{\theta}}(a_{\hat{\theta}}^v) - c(a_{\hat{\theta}}^v) > \max_a [u(a) - c(a)].] \tag{16}$$

Otherwise, the principal could strictly raise his expected profits by removing $\hat{t}$ from the menu and introducing t^* (in case t^* did not belong to the original menu), thus contradicting the assumption that the original menu is optimal. The reason is that t^* gives a zero surplus to *all* types. Therefore, if a type chose an exploitative contract other than $\hat{t}$ from the original menu, he would not choose t^* from the new menu.

It follows that every exploitative contract generates a higher profit than t^*. Consider amending the original menu by replacing all non-exploitative contracts with t^*. Since this contract gives a zero surplus to all types, any type who chose an exploitative contract from the original menu has no incentive to deviate to t^*. In addition, any type who chose a non-exploitative contract from the original menu and switches to an exploitative contract given the new menu will increase the principal's expected profits, by (16). This contradicts the assumption that the original menu is optimal. ‖

Proof of Remark 1. Assume the contrary, that is, that the optimal menu satisfies

$$U(\underline{\theta}, \underline{\theta}) = \underline{\theta}(D^u_{\underline{\theta}} - D^v_{\underline{\theta}}) - D^v_{\underline{\theta}} > 0.$$

Because $t_{\underline{\theta}}$ is exploitative, $D^v_{\underline{\theta}} < 0$. Hence, by $IR_{\underline{\theta}}$, $D^u_{\underline{\theta}} > 0$. This implies that

$$U(\underline{\theta}, \theta) = \theta(D^u_{\underline{\theta}} - D^v_{\underline{\theta}}) - D^v_{\underline{\theta}} > 0$$

for all $\theta > \underline{\theta}$. Hence, by $IC_{\theta,\underline{\theta}}$, $U(\theta,\theta) > 0$ for all $\theta > \underline{\theta}$. The principal can then modify all the exploitative contracts in the original menu as follows: for every $\theta \geq \underline{\theta}$, raise both $t(a^u_\phi)$ and $t(a^v_\phi)$ by $U(\underline{\theta}, \underline{\theta})$. This modification leaves all the IR, IC, UR, and VR constraints intact and generates a strictly higher profit, a contradiction. ‖

Proof of Proposition 2. The proof proceeds by a series of lemmas. ‖

Lemma 1. *$v(a^v_\theta) - u(a^v_\theta) > v(a^u_\theta) - u(a^u_\theta)$ for every exploited type θ.*

Proof. Assume not. Then, $v(a^v_\theta) - u(a^v_\theta) \leq v(a^u_\theta) - u(a^u_\theta)$. By VR_θ, $v(a^v_\theta) - t(a^v_\theta) \geq v(a^u_\theta) - t(a^u_\theta)$. Taken together, these inequalities imply that $D^v_\theta \geq D^u_\theta$. Because t_θ is exploitative, $D^v_\theta < 0$. But this implies that $U(\theta,\theta) < 0$, in contradiction to IR_θ. ‖

Lemma 2. *W.l.o.g. we may restrict attention to optimal menus with the property that for every $\theta \geq \underline{\theta}$,*

$$a^u_\theta \in \arg\max_a [u(a) - v(a)].$$

Proof. Consider some optimal menu with cut-off $\underline{\theta}$. Assume $a^u_\theta \notin \arg\max_a[u(a) - v(a)]$ for some $\theta \geq \underline{\theta}$. By Lemma 1, $a^v_\theta \notin \arg\max_a[u(a) - v(a)]$. The principal can then modify t_θ into t'_θ, such that a^u_θ is replaced with $a^* \in \arg\max_a[u(a) - v(a)]$ (*i.e.* impose an infinitely large fine on a^u_θ, whereas originally such a fine was imposed on a^*) and adjust $t'_\theta(a^*)$ such that $u(a^*) - t'_\theta(a^*) = u(a^u_\theta) - t_\theta(a^u_\theta)$. In this way, all the IR and IC constraints are preserved. The only thing that remains to be verified is that VR_θ is satisfied. That is, picking a^v_θ over a^* should be consistent with maximizing v in the second period. By assumption, this condition is satisfied by the original contract: $v(a^v_\theta) - t_\theta(a^v_\theta) \geq v(a^u_\theta) - t_\theta(a^u_\theta)$. Since $t'_\theta(a^*) = t_\theta(a^u_\theta) + u(a^*) - u(a^u_\theta)$ and $v(a^u_\theta) - u(a^u_\theta) > v(a^*) - u(a^*)$, VR_θ continues to hold. ‖

Lemma 3. *$t(a^u_\theta)$ is non-increasing in θ in the range $\theta \geq \underline{\theta}$.*

Proof. Rewriting $IC_{\theta,\phi}$ and $IC_{\phi,\theta}$ for $\underline{\theta} \leq \theta < \phi \leq 1$, we obtain

$$\frac{\theta}{1-\theta} D^u_\theta + D^v_\theta - \frac{\theta}{1-\theta} D^u_\phi - D^v_\phi \geq 0$$
$$\frac{\phi}{1-\phi} D^u_\phi + D^v_\phi - \frac{\phi}{1-\phi} D^u_\theta - D^v_\theta \geq 0.$$

For $\phi = 1$, we need not divide the inequality by $1 - \phi$.

By adding the above inequalities and using Lemma 2 (which implies that $a^u_\theta = a^u_\phi$), we get

$$[t_\theta(a^u_\theta) - t_\phi(a^u_\phi)]\left(\frac{\phi}{1-\phi} - \frac{\theta}{1-\theta}\right) \geq 0.$$

Hence, $t_\theta(a^u_\theta) \geq t_\phi(a^u_\phi)$. ‖

The following lemma will be instrumental in characterizing how $t_\theta(a^v_\theta) - c(a^v_\theta)$ changes with θ.

Lemma 4. *Suppose that $\theta_3 > \theta_2 > \theta_1$, and all three types choose exploitative contracts, such that t_{θ_2} and t_{θ_1} are distinct. If IC_{θ_1,θ_2} and IC_{θ_2,θ_1} hold, then $U(\theta_2,\theta_3) \geq U(\theta_1,\theta_3)$.*

Proof. Consider a menu of contracts that satisfies IC_{θ_1,θ_2} and IC_{θ_2,θ_1}. Rewriting IC_{θ_2,θ_1}, we obtain

$$\theta_2[(D^u_{\theta_2} - D^u_{\theta_1}) + (D^\upsilon_{\theta_1} - D^\upsilon_{\theta_2})] \geq D^\upsilon_{\theta_1} - D^\upsilon_{\theta_2}. \tag{17}$$

By Lemma 2, $u(a^u_{\theta_1}) = u(a^u_{\theta_2})$, and by Lemma 3, $t_{\theta_2}(a^u_{\theta_2}) \leq t_{\theta_1}(a^u_{\theta_1})$. Therefore, $D^u_{\theta_2} \geq D^u_{\theta_1}$. Suppose $D^\upsilon_{\theta_2} > D^\upsilon_{\theta_1}$. Then, $U(\theta_2,\theta_1) > U(\theta_1,\theta_1)$, violating IC_{θ_1,θ_2}. Hence, $D^\upsilon_{\theta_2} \leq D^\upsilon_{\theta_1}$, which implies that the L.H.S. of (17) is non-negative. Because $\theta_3 > \theta_2$,

$$\theta_3[(D^u_{\theta_2} - D^u_{\theta_1}) + (D^\upsilon_{\theta_1} - D^\upsilon_{\theta_2})] > D^\upsilon_{\theta_1} - D^\upsilon_{\theta_2}.$$

Rewriting this inequality, we obtain $U(\theta_2,\theta_3) \geq U(\theta_1,\theta_3)$. ‖

Using this lemma, we establish the following result.

Lemma 5. *$t(a^\upsilon_\theta) - c(a^\upsilon_\theta)$ is non-decreasing in θ.*

Proof. Assume the converse. Suppose that $t(a^\upsilon_\theta) - c(a^\upsilon_\theta)$ attains a global maximum at some $\theta < 1$. Modify the menu by omitting all t_ϕ for $\phi > \theta$. Since t_θ satisfies IR_θ, $U(\theta,\phi) > 0$. Given that $IC_{\theta,\omega}$ is satisfied for all $\omega < \theta$, it follows from Lemma 4 that every $\phi > \theta$ prefers t_θ to all t_ω. Clearly, this modification does not decrease expected profits. Let $\bar{\theta}$ be the highest type below θ for which $t(\cdot) - c(\cdot)$ attains a local maximum. If no such $\bar{\theta}$ exists, the proof is complete. If there exists such a $\bar{\theta}$, omit all contracts t_ϕ for $\phi \in (\bar{\theta},\theta)$ for which $t(a^\upsilon_\phi) - c(a^\upsilon_\phi) < t(a^\upsilon_{\bar{\theta}}) - c(a^\upsilon_{\hat{\theta}})$. By the same argument as above, $U(\bar{\theta},\phi) > 0$ and no type ϕ would want to deviate to a contract t_ω, $\omega < \bar{\theta}$. If any type ϕ who originally chose $t_{\bar{\theta}}$ now chooses a contract t_ω with $\omega > \bar{\theta}$, the principal's expected profit will only increase. Again, this modification leads to strictly higher expected profits. Iterating this argument, we eliminate all local maxima. ‖

This completes the proof of the Proposition. □

Proof of Proposition 3. *Proof of* (i). We adopt Krishna's (2002, pp. 63–66) derivation of incentive compatibility for direct mechanisms. For every $\theta \in [0,1]$, define $m(\theta) \equiv D^\upsilon_\theta$. The optimal menu is incentive compatible if for all types θ and ϕ,

$$V(\theta) \equiv \theta q(\theta) - m(\theta) \geq \phi q(\theta) - m(\phi).$$

By Observation 2, $q(\theta) \geq 0$ for all $\theta \geq \underline{\theta}$ (by the definition of $\underline{\theta}$, $q(\theta) = 0$ for all $\theta < \underline{\theta}$). Hence, the L.H.S. of the above inequality is an affine function of the true value θ. Incentive compatibility implies that for all $\theta \geq \underline{\theta}$,

$$V(\theta) = \max_{\phi \in [0,1]} \{\phi q(\theta) - m(\phi)\}.$$

That is, $V(\theta)$ is a maximum of a family of affine functions, and hence it is convex on $[\underline{\theta},1]$.[11]

Incentive compatibility is equivalent to the requirement that for all $\theta,\phi \in [\underline{\theta},1]$,

$$V(\phi) \geq V(\theta) + q(\theta)(\phi - \theta).$$

This implies that for all $\theta > \underline{\theta}$, $q(\theta)$ is the slope of a line that supports the function $V(\theta)$ at the point θ. Because $V(\theta)$ is convex, it is absolutely continuous and thus differentiable almost everywhere in the interior of its domain. Hence, at every point that $V(\theta)$ is differentiable, $V'(\theta) = q(\theta)$. Since $V(\theta)$ is absolutely continuous, we obtain that for all $\theta > \underline{\theta}$,

$$V(\theta) = V(\underline{\theta}) + \int_{\underline{\theta}}^{\theta} q(x)dx.$$

By Remark 1, $IR_{\underline{\theta}}$ is binding. Hence, $V(\underline{\theta}) = 0$, and we obtain (5).

Proof of (ii). For each type θ, denote the slack in the VR_θ constraint by δ_θ, that is,

$$\delta_\theta \equiv [\upsilon(a^\upsilon_\theta) - t_\theta(a^\upsilon_\theta)] - [\upsilon(a^u_\theta) - t_\theta(a^u_\theta)].$$

11. Because all types lower than $\underline{\theta}$ are assigned a non-exploitative contract, $V(\theta) = 0$ for all $\theta < \underline{\theta}$.

By (i), any contract that satisfies $IC_{\theta,\phi}$ for all ϕ must also satisfy

$$U(\theta,\theta) = \int_{\underline{\theta}}^{\theta} \left[\Delta^* + v\left(a_x^v\right) - u\left(a_x^v\right) - \delta_x\right] dx. \tag{18}$$

Assume that the optimal menu satisfies that $\delta_\theta > 0$ for some positive measure of types in $[\underline{\theta}, 1]$. Consider amending the menu by changing only $t_\theta(a_\theta^v)$ for all types $\theta \geq \underline{\theta}$ such that the new transfer is equal to $t_\theta(a_\theta^v) + \delta_\theta$, making the VR_θ constraint binding for all these types. Clearly, this change does not violate the UR_θ constraint of these types. It follows from (18) that the incentive compatibility constraints are not violated, and that this change only raises $U(\theta,\theta)$. Hence, the IR_θ constraint is also not violated. Since $t_\theta(a_\theta^v)$ has increased for a positive measure of types, while $c(a_\theta^v)$ has remained unchanged for each θ, the principal's expected profits increase, a contradiction. ‖

Proof of Proposition 4. Because there is no link between the different types (the incentive compatibility constraints have already been incorporated into the objective function), we can solve for a_θ^v by point-by-point optimization. ‖

Proof of Proposition 5. Substituting (9) into (8), we obtain the following expression for the principal's objective function:

$$F(\underline{\theta}) \max_a [u(a) - c(a)] + \int_{\underline{\theta}}^{1} [\psi(x)q(x) + u(a_x^v) - c(a_x^v)] \cdot f(x) dx, \tag{19}$$

where

$$\psi(x)q(x) + u(a_x^v) - c(a_x^v) \equiv \max_a [\psi(x) \cdot (\Delta^* + v(a) - u(a)) + u(a) - c(a)].$$

By (19), if

$$\psi(x)q(x) + u(a_x^v) - c(a_x^v) < \max_a [u(a) - c(a)] \tag{20}$$

for some x, then the principal is better off assigning x a non-exploitative contract. Hence, if (20) holds for all $x \in [0,1]$, then no exploitative contract should be offered, that is, $\underline{\theta} = 1$.

Suppose (20) holds with equality for some set of types. To determine the cut-off $\underline{\theta}$, we note that the L.H.S. of (20) is simply $t_x(a_x^v) - c(a_x^v)$, which by Proposition 2(iii), is non-decreasing in x. Therefore, $\underline{\theta}$ should be set equal to the highest type in this set, provided $\psi(x)q(x) + u(a_x^v) - c(a_x^v)$ is non-decreasing at that value. If no such solution exists, then there are no exploitative contracts in the optimal menu: $\underline{\theta} = 1$. Finally, note that since $\psi(0) < 0$, the function $\psi(\theta)q(\theta) + u(a_\theta^v) - c(a_\theta^v)$ cannot lie *above* $\max_a [u(a) - c(a)]$ for all θ. ‖

Proof of Proposition 6. By part (ii) of Proposition 3, $VR_{\underline{\theta}}$ is binding. By Remark 1, $IR_{\underline{\theta}}$ is binding. The combination of these two binding constraints allows us to solve for $t_{\underline{\theta}}(a^*)$ and $t_{\underline{\theta}}(a_{\underline{\theta}}^v)$. To solve for $t_\theta\left(a_\theta^v\right)$ and $t_\theta\left(a^*\right)$ for $\theta > \underline{\theta}$, we use the following pair of equations. The first equation is given by part (ii) of Proposition 3, which states that VR_θ is binding. The second equation is given by the two alternative formulations of $U(\theta,\theta)$: (1) and (5). ‖

Proof of Proposition 7. (*Sufficiency*) Suppose that (12) holds, and yet the optimal menu does not include an exploitative contract. Then, the principal extracts $\max_a [u(a) - c(a)]$ from all types, using a single contract with $t(a) = u(a)$ for $a \in \arg\max_a [u(a) - v(a)]$ (and infinite fines on all other actions).

To reach a contradiction, we show that the principal can add to the (singleton) menu an exploitative contract with the following properties:

(i) the contract extracts more than $\max_a [u(a) - c(a)]$ from all types who choose it,
(ii) the contract satisfies the UR and VR constraint of all types who choose it, and
(iii) there exists a strictly positive measure of types who prefer this contract to the non-exploitative contract described above.

For a given pair of utility functions, u and v, let $f(u,v)$ denote the L.H.S. of (12). By (12), there exists $\alpha \in (0,1)$ that solves $\alpha f(u,v) = \max_a [u(a) - c(a)]$. Let $\beta \in (\alpha, 1)$, $a^v \in \arg\max_a [v(a) - c(a)]$, $a^u \in \arg\max_a [v(a) - c(a)]$ and consider the following contract:

$$t(a) = \begin{cases} \beta f(u,v) + c(a^v) & \text{if} \quad a = a^v \\ v(a^u) - v(a^v) + \beta f(u,v) + c(a^v) & \text{if} \quad a = a^u \\ \infty & \text{if} \;\; a \notin \{a^u, a^v\} \end{cases}.$$

Note that this contract is exploitative since

$$\beta f(u,v) > \max_a[u(a)-c(a)] \geq u(a^v)-c(a^v).$$

By construction, (i) is satisfied. To establish (ii), note first that the VR constraint is binding by construction. To see why the UR constraint is satisfied, note that this constraint holds if $t(a^v)-t(a^u) \geq u(a^v)-u(a^u)$. But this inequality must hold since $t(a^v)-t(a^u) = v(a^v)-v(a^u)$, and $u(a^u)-v(a^u) \geq u(a^v)-v(a^v)$, by the definition of a^u.

Finally, to establish (iii), it suffices to show that there exists a strictly positive measure of types for whom the above exploitative contract gives a strictly positive surplus because the non-exploitative contract in the original menu gives zero surplus to all types. By the definition of a^u, a^v, and $t(a^u)$, $u(a^u)-t(a^u) = (1-\beta)f(u,v) > 0$. Hence, there exists $\theta^* < 1$, such that every type above θ^* earns a strictly positive surplus from $t(\cdot)$.

(*Necessity*) Suppose that the optimal menu contains an exploitative contract $t(\cdot)$. The profit generated by this contract must be at least as high as the highest profit that can be attained with a single non-exploitative contract:

$$t(a^v)-c(a^v) \geq \max_a[u(a)-c(a)], \tag{21}$$

where $a^v \in \arg\max_a[v(a)-t(a)]$. By Lemma 3, there is no loss of generality in restricting attention to optimal menus in which all exploitative contracts satisfy $a^u = \arg\max_a[u(a)-v(a)]$. By the VR constraint,

$$v(a^v)-v(a^u)+u(a^u)-c(a^v) \geq t(a^v)-t(a^u)+u(a^u)-c(a^v). \tag{22}$$

Because $t(\cdot)$ is exploitative and because IR_θ needs to hold for any type θ who chooses $t(\cdot)$, $u(a^u) > t(a^u)$. This, together with (21), implies that the R.H.S. of (22) is *strictly* greater than $\max_a[u(a)-c(a)]$. But since

$$\max_a[v(a)-c(a)]+\max_a[u(a)-v(a)] \geq v(a^v)-v(a^u)+u(a^u)-c(a^v),$$

inequality (12) must hold. ‖

Proof of Proposition 8. By the definition of exploitative contracts, $D^v_\theta < D^v_\phi$ for any $\theta \geq \underline{\theta} > \phi$. The proof of Lemma 4 establishes that for any $\theta > \phi \geq \underline{\theta}$, $D^v_\phi \geq D^v_\theta$. ‖

Proof of Proposition 9. Assume that the optimal menu is implementable by a menu of two-part tariffs. Let $k+pa$ be the two-part tariff that implements the non-exploitative contract. Let a^c be the interior solution to $\max_{a\in[0,1]}[u(a)-c(a)]$ (since u is strictly concave and c is convex, there exists a unique solution). In order for the agent to take the same action whether his utility is u or v , it must be the case that $p = u'(a^c) = v'(a^c)$. But this violates assumption (iii).

We now show that the optimal menu is implementable with three-part tariffs. Consider first the three-part tariff t^{soph} given by the threshold action $a^* = a^c$, the per-unit price $p_1 = u'(a^c)$ for $a \leq a^c$, the per-unit price $p_2 = v'(a^c)$ for $a > a^c$, and the fixed fee $k = u(a^c) - p_1 \cdot a^c$. It follows that

$$a^c = \arg\max_{a\in[0,1]}[u(a)-t^{\text{soph}}(a)] = \arg\max_{a\in[0,1]}[v(a)-t^{\text{soph}}(a)]$$

and $u(a^c)-t(a^c) = 0$. Hence, t^{soph} implements the non-exploitative contract in the optimal menu.

Suppose the optimal menu includes at least one exploitative contract. By Proposition 2(i), we may restrict attention to a menu in which $a^u_\theta = \arg\max_{a\in[0,1]}[u(a)-v(a)]$ for all $\theta \geq \underline{\theta}$. By our assumptions on u and v, $a^u_\theta = 0$. Consider the exploitative contract t_θ assigned by the optimal menu to some type $\theta \geq \underline{\theta}$. Let a^v_θ be the real action associated with this contract. Define $p^\theta_2 \equiv v'(a^v_\theta)$ and let b^u_θ denote the action satisfying $u'(b^u_\theta) = p^\theta_2$. By our assumptions on u and v, $b^u_\theta < a^v_\theta$. Define a^θ to be the action that solves the following equation:

$$v(a^v_\theta) = v(0)+p_1 \cdot a^\theta + p_2 \cdot (a^v_\theta - a^\theta), \tag{23}$$

where $p_1 > v'(0)$ and is set to be sufficiently large so that $a^\theta < b^u_\theta$ (note that by our assumptions on v, $a^\theta > 0$). Finally, define k^θ by the equation

$$k^\theta + p_1 a^\theta + p_2 \cdot (a^v_\theta - a^\theta) = t_\theta(a^v_\theta).$$

We claim that the three-part tariff $(k^\theta, p^\theta_1, p^\theta_2, a^\theta)$ implements t_θ. First, by construction it replicates $t_\theta(a^v_\theta)$. In addition, by substituting (23) into the definition of k^θ, we obtain that $k^\theta = t_\theta(a^v_\theta) - v(a^v_\theta) + v(0) = t_\theta(0)$, where the last equality follows from the result (see Proposition 3(ii)) that VR_θ is binding.

Second, we need to show that given this tariff, $a = 0$ is the best action for an agent with utility u. Because $p_1 > v'(0)$ and $v'(0) > u'(0)$, it follows that

$$\arg\max_{a\in[0,a^\theta]}[u(a)-k^\theta - p_1 \cdot a] = 0.$$

On the other hand,

$$\arg\max_{a\in[a^\theta,1]}[u(a)-k^\theta-p_1a^\theta-p_2\cdot(a-a^\theta)]=b^u_\theta.$$

Because $a^\theta<b^u_\theta$, our assumptions on u and υ imply that

$$\upsilon(a^\upsilon_\theta)-p_2\cdot a^\upsilon_\theta>u(b^u_\theta)-p_2\cdot b^u_\theta. \tag{24}$$

Subtracting $[a^\theta(p_1-p_2)+k^\theta]$ from both sides of (24) yields

$$\upsilon(a^\upsilon_\theta)-k^\theta-p_1a^\theta-p_2\cdot(a^\upsilon_\theta-a^\theta)>u(b^u_\theta)-k^\theta-p_1a^\theta-p_2\cdot(b^u_\theta-a^\theta).$$

Because VR_θ is binding, the L.H.S. of this inequality is simply $\upsilon(0)-k^\theta$. But since $u(0)=\upsilon(0)$, we have

$$u(0)-k^\theta\geq u(b^u_\theta)-k^\theta-p_1a^\theta-p_2\cdot(b^u_\theta-a^\theta).$$

It remains to show that given $(k^\theta,p^\theta_1,p^\theta_2,a^\theta)$, the optimal action for an agent with a utility of υ is a^υ_θ. First, note that

$$\arg\max_{a\in[a^\theta,1]}[\upsilon(a)-k^\theta-p_1a^\theta-p_2\cdot(a-a^\theta)]=a^\upsilon_\theta$$

and

$$\arg\max_{a\in[0,a^\theta]}[\upsilon(a)-k^\theta-p_1a]=0.$$

Second, equation (23) implies that

$$\upsilon(a^\upsilon_\theta)-k^\theta-p_1a^\theta-p_2\cdot(a^\upsilon_\theta-a^\theta)=\upsilon(0)-k^\theta$$

such that a^υ_θ is optimal in the entire domain of actions. ‖

Acknowledgements. We are most grateful to Juuso Välimäki and two anonymous referees for helping us improve this paper. We also thank Mark Armstrong, Heski Bar-Isaac, Eddie Dekel, Stefano DellaVigna, Susanna Esteban, Ronny Razin, and seminar participants at Berkeley, LSE, Penn State, Rutgers, and Tel Aviv University for helpful comments. Financial support from the U.S.–Israel Binational Science Foundation, Grant No. 2002298 is gratefully acknowledged.

REFERENCES

AMADOR, M., WERNING, I. and ANGELETOS, G.-M. (2004), "Commitment vs. Flexibility" (Mimeo, Standford GSB).

ARMSTRONG, M. (1996), "Nonlinear Pricing with Imperfectly Informed Consumers" (Mimeo, Nuffield College).

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

AUSUBEL, L. M. (1999), "Adverse Selection in the Credit Card Market" (Mimeo, University of Maryland).

BENABOU, R. and TIROLE, J. (2002), "Self-Confidence and Personal Motivation", *Quarterly Journal of Economics*, **117**, 871–915.

COURTY, P. and LI, H. (2000), "Sequential Screening", *Review of Economic Studies*, **67**, 697–718.

DELLAVIGNA, S. and MALMENDIER, U. (2004), "Contract Design and Self-Control: Theory and Evidence", *Quarterly Journal of Economics*, **119**, 353–402.

DELLAVIGNA, S. and MALMENDIER, U. (2006), "Paying Not to Go to the Gym", *American Economic Review* (in press).

ESTEBAN, S., MIYAGAWA, E. and SHUM, M. (2003), "Nonlinear Pricing with Self-Control Preferences" (Mimeo, Columbia University).

FANG, H. and MOSCARINI, G. (2005), "Morale Hazard", *Journal of Monetary Economics*, **52**, 749–777.

GABAIX, X. and LAIBSON, D. (2005), "Shrouded Attributes, Consumer Myopia, and Information Suppression in Competitive Markets" (Mimeo, Harvard University).

GILPATRIC, S. (2003), "Present-Biased Preferences, Self-Awareness and Shirking in a Principal–Agent Setting" (Mimeo, University of Tennessee).

GUL, F. and PESENDORFER, W. (2001), "Temptation and Self-Control", *Econometrica*, **69**, 1403–1435.

KRISHNA, V. (2002) *Auction Theory* (San Diego: Academic Press).

LAIBSON, D. and YARIV, L. (2004), "Safety in Markets: An Impossibility Theorem for Dutch Books" (Mimeo, UCLA).

LAMONT, D. D. (1995), "Negative Option Offers in Consumer Service Contracts: A Principled Reconciliation of Commerce and Consumer Protection", *UCLA Law Review*, **42**, 1315–1388.

LOEWENSTEIN, G., O'DONOGHUE, T. and RABIN, M. (2003), "Projection Bias in Predicting Future Utility", *Quarterly Journal of Economics*, **118**, 1209–1248.

O'DONOGHUE, T. and RABIN, M. (1999*a*), "Incentives for Procrastinators", *Quarterly Journal of Economics*, **114**, 103–124.

O'DONOGHUE, T. and RABIN, M. (1999*b*), "Doing It Now or Later", *The American Economic Review*, **89**, 103–124.

O'DONOGHUE, T. and RABIN, M. (2001), "Choice and Procrastination", *Quarterly Journal of Economics*, **116**, 121–160.

PICCIONE, M. and RUBINSTEIN, A. (2003), "Modeling the Economic Interaction of Agents with Diverse Abilities to Recognize Equilibrium Patterns", *Journal of European Economic Association*, **1**, 212–223.

RUBINSTEIN, A. (1993), "On Price Recognition and Computational Complexity in a Monopolistic Model", *Journal of Political Economy*, **101**, 473–484.

SARAFIDIS, Y. (2004), "Inter-temporal Price Discrimination with Time Inconsistent Consumers" (Mimeo, INSEAD).

SHAFIR, E. (1993), "Choosing vs. Rejecting: Why Some Options are Both Better and Worse than Others", *Memory & Cognition*, **21**, 546–556.

SHUI, H. and AUSUBEL, L. (2004), "Time Inconsistency in the Credit Card Market" (Mimeo, University of Maryland).

SPIEGLER, R. (2004), "Competition Over Agents With Boundedly Rational Expectations" (Mimeo, Tel-Aviv University).

VAN DEN STEEN, E. (2001), "Essays on the Managerial Implications of Differing Priors" (Ph.D. Dissertation, Stanford Business School).

WHITE, J. J. and MUNGER, F. W. (1971), "Consumer Sensitivity to Interest Rates: An Empirical Study of New-Car Buyers and Auto Loans", *Michigan Law Review*, **69**, 1207–1258.

YILDIZ, M. (2003), "Bargaining Without a Common Prior—An Immediate Agreement Theorem", *Econometrica*, **71**, 793–811.