DISCUSSION PAPER SERIES

No. 4573

CONTRACTING WITH DIVERSELY NAIVE AGENTS

Kfir Eliaz and Ran Spiegler

INDUSTRIAL ORGANIZATION

Centre for Economic Policy Research

www.cepr.org

Available online at: **www.cepr.org/pubs/dps/DP4573.asp**

ISSN 0265-8003

CONTRACTING WITH DIVERSELY NAIVE AGENTS

Kfir Eliaz, New York University and CEPR
Ran Spiegler, Tel Aviv University

Discussion Paper No. 4573
August 2004

Centre for Economic Policy Research
90–98 Goswell Rd, London EC1V 7RR, UK
Tel: (44 20) 7878 2900, Fax: (44 20) 7878 2999
Email: cepr@cepr.org, Website: www.cepr.org

CEPR Discussion Paper No. 4573

August 2004

ABSTRACT

Contracting with Diversely Naive Agents*

A principal contracts with agents who have diverse abilities to forecast changes in their future tastes. While the principal knows that the agent's tastes are changing, the agent believes that with probability θ, their future preferences will be identical to their present preferences. The principal does not observe θ, but knows the probability distribution from which it is drawn. Thus, the agent's prior probability θ is their 'private type', and the principal has to offer a menu of contracts in order to screen the agent's type. We provide a full characterization of the principal's optimal menu. The results allow us to interpret some real-life contractual arrangements in a variety of examples.

JEL Classification: L12 and L14
Keywords: contracts, dynamic inconsistency, naivety and non-common priors

Kfir Eliaz
Economics Department
New York University
269 Mercer Street
New York, NY 10003
USA
Tel: (1 212) 998 8912
Fax: (1 212) 995 4186
Email: kfir.eliaz@nyu.edu

For further Discussion Papers by this author see:
www.cepr.org/pubs/new-dps/dplist.asp?authorid=153104

Ran Spiegler
School of Economics
Tel Aviv University
Tel Aviv 69978
ISRAEL
Email: rani@post.tau.ac.il

For further Discussion Papers by this author see:
www.cepr.org/pubs/new-dps/dplist.asp?authorid=148706

*We thank Heski Bar-Isaac, Eddie Dekel, Susanna Esteban, Ronny Razin and Yona Rubinstein for helpful comments. Financial support from the US-Israel Binational Science Foundation, Grant No. 2002298 is gratefully acknowledged.

Submitted 13 July 2004

Contracting with Diversely Naive Agents*

Kfir Eliaz[†] and Ran Spiegler[‡]

July 13, 2004

Abstract

A principal contracts with agents who have diverse abilities to forecast changes in their future tastes. While the principal knows that the agent's tastes are changing, the agent believes that with probability θ, his future preferences will be identical to his present preferences. The principal does not observe θ, but he knows the probability distribution from which it is drawn. Thus, the agent's prior probability θ is his "private type", and the principal has to offer a menu of contracts in order to screen the agent's type. We provide a full characterization of the principal's optimal menu. The results allow us to interpret some real-life contractual arrangements in a variety of examples.

1 Introduction

In standard contract-theoretic models, agent types differ in their preferences. Usually, the agent's type is a parameter that characterizes his willingness to pay or his 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.

The literature has discussed a variety of sources of dynamically inconsistent preferences. First, the mere passage of time may affect an individual's rate of substitution between consumption at times t and $t+1$. This source of time-inconsistency has been extensively studied

*We thank Heski Bar-Isaac, Eddie Dekel, Susanna Esteban, Ronny Razin and Yona Rubinstein for helpful comments. Financial support from the US-Israel Binational Science Foundation, Grant No. 2002298 is gratefully acknowledged.

[†]Dept. of Economics, NYU. 269 Mercer St., New York, NY 10003, E-mail: kfir.eliaz@nyu.edu, URL: http://homepages.nyu.edu/~ke7

[‡]School of Economics, Tel Aviv University, Tel Aviv 69978, Israel. E-mail: rani@post.tau.ac.il. URL: http://www.tau.ac.il/~rani

in the economic literature. For a recent collection of works on this topic, see Lowenstein, Read and Baumeister (2002). Reference point effects may also result in dynamically inconsistent preferences. For example, an agent's evaluation of a transaction after he accepts it may differ from his ex-ante evaluation (due to an "endowment effect" or a "status quo effect"). For psychological studies on this subject, see Kahneman, Knetch and Thaler (1991), Tversky and Kahneman (1991), Kahneman (1992) and Shafir (1993). Preferences may also change because of addiction (see Gruber and Koszegi (2001), Gul and Pesendorfer (2004) and the references therein).

In recent years, economists have become increasingly interested in the implications of dynamically inconsistent preferences on economic behavior in a variety of contexts. A number of authors (O'Donoghue and Rabin (1999b), Gilpatric (2003), Sarafidis (2004), and especially, DellaVigna and Malmendier (2003,2004)) have begun exploring the question of how a rational agent (whether a profit-maximizing firm or a mechanism designer) would want to 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. The agent has quasi-linear utility over action-transfer pairs. However, his 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 the variety of sources for changing tastes alluded to above.

The following are some situations with dynamically inconsistent agents, which will serve as leading examples in this paper:

- The principal offers a variety of cable TV packages, ranging from the basic package to an extended package containing all channels. In the absence of cable TV, the agent has a low willingness to pay for the extended package. However, once he is exposed to it, he becomes hooked and raises his willingness to pay for this high-end service.

- The principal lends the agent a sum of money in period 1. In period 2, the agent chooses how to allocate the repayment between that period and period 3. In period 1, the agent would like to repay all the loan in period 2. However, in period 2 he prefers to delay some of the repayment.

- The principal provides cell phone services. Prior to owning a cell phone, the agent believes that he will only use it in emergencies. As a cell phone owner, the agent is tempted to use the cell phone for other, non-emergency purposes.

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. 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 revenue.

From a formal point of view, our model is non-standard in the sense that it departs from the common-prior assumption. The agent's type is simply his prior belief over his second-period preferences. Except for the type $\theta = 0$, the agent's belief differs from the principal's. We interpret the model as if the principal's prior is correct and the agent's prior is biased. This judgment is arbitrary, as far as analyzing 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., his vulnerability to temptation, propensity to procrastinate, or sensitivity to reference points) than some of the agents themselves. Moreover, there is a growing body of evidence suggesting that individuals form biased assessments of the probabilities of future events that impact their preferences (see Weinstein (1980), Buunk *et al.* (1990), Helgeson and Taylor (1993) and Weeks *et al.* (1998)).

It is important to note that the state of nature, over which the principal and the agents disagree in terms of their priors, is the agent's second-period utility function. We assume that this utility function is *not verifiable*, hence the principal and the agent cannot make infinite bets on the realization of this state.

Drawing on the terminology of O'Donoghue and Rabin (1999c), 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 formalization of the agent's "degree of naiveté" as a biased prior belief is a distinguishing feature of our model. O'Donoghue and Rabin (2001) and Lowenstein, O'Donoghue and Rabin (2003) proposed alternative notions of biased perception of future tastes. In Section 5.4 we discuss the difference between these notions. Our approach has three merits. First, it allows us to disentangle the effect of the agent's naiveté 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 over-confidence and over-optimism in contracting relations and bargaining interactions (Benabou and Tirole (2002), Fang and Moscarini (2003), Van den Steen (2001), Yildiz (2003)).

How does a partially naive, dynamically inconsistent agent perceive 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. That is, the agent's dynamic inconsistency is reflected by the fact that he evaluates future actions according to his period 1 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.

To better understand the notion of exploitative contracts, consider a cable TV provider facing an agent who is interested in the basic cable package. The agent is currently unwilling to pay the provider's rate for the extended package. However, once he is exposed to the extended package, he changes his taste and his willingness to pay for that package rises above its rate. A partially naive agent believes that with some probability, he will not change his taste for the extended package. Suppose the cable TV provider offers a "free trial" period for the extended package. The "imaginary" action is to accept the offer and cancel the extended package at the end of the free trial period. The "real" action is to accept the offer and not cancel the extended package. Therefore, a relatively naive agent who accepts the offer ends up being exploited by the cable TV provider: he pays more than his current willingness to pay for the extended package. A relatively sophisticated agent will refuse the offer. This example suggests that the principal may wish to design a menu of contracts, some of which target the relatively naive types and some target the relatively sophisticated types.

Section 3 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 (or "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 extracts their highest first-period willingness to pay, $\max_a u(a)$, by forcing them to consume $\arg\max_a u(a)$. The relatively naive types (high θ) choose exploitative contracts.

Discrimination among the naive types. The number of exploitative contracts in the menu depends on the nature of the conflict of interests between the agent's first-period and second-period selves - specifically, the slopes of u and v. For example, when $u'(a)/v'(a)$ is constant, there is at most one exploitative contract. On the other hand, when $u'(a)/v'(a)$ is decreasing (reflecting situations like consumption of an addictive good, in which the agent considers a marginal unit to be increasingly harmful ex-ante and increasingly tempting once being exposed to it), there may be a continuum of exploitative contracts.

Monotonicity of the payment scheme. The actual payment to the principal increases with the agent's naiveté. However, the transfer associated with the imaginary action decreases with agent's naiveté.

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

Our results reveal a special structure of the optimal menu: it offers two distinct types of contracts. The first type of contract essentially commits the agent to a specific action. This is the non-exploitative contract, which is directed at the relatively sophisticated types. The second type of contracts endows the agent with more flexibility, and it resembles a non-linear pricing scheme. In standard models of price discrimination, non-linear pricing emerges as a way to discriminate between agents who differ in their utility from actions: different types will differ in their consumption levels. In our model, the objective of non-linear pricing schedules is to lure the naive types: one consumption level is an imaginary action, for which the payment is small (and possibly negative), while another consumption level is the real action, for which the payment is large.

This characterization is consistent with some real-life contractual arrangements. Internet and cable TV providers, as well as credit fraud protection agencies, offer free trial periods, which in essence reward a low level of consumption. Phone companies offer pre-paid packages side by side with variable-rate packages having zero marginal rate for low-intensity usage.[1] Book publishers offer readers two alternative price schedules: the standard cover price schedule, and a "book club" schedule that offers significant discounts for a limited number of purchases. CD and DVD clubs operate in a similar way.

We interpret these institutions as means of discriminating between sophisticated and naive types. The naive types are tempted by the rewards of the variable-rate schedule, believing they will stick to low levels of consumption. However, they end up with high levels of consumption. In contrast, relatively sophisticated types choose to avoid the free trials, book clubs and variable-rate schedules, and opt for the rigid contract that serves as a commitment device. Of course, there are alternative explanations for these institutions, which are based on more standard models with common priors and time-consistent preferences (see Section 3.3). However, these explanations require a distinct model for each institution, whereas our interpretation is based on a single perspective.

In Section 3.2 we provide a detailed algorithm for computing the optimal menu. The algorithm is an adaptation of standard tools of optimal mechanism design. Our main analytic task is to show that indeed, one can apply these tools in our framework. One merit of this technique is that is provides an elegant representation of the speculative component of the transactions between the principal and the agents. 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. Our algorithm provides a simple expression for the "surplus of the speculative transaction" between the parties.

As mentioned earlier, our notion of partial naiveté allows us to isolate the role played by the dynamic inconsistency assumption. We do this in Section 4 where we study a variant of our model in which agents are dynamically consistent. What this means is that each agent evaluates his indirect utility from a contract as the expectation of a state-dependent utility function: in one state his utility is u, in another state his utility is v. We interpret this model as a situation with heterogeneous priors, in which the agents are uncertain of their future tastes, but the principal believes he knows what those future tastes will be.

The characterization of the optimal menu in the modified model with dynamically *consistent* agents differs from our original characterization in several important respects:

(i) There is at most one exploitative contract (no discrimination among the naive types).
(ii) Sophisticated, unexploited types may exert an informational externality on relatively naive types. Therefore, sophisticated types may be crowded out.

[1] See DellaVigna and Malmendier (2004) for details.

(iii) Alternatively, when sophisticated types are not crowded out, the non-exploitative contract offered to them may be flexible and involve an "imaginary" action.

In Section 5 we briefly discuss other variations on the model. First, we examine a model in which the agent's type is the true probability distribution over his second-period utility. We show that in this model, the optimal menu consists of a single contract chosen by all types. This exercise demonstrates that the non-common priors assumption is necessary for our analysis. Second, we discuss the effect of introducing competition among principals. Third, we examine the implication of relaxing our assumption that the principal can perfectly verify the agent's second-period action. Finally, we compare our notion of partial naiveté to alternative definitions in the literature.

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 (1999b) 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 work that is most closely related to the present study is DellaVigna and Malmendier (2004) (DM, henceforth). In this paper, a monopolistic firm offers a two-part tariff to an agent whose preferences are given by the (β, δ) functional form. 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 the true value. 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 paper and DM is of course our 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 naiveté. 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 impose no a priori restrictions on the space 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 naiveté. The optimal menu contains a variety of contractual forms, where different forms target agents with different degrees of naiveté.

Our paper extends this literature in further dimensions. First, while the above papers focus on a particular type of dynamic inconsistency, namely (β, δ) preferences, we allow for a more general class of dynamic inconsistencies. Therefore, we are able to accommodate phenomena which are not captured by the (β, δ) model. Second, we employ a different formalization of partial naiveté, which allows us to distinguish in a clear manner between the effects that arise from dynamically inconsistent preferences and the effects that arise from the agent's naiveté.

Esteban et. al. (2003) study a model of contracting with agents having self-control problems, taking a different approach. Specifically, they analyze a non-linear pricing problem, in which consumers' preferences are given by the functional form introduced by Gul and Pesendorfer (2001). Different consumer types share the same functional form, but differ in the parameter values. Recall that in the Gul-Pesendorfer model, there is no dynamic inconsistency. Rather, the agent's preferences are defined over an extended consequence space, such that the description of a consequence contains not only the actual consumption decision, but also another, maximally tempting consumption option.

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 solve this problem by reformulating it as a mechanism-design problem where the agent's private type is the shock to his future utility. The authors take two approaches to modeling 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. In both cases the agent's private type is a standard preference parameter. In addition, in the model with time-inconsistent preferences, the disagreement between the present and future selves (captured by the parameter β) is known to the social planner.

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.

Other papers have studied the question of whether market interaction between profit-maximizing firms and perceptually biased agents will result in the exploitation of the latter. Laibson and Yariv (2004) analyze an intertemporal competitive economy, in which firms compete in "dutch books" over agents with inconsistent preferences. Spiegler (2003,2004) analyzes competition over agents with bounded ability to perceive stochastic environments.

2 The model

A principal faces a continuum of agents. The principal can provide each agent with the opportunity to choose an action from the set $[0,1]$. However, 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 believes that with probability θ his second-period utility from actions will remain u, and with probability $1-\theta$ it will change into v. Agents differ in their value of θ. We assume that θ is distributed according to a continuous c.d.f. $F(\theta)$ with support $[0,1]$. The principal does not observe θ. We assume that all of the above is common knowledge between the principal and the agents.

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. For example, consider the two types $\theta = 0$ and $\theta = 1$. Type 0 is a "fully sophisticated" agent, who believes that his preferences will change with certainty. Given a contract $t(\cdot)$, he will choose the action $a^v \equiv \arg\max_a [v(a) - t(a)]$ in period 2. Therefore, type 0's first-period indirect utility from $t(\cdot)$ is $u(a^v) - t(a^v)$. In contrast, type 1 is "fully naive": he believes his utility function will remain $u(\cdot)$. Given the contract $t(\cdot)$, he will choose the action $a^u \equiv \arg\max_a [u(a) - t(a)]$ in period 2. Therefore, type 1's first-period indirect utility from $t(\cdot)$ is $u(a^u) - t(a^u)$. More generally, the first-period indirect utility of a "partially naive" type $\theta \in [0,1]$ from a contract $t(\cdot)$ is

$$\theta \left[u\left(a^{u}\right)-t\left(a^{u}\right)\right]+(1-\theta)\left[u\left(a^{v}\right)-t\left(a^{v}\right)\right]$$

The principal does not have any intrinsic preference over the agent's second-period actions. His objective is to maximize expected revenue. Because the principal does not observe θ, he offers the agents a menu of contracts, where a menu is set of transfer functions $t:[0,1] \rightarrow \mathbb{R}$. Given a menu, each agent type picks his optimal contract. 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 $t_{\phi}:[0,1] \rightarrow \mathbb{R}$. The principal's problem is then to find the optimal set of functions $\{t_{\theta}(a)\}_{\theta \in [0,1]}$.

Discussion

Formally, we analyze 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. For example, the principal may have learned (from experience, marketing research, consulting, etc.) that consumers experience a systematic endowment effect, or that they may become used to a good or service once they try it for free. The consumer, on the other hand, may have no experience with the particular good or service being provided, and he has not conducted extensive psychological research on consumer behavior. Alternatively, the consumer may be aware that other consumers are subject to these psychological effects, yet he is confident that he is impervious to them. Hence, we interpret our model as a situation in which the principal knows the true state of nature, whereas the agent holds an erroneous belief.[2]

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 – i.e., the agents' second-period utility function – is not verifiable. Only the actions that the agents take are verifiable.

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. This change may be caused by several factors. Preferences may change simply because of the passage of time, as in the (β, δ) models of time-inconsistency. Preferences may also change because of some action taken by the principal at the time in which the contract is signed. For example, the principal may provide a good or service for a "free trial" period, and the agent may become "addicted", or experience an "endowment effect". To allow for a variety

[2]Because it is not common knowledge that the principal knows the state of nature, this is not a principal-agent model with an informed principal.

of sources of dynamic inconsistency, and to keep the model as simple as possible, we simply assume that the agents' preferences change from one period to the next (conditional on signing the contract) without explicitly modeling the source of that change.

Finally, we comment on our choice of a continuum action space. In solving for the optimal menu of contracts, we would like to enable the principal to carry out as fine a discrimination as possible between the different agent types. Because the type space is the unit interval, a finite action space might artificially induce "bunching" of types.

3 Characterizing the optimal menu

In this section we provide a full characterization of the optimal menu of contracts. In Section 3.1, we state a number of important general properties of the optimal menu. In Section 3.2, we show how to adapt standard tools of mechanism design to derive an algorithm for computing the optimal menu. In Section 3.3, we present a number of examples.

3.1 Qualitative features of the menu

The following simple example illustrates how a principal may discriminate between agents based on their degree of naiveté.

Example 3.0. Suppose there are only two types of agents: $\theta = 1$ (the fully naive type) and $\theta = 0$ (the fully sophisticated type). Assume $v(\cdot)$ and $u(\cdot)$ are increasing with $v(a) > u(a)$ for all $a > 0$. Let $v(0) = u(0) = 0$. Clearly, the principal cannot extract more than $u(1)$ from the sophisticated agent. Hence, the best contract for this agent is a transfer function $t_0(\cdot)$ satisfying

$$t_0(a) = \begin{cases} u(1) & if \quad a = 1 \\ \infty & if \quad other \end{cases}$$

The question is, can the principal extract more than $u(1)$ from the naive agent? In particular, can he extract $v(1)$? Consider the contract $t_1(\cdot)$,

$$t_1(a) = \begin{cases} v(1) - 2\varepsilon & if \quad a = 1 \\ \varepsilon & if \quad a = 0 \\ \infty & if \quad other \end{cases}$$

where $\varepsilon > 0$. This contract promises the agent a free gift if he chooses the zero action, but demands a payment if $a = 1$ is chosen instead. The naive agent, who believes his period 2 utility is given by $u(a)$, expects to get the free gift from $t_1(\cdot)$. However, conditional on choosing this contract, the agent prefers to choose $a = 1$ in period 2 and receive a net surplus of 2ε. Knowing this, the sophisticated agent associates a negative indirect utility with the contract $t_1(\cdot)$ (recall that $u(a) < v(a)$ for all $a > 0$). Since $t_0(\cdot)$ guarantees zero surplus to both agents, the naive agent ranks $t_1(\cdot)$ above $t_0(\cdot)$, while the sophisticated agent has the opposite ranking. It follows that by offering the menu $\{t_0(\cdot), t_1(\cdot)\}$ and setting ε to be arbitrarily close to zero, the principal can extract self 1's highest willingness to pay from the sophisticated agent, and close to self 2's highest willingness to pay from the naive agent.

While Example 3.0 is quite simplistic, it conveys the basic intuition underlying the design of the optimal menu. First, the contract intended for each type must yield that type a non-negative indirect utility. Second, the most the principal can extract from sophisticated types is $\max_a u(a)$, the highest willingness to pay of self 1. To extract more than this amount from the relatively naive types, the principal must offer them contracts that promise some sort of a gift. However, to avoid having to pay for gifts, the principal needs to make sure that in period 2 the agent would prefer to choose a different action than the one that promises a gift. This intuition is made precise in the following observation.

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

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

subject to the constraints,

$$\theta\left[u(a_\theta^u) - t_\theta(a_\theta^u)\right] + (1-\theta)\left[u(a_\theta^v) - t_\theta(a_\theta^v)\right] \geq 0 \qquad (IR_\theta)$$

$$\theta\left[u(a_\theta^u) - t_\theta(a_\theta^u)\right] + (1-\theta)\left[u(a_\theta^v) - t_\theta(a_\theta^v)\right] \geq \theta\left[u\left(a_\phi^u\right) - t_\phi(a_\phi^u)\right] + (1-\theta)\left[u\left(a_\phi^v\right) - t_\phi(a_\phi^v)\right] \qquad (IC_{\theta,\phi})$$

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

$$a_\theta^u = \arg\max_{a\in A}\left[u(a) - t_\theta(a)\right] \qquad (UR_\theta)$$

$$a_\theta^v = \arg\max_{a\in A}\left[v(a) - t_\theta(a)\right] \qquad (VR_\theta)$$

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 than 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 v (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 v. This is precisely the condition VR_θ.

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 one 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*". Without loss of generality, 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, i.e.,

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

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 $t_\theta\left(a_\theta^v\right) > u(a_\theta^v)$.*

Definition 1 formalizes the notion that an exploitative contract extracts more than the willingness to pay of a fully sophisticated agent. For instance, the contract chosen by the naive type, $t_1(\cdot)$, in Example 3.0 is exploitative: self 1's willingness to pay for $a = 1$ is $u(1)$, while $t_1(\cdot)$ induces $a = 1$ and extracts close to $v(1) > u(1)$. The contract chosen by the sophisticated type, $t_0(\cdot)$, is non-exploitative: it extracts exactly the agent's willingness to pay in period 1.

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, every non-exploitative contract satisfies* $t_\theta(a_\theta) = \max_a u(a)$ *and* $t_\theta(a) = \infty$ *for every other action* 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: it demands a payment of $\max_a u(a)$ for choosing $\arg\max_a u(a)$ and sets a prohibitive fine on all other actions. Only exploitative contracts can generate a higher revenue for the principal. Note that the above 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.

Second, there exists a cutoff $\underline{\theta}$, 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.*

We conclude that relatively sophisticated types, who choose the non-exploitative contract, exert no informational externality on the relatively naive types, who choose exploitative contracts. This is a non-standard effect: in conventional models of price discrimination "low types" exert an informational externality on "high types". That is, in those models high types receive informational rents because they have an incentive to mimic the low types. In contrast, our model has the feature that all types share the same first-period preferences: they only differ in their beliefs regarding future preferences. Because the non-exploitative contract forces the agent to play $\arg\max_a u(a)$ in period 2, all types derive the same indirect utility from this contract. In particular, all types have an indirect utility of zero from this contract. Hence, "high types" (the naive types) do not require an informational rent to prevent them from choosing the non-exploitative contract intended for the "low types" (the sophisticated agents). As we shall see in Section 4, this will no longer be the case when agents are dynamically *consistent*.

This no-externality effect has the following important implication:

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

Thus, as long as there is a surplus in the interaction with fully sophisticated types, the principal's lack of information regarding the agent's type does not lead to crowding out of sophisticated types. On the other hand, we shall see below that if $\max_a u(a) = 0$, crowding out of low types is possible.

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^v_θ. Intuitively, one would expect that the greater the agent's naiveté, the higher the gift he would require, but also the higher the payment he would actually end up making. This intuition is verified in our next result.

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.* $a^u_\theta = \arg\max_a(u(a) - v(a))$*,*
(ii) $t(a^u_\theta)$ *is non-increasing in* θ *in the range* $\theta > \underline{\theta}$*, and*
(iii) $t(a^v_\theta)$ *is non-decreasing in* θ*.*

To understand the intuition for *(i)*, recall that the transfer associated with the imaginary action is not part of the principal's revenue. The role of the imaginary action is to allow the principal to extract the highest amount possible 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 $\arg\max_a(u(a) - v(a))$ in period 2 would be willing to pay a high amount to choose a different action when he discovers that his period 2 utility is $v(\cdot)$ 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 $u(a^v) - t(a^v)$ and $u(a^u) - t(a^u)$, where a^v 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 $u(a^u_\phi) - t(a^u_\phi) \geq u(a^v_\theta) - t(a^v_\theta)$. 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 more standard single-crossing argument. Suppose $t(a^v_\theta)$ has a single peak at $\theta < 1$. Then the principal can increase his revenue 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.

3.2 Computing the 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^* and the difference between its first and second period evaluation, $u(a^*) - v(a^*)$, by Δ^*. 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:

$$U(\theta,\theta) = \theta\left[u(a^u_\theta) - t_\theta(a^u_\theta) - u(a^v_\theta) + t_\theta(a^v_\theta)\right] + \left[u(a^v_\theta) - t_\theta(a^v_\theta)\right] \quad (1)$$

We then define

$$q(\theta) \equiv \left[u(a^u_\theta) - t_\theta(a^u_\theta)\right] - \left[u(a^v_\theta) - t_\theta(a^v_\theta)\right] \quad (2)$$

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 of nature would occur, which the principal had not anticipated. In this state he expects to obtain a net surplus (as evaluated in period 1) of $u\left(a_\theta^u\right) - t_\theta\left(a_\theta^u\right)$. 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:

$$U\left(\theta, \theta\right) = \left[u\left(a_\theta^v\right) + \theta q\left(\theta\right)\right] - t_\theta\left(a_\theta^v\right) \tag{3}$$

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, $u\left(a_\theta^v\right)$, and the "speculative surplus" $q\left(\theta\right)$, appropriately weighted by the agent's degree of naiveté, θ. For the agent to accept an exploitative contract, the speculative surplus should be non-negative:

Observation 2. $q\left(\theta\right) \geq 0$ *for any type* θ *who chooses an exploitative contract.*

We have already noted that by the definition of exploitative contracts, $u\left(a_\theta^v\right) - t_\theta\left(a_\theta^v\right) < 0$, and therefore, by IR_θ, $u\left(a_\theta^u\right) - t_\theta\left(a_\theta^u\right) > 0$. This immediately implies Observation 2. It follows that for any type θ who chooses an exploitative contract we can write:

$$q\left(\theta\right) = \left[u\left(a_\theta^u\right) - v\left(a_\theta^u\right)\right] + \left[v\left(a_\theta^v\right) - u\left(a_\theta^v\right)\right] - \delta_\theta \geq 0 \tag{4}$$

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*

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

(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 $t_\theta\left(a_\theta^v\right)$, the transfer that type θ pays to the principal:

$$t_\theta\left(a_\theta^v\right) = u\left(a_\theta^v\right) + \theta q\left(\theta\right) - \int_{\underline{\theta}}^{\theta} q\left(\theta\right) dx \tag{6}$$

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 $u(a_{\underline{\theta}}^v) + \underline{\theta} q\left(\underline{\theta}\right)$ 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 naiveté $\underline{\theta}$. 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 $\int_{\underline{\theta}}^{\theta} q\left(\theta\right)$. 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:

$$q\left(\theta\right) = \left[u\left(a_\theta^u\right) - v\left(a_\theta^u\right)\right] + \left[v\left(a_\theta^v\right) - u\left(a_\theta^v\right)\right] \tag{7}$$

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 behavior. The disagreement is that the agent believes he will choose a_θ^u and the principal believes that the agent will choose a_θ^v. If the agent is correct, then the agent's speculative gain is the difference between his first-period evaluation of a_θ^u 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_θ^v 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 $\left(a_x^v\right)_{x\in[\underline{\theta},1]}$ that maximize

$$F\left(\underline{\theta}\right) \max_a u\left(a\right) + \left[1 - F\left(\underline{\theta}\right)\right] \cdot E\left[t_\theta\left(a_\theta^v\right) | \theta \geq \underline{\theta}\right] \tag{8}$$

where $E\left[t_\theta\left(a_\theta^v\right)|\theta \geq \underline{\theta}\right]$ is given by the expression

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

with

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

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.

Proposition 4 *The principal's problem is solved in four steps:*
Step 1. Derive a_θ^v for each $\theta \geq \underline{\theta}$ where

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

Step 2: Derive $\underline{\theta}$ by solving the equation

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

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

Step 3: Having obtained $\underline{\theta}$ and $a_{\underline{\theta}}^v$ in the previous steps, solve for $t_{\underline{\theta}}\left(a^*\right)$ and $t_{\underline{\theta}}\left(a_{\underline{\theta}}^v\right)$ using the fact that $IR_{\underline{\theta}}$ and $VR_{\underline{\theta}}$ are binding.

$$\begin{aligned} 0 &= \underline{\theta}\left[u\left(a^*\right) - t_{\underline{\theta}}\left(a^*\right)\right] + (1-\underline{\theta})\left[u\left(a_{\underline{\theta}}^v\right) - t_{\underline{\theta}}\left(a_{\underline{\theta}}^v\right)\right] \\ v\left(a_{\underline{\theta}}^v\right) - t_{\underline{\theta}}\left(a_{\underline{\theta}}^v\right) &= v\left(a^*\right) - t_{\underline{\theta}}\left(a^*\right) \end{aligned}$$

Step 4: Finally, compute $t_\theta\left(a_\theta^v\right)$ and $t_\theta\left(a^*\right)$ for each $\theta > \underline{\theta}$ by solving

$$\theta\left[u\left(a^*\right)-t_\theta\left(a^*\right)\right]+(1-\theta)\left[u\left(a_\theta^v\right)-t_\theta\left(a_\theta^v\right)\right]=(\theta-\underline{\theta})\,\Delta^*+\int_{\underline{\theta}}^{\theta}\left[v\left(a_x^v\right)-u\left(a_x^v\right)\right]dx$$

and

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

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. Therefore, our original problem is amenable to the "textbook recipe".

The first step in the "recipe" yields the real action for each exploited type θ. The cutoff $\underline{\theta}$ is determined roughly as follows. In optimum, the benefit from marginally raising the cutoff is exactly offset by the loss. Suppose the current cutoff is at some ϕ. Then, if the principal raises the cutoff, he loses the surplus he could have extracted from ϕ, given that this is the lowest exploited type. This loss is equal to $\phi q\left(\phi\right)+u\left(a_\phi\right)$. However, the benefit from raising the cutoff above ϕ is twofold. First, the principal extracts $\max_a u(a)$ from type ϕ, which is the surplus extracted from every non-exploited 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 $\frac{1-F(\theta)}{f(\theta)}q\left(\theta\right)$. It follows that the principal should set $\underline{\theta}=\phi$ if

$$\phi q\left(\phi\right)+u\left(a_\phi\right)=\max_a u(a)+\frac{1-F\left(\phi\right)}{f\left(\phi\right)}q\left(\phi\right)$$

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 $t_\theta\left(a_\theta^v\right)$ and $t_\theta\left(a_\theta^u\right)$ for all $\theta > \underline{\theta}$, we equate (5) with (1) and use again the result that VR_θ is binding for all θ (Proposition 4, part (ii)).

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.

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

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

The intuition for this result is as follows. The maximal possible revenue from an exploitative contract is the revenue that can be obtained when facing only fully naive agents. The L.H.S of (12) represents this revenue.[3] Therefore, the principal offers an exploitative contract in the optimal menu if and only if this revenue exceeds the optimal non-exploitative contract.

3.3 Examples

The following set of examples serves a double role. First, it demonstrates how to apply the "recipe" of Proposition 4. Second, the optimal menu in each example resembles a commonly observed contractual arrangement. Throughout the sub-section, we assume $F(\theta) = \theta$.

Example 3.1. *"Freebees"*

Consider a situation in which agents exhibit a "reference-point effect": their evaluation of available actions depends on whether they are *choosing* a contract or *modifying* an existing contract (for experimental evidence, see Shafir (1993)). When faced with the problem of

[3] The L.H.S of this inequality is obtained by noting that when there is a single agent type, $\theta = 1$, then $\underline{\theta} = 1$. Following our discussion of Proposition 4, the surplus extracted from type $\underline{\theta}$ is $\underline{\theta} q(\underline{\theta}) + u\left(a^v_{\underline{\theta}}\right)$.

choosing a contract, the agent cannot find a good enough reason for why any action $a > 0$ is more valuable than the outside option, hence $u(a) = 0$ for all a. After signing a contract, the agent's reference point changes, and cannot find any good reason not to increase a, i.e., $v(a) = a - 1$.

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, he finds greater coverage to be desirable.

Let us characterize the optimal menu. The optimal non-exploitative contract is trivial because $\max_a u(a) = 0$. Therefore, the optimal menu contains only exploitative non-trivial contracts. Let us go through the steps in the recipe.

In step 1, we compute the real action for all exploited types. Due to the linearity of v and u, $\psi(\theta) \cdot [\Delta^* + v(a) - u(a)] + u(a)$ is monotonically increasing with a as long as $\psi(\theta) > 0$. Therefore, for every exploited type θ, $a^v = 1$. In addition, for every exploited type θ, $a^u_\theta = \arg\max_a [u(a) - v(a)] = 0$. The identity of the real action for all exploited types implies that there is a unique exploitative contract.

The remaining steps are straightforward. First, the cutoff is $\underline{\theta} = \frac{1}{2}$. Second, we use the facts that VR_θ is binding for $\theta > \frac{1}{2}$ and that $IR_{\frac{1}{2}}$ is binding, to complete the characterization of the exploitative contract: $t(1) = \frac{1}{2}$, $t(0) = -\frac{1}{2}$, and $t(b) = \infty$ for every $b \neq 0, 1$. All types $\theta < \frac{1}{2}$ choose to opt out, whereas all types $\theta > \frac{1}{2}$ choose the exploitative contract and end up consuming $a = 1$.

The exploitative contract can be interpreted as follows. The gross transaction price is 1. It gives the agent a gift of $\frac{1}{2}$ upon signing the contract, coupled with the possibility to cancel the transaction while keeping the gift. Naive agents believe that with high probability their reference point will not change and they will cancel the deal and earn the gift. 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.[4]

Example 3.2. *"Cell-phone packages" and "book clubs"*

[4]Note that there may be other types of dynamic inconsistency, other than reference-point effects, which may explain this contract. For example, the agent may have self control problems which cause him to procrastinate cancelling the deal.

Consider a situation in which the agent initially believes that an intermediate level of consumption of some good is best for him. However, when confronted with the actual choice of consumption, the agent's preferences monotonically increase with a. The cell phone example alluded to in the Introduction is a case in point. Book consumption provides another example. A consumer's initial inclination may be to purchase only a few books, but when faced with a vast selection, he is tempted to purchase as many books as possible. Formally, let $u(a) = \frac{1}{2} - \left|a - \frac{1}{2}\right|$ and $v(a) = \frac{1}{2}a$.

By following the recipe, it is straightforward to obtain the following optimal menu. All types lower than $\frac{5}{6}$ choose the non-exploitative contract $t^{NE}(\frac{1}{2}) = \frac{1}{2}$ and $t^{NE}(a) = \infty$ for all $a \neq \frac{1}{2}$. All higher types choose the exploitative contract $t^{E}(\frac{1}{2}) = \frac{3}{8}$, $t^{E}(1) = \frac{5}{8}$ and $t^{E}(a) = \infty$ for all $a \notin \{\frac{1}{2}, 1\}$. Once again, the uniqueness of the exploitative contract is due to the linearity of u and v in the relevant domain.

The optimal menu resembles a number of real-life contractual arrangements. Several cell-phone companies provide a menu that consists of two types of packages: a pre-paid package which puts an upper bound on the number of minutes, and a flexible package with no limit on the number of minutes. The flexible packages offer zero marginal price for low-intensity usage and a positive marginal price for high-intensity usage. The effective price for low-intensity usage is higher under the pre-paid package than under the flexible package.[5]

We interpret this menu as a means to discriminate between sophisticated and naive types. Sophisticated types anticipate that their taste for cell-phone conversations will increase, hence, they view the pre-paid package as a commitment device. Naive types believe that their taste will not change, and choose the flexible package because it is more attractive for their intended intensity of usage. However, they end up using the cell phone more intensively than they wished ex-ante.

Book publishers offer similar contractual arrangements. Readers are offered two alternative pricing schemes. Under one scheme, each book is sold according to its cover price. Under another scheme, readers are asked to join a book club. By paying a small membership fee, readers can receive a number of books (essentially) for free and are given the right to purchase additional books at discounted prices. In addition, members receive book catalogs and special promotions.[6] A naive reader who is only interested in purchasing a limited number of books will be tempted to join the book club. However, once he becomes a member, the reader will end up purchasing more books than he initially intended. In contrast, a sophisticated reader will choose to purchase books individually at their cover price, because this will prevent him

[5]Detailed examples of such menus are available on the following websites: www.t-mobile.com, www.vodafone.co.uk, and www.cellcom.co.il.

[6]See for example, www.doubledaybookclub.com.

from purchasing more books than he currently wishes.

Example 3.3. *A continuum of exploitative contracts*

The objective of this example is to illustrate an optimal contract that contains a continuum of exploitative contracts.

Let $v(a) = a(1-2a)$ and $u(a) = a(1-a)$. Note that the present and future selves of the agent have a conflict of interests over the range of actions $[\frac{1}{4}, \frac{1}{2}]$: while $u(\cdot)$ increases over this range, $v(\cdot)$ decreases. Let us apply the "recipe" of Proposition 4 in detail.

Step 0: Computing a^*

$$a^* = \arg\max_a [a(1-a) - a(1-2a)] = \arg\max_a \left(a^2\right) = 1$$

Therefore, $\Delta^* = 1$.

Step 1: Computing a^v_θ for all $\theta \geq \underline{\theta}$

$$\psi(\theta) \cdot [\Delta^* + v(a) - u(a)] + u(a) = (2\theta - 1) + a_\theta - 2\theta a_\theta^2$$

It follows that

$$a^v_\theta = \arg\max_{a \in [0,1]} \left[(2\theta - 1) + a - 2\theta a^2\right] = \min\left\{\frac{1}{4\theta}, 1\right\}$$

Hence, $a^v_\theta = \frac{1}{4\theta}$ for $\theta \geq \frac{1}{4}$.

Step 2. Computing $\underline{\theta}$

Because $\max_a \left\{(2\theta - 1) + a_\theta - 2\theta a_\theta^2\right\}$ must be increasing on $[\underline{\theta}, 1]$, it cannot be the case that $\underline{\theta} < \frac{1}{4}$. Therefore,

$$\max_a \left\{(2\theta - 1) + a_\theta - 2\theta a_\theta^2\right\} = 2\theta - 1 + \frac{1}{8\theta}$$

for $\theta \geq \underline{\theta} \geq \frac{1}{4}$. To solve for $\underline{\theta}$ we equate the R.H.S of the above equation to $\arg\max_a u(a) = \frac{1}{4}$ and obtain two solutions, $\frac{1}{8}$ and $\frac{1}{2}$. Since $2\theta - 1 + \frac{1}{8\theta}$ is increasing on $[\frac{1}{2}, 1]$, we set $\underline{\theta} = \frac{1}{2}$.

It follows that agents with $\theta < \frac{1}{2}$ end up choosing $\arg\max_a u(a) = \frac{1}{2}$, while agents with $\theta > \frac{1}{2}$ end up choosing an action in the interval $[\frac{1}{4}, \frac{1}{2})$. The more naive the agent, the closer his action to $\frac{1}{4}$.

Step 3: Solving for $t_{\frac{1}{2}}(\frac{1}{2})$ and $t_{\frac{1}{2}}(1)$. By $IR_{\frac{1}{2}}$,

$$\frac{1}{2}\left[0 - t_{\frac{1}{2}}(1)\right] + \frac{1}{2}\left[\frac{1}{4} - t_{\frac{1}{2}}(\frac{1}{2})\right] = 0$$

By $VR_{\frac{1}{2}}$,

$$-1 - t_{\frac{1}{2}}(1) = 0 - t_{\frac{1}{2}}(\frac{1}{2})$$

Hence, $t_{\frac{1}{2}}(\frac{1}{2}) = \frac{5}{8}$ and $t_{\frac{1}{2}}(1) = -\frac{3}{8}$.

Step 4: Solving for $t_\theta(a^v_\theta)$ and $t_\theta(1)$

$$\theta\left[-t_\theta(1)\right] + (1-\theta)\left[\frac{1}{4\theta}\left(1 - \frac{1}{4\theta}\right) - t_\theta(a^v_\theta)\right] = \theta - \frac{1}{2} - \int_{\frac{1}{2}}^{\theta} \frac{1}{16x^2}dx$$

and

$$-1 - t_\theta(1) = \frac{1}{4\theta}\left(1 - \frac{1}{2\theta}\right) - t_\theta(a^v_\theta)$$

For instance, the fully naive type $\theta = 1$ chooses a contract with $t_1(1) = -\frac{7}{16}$ and $t_1\left(\frac{1}{4}\right) = \frac{11}{16}$ (notice that in accordance to claims *(ii)* and *(iii)* of Proposition 3, $t_1(1) < t_{\frac{1}{2}}(1)$ and $t_1\left(\frac{1}{4}\right) > t_{\frac{1}{2}}\left(\frac{1}{4}\right)$).

The multiplicity of exploitative contracts is due to the nature of the conflict of interests between the present and future selves of the agent - specifically, the fact that $u'(a)/v'(a)$ is increasing with a. Rewriting step 2 in the "recipe", the principal chooses a^v_θ so as to maximize $\psi(\theta)v(a)+(1-\psi(\theta))u(a)$. When $u'(a)/v'(a)$ is not constant, this expression may have a local maximum which varies smoothly with θ. In contrast, when $u'(a)/v(a')$ is constant, there is a unique exploitative contract.

This example fits the following scenario. A borrower considers signing a debt contract with a lender. The borrower decides in period 2 how to allocate the repayment between periods 2 and 3. At period 1, the borrower has single peaked preferences: ideally, he would like to repay the loan in two equal installments. However, after receiving and using the loan, the agent's single-peaked preferences change, such that he prefers to repay only 25% of the loan in period 2.

In light of this scenario, the optimal menu may be interpreted as follows. The principal offers two types of debt contracts. One debt contract is effectively a mortgage that uses collateral as a tool to force lenders to repay half the loan in the second period. Another contract is more flexible and does not require collateral. Instead, the contract levies fines on lenders who do not repay the entire loan in the second period. However, the fines are not too high as to deter any partial repayment. In addition, lenders who repay the *entire* loan in the second period are rewarded by the value of future benefits from a meticulous credit history.

Discussion

It should be emphasized that there are alternative, more conventional explanations for the contractual designs described above. For instance, the menu described in the first example, which consists of a contract that offers a free gift, may be explained by "customer poaching". The menu described in the second example, which is analogous to that which is offered by cell phone companies, may be explained by standard price discrimination according to consumers' credit worthiness and willingness to pay (in the usual, dynamically consistent sense). We do not wish to argue that any of the two explanations is superior. Our objective is merely to show that our model does lead to a realistic contract design.

Note, however, that in order to explain the menus in the examples using conventional arguments, one may need to adopt a different framework for each example. For instance, one would need a model of customer poaching for the first example, and a different model of price discrimination for the second example. In contrast, we propose a single perspective for interpreting all the contractual arrangements in the above examples.

Is it possible to decide empirically between the various explanations? In principle, one could adopt the empirical approach taken by DellaVigna and Malmendier (2003): examining second-period behavior and checking whether it rationalizes first-period behavior. However, in the above examples, agents who choose the flexible scheme in the first period end up consuming more in the second period than agents who choose the rigid scheme. Therefore, it seems hard to rule out the possibility that the former agents simply have a stronger preference for the principal's good or service than the latter.

4 Contracting with dynamically consistent agents

The model of Section 2 was based on two important assumptions: (i) the principal and the agent have conflicting prior beliefs regarding the agent's future preferences, and (ii) the agent is dynamically inconsistent: he evaluates future actions according to his current self's utility. In this section, we relax the second assumption by analyzing a model in which the agent has *dynamically consistent preferences* (DCP), yet he continues to hold incorrect prior beliefs about them.

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 $v(\cdot)$ with probability $1-\theta$. Hence, the indirect utility of an agent of type θ, who pretends to be of type ϕ is as follows:

$$U(\phi,\theta) = \theta\left[u\left(a^u_\phi\right) - t_\phi(a^u_\phi)\right] + (1-\theta)\left[v\left(a^v_\phi\right) - t_\phi(a^v_\phi)\right]$$

where $a^u_\phi \equiv \arg\max_{a\in[0,1]}\left[u(a) - t_\phi(a)\right]$ and $a^v_\phi \equiv \arg\max_{a\in A}\left[v(a) - t_\phi(a)\right]$. Compare this expression with the expression for $U(\phi, \theta)$ in the model of Section 2. The agent's evaluation of the v-optimal action at period 1 is made according to v, rather than according to u.

As can be seen from the definition of $U(\phi, \theta)$, the difference between the current model and the model of Section 2 lies in the agent's first-period evaluation of his second-period actions. In the model with dynamically inconsistent preferences (DIP), there is a period 1 self with a utility function u, who is not certain about the preferences of the period 2 self. With DCP there is a single self, who is uncertain of the state of the world, which is revealed to him only in period 2. In fact, the dynamic element in the DCP model is superfluous from a formal point of view. The model may be interpreted as a static principal-agent problem with non-common priors. A general analysis of such a model lies beyond the scope of this paper. In this section, we are only interested in the relation between this model and the DIP model, because of their formal resemblance.

There is also a practical economic reason to be interested in this alternative model. It is not always obvious, given a particular scenario, which of the two models better fits the situation. For example, imagine a consumer who is about to enter a DVD rental store. He is unsure about how frequently he will want to watch a DVD in the future. When evaluating his willingness to pay for a monthly membership, the agent's calculation seem better approximated by the DCP model, because there is no clash between present and future selves. But now suppose that the rental store specializes in *pornographic* DVDs. In this case, it may seem more apt to assume that at the present, the consumer assigns a low value to such movies, but in the future, he will develop an insatiable taste for them. This case seems to better fit the DIP model. The economist's decision whether to categorize the situation as a DCP model or a DIP model depends on the specific details of the situation.

In the DCP model, the principal's menu-design problem is the following.

Observation 3. *The optimal menu of contracts* $\{t_\theta(a)\}_{\theta\in[0,1]}$ *is given by the solution to the following maximization problem:*

$$\max_{\{t_\theta(a)\}_{\theta\in[0,1]}} \int_0^1 t_\theta\left(a_\theta^v\right) dF(\theta)$$

subject to the constraints,

$$U(\theta, \theta) \geq 0 \qquad (IR_\theta)$$

$$U(\theta, \theta) \geq U(\phi, \theta) \qquad (IC_{\theta,\phi})$$

for all $\phi \in [0, 1]$.

Note that the constraints UR_θ and VR_θ are incorporated into the definition of $U(\theta, \theta)$.

Because the DCP model is not the focus of our paper, we do not provide a full characterization of the optimal menu. Instead, we highlight the similarities and differences between the DCP and DIP models.

As in the DIP model, a contract is said to be exploitative, if it manages to extract more than what a fully sophisticated agent would be willing to pay for the action he is induced to take. This is made precise in the following definition.

Definition 2 *A contract is exploitative if* $t_\theta(a_\theta^v) > v(a_\theta^v)$

Note that this definition is formally different from the analogous definition in the DIP model. Nevertheless, it captures the same idea: a fully sophisticated agent would never accept an exploitative contract. An agent may accept such a contract only if he has an incorrect belief of his future taste.

To see the difference between Definitions 1 and 2, consider the specification of v and u given in Example 3.1: $u(a) = 0$ for every a, $v(\cdot)$ is strictly increasing. In the DIP model, the optimal menu consists of one exploitative contract, which generates a revenue below $v(1)$. In the DCP model, the principal can offer the following contract: $t(a) = a$ for every a. This contract yields zero indirect utility for every type. Therefore, it is not exploitative. Because every type is willing to accept this contract, it generates the first-best revenue, $v(1)$.

Recall that in the DIP model, the optimal menu partitions the set of types into (at most) two intervals: the set of relatively naive types who choose exploitative contracts and the set of relatively sophisticated types who choose non-exploitative contracts. This property also holds in the DCP model:

Proposition 6 *There exists a type* $\underline{\theta} \in [0,1]$, *such that for every* $\theta > \underline{\theta}$, t_θ *is exploitative, and for every* $\theta < \underline{\theta}$, t_θ *is non-exploitative.*

Recall two properties of the optimal menu in the DIP model: (i) the menu contains a unique non-exploitative contracts, and (ii) the menu may contain multiple exploitative contracts. That is, there is no discrimination among the relatively sophisticated, unexploited types, but there may be discrimination among the relatively naive, exploited types. While the first feature is retained in the DCP model, the second feature is no longer true.

Proposition 7 *The optimal menu contains at most one non-exploitative contract and at most one exploitative contract. If the menu includes an exploitative contract, then the imaginary action associated with that contract is* $a^* \equiv \arg\max_a(u(a) - v(a))$*, while the real action is* $\arg\max_a v(a)$*.*

What is the intuition for the result that rules out multiple exploitative contracts in the DCP model? As in the DIP model, the transaction between the principal and an exploited type involves a speculative component, because the parties hold different priors. As in the DIP model, the "speculative surplus" is the sum of the speculative gains that the two parties expect from their transaction. 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 period 1 evaluation of a^u_θ and what the principal believes to be the agent's second-period evaluation of this action. This difference is $u(a^u_\theta) - v(a^u_\theta)$, as in the DIP model. 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 period 1 evaluation of this action. However, *this difference is null in the DCP model, because if the principal is correct, the agent's period* 1 *evaluation of* a^v_θ *is identical to that of the principal,* $v(a^v_\theta)$. Therefore, the "speculative surplus" is $u(a^u_\theta) - v(a^u_\theta)$. But in Proposition 7 we show that this term is constant across all exploited types in the optimal menu. Hence, the principal offers the same exploitative contract to all exploited types.

Thus, there is a major difference between the element of speculative trade in the two models. The speculative component in the DIP model may vary with the exploited agent's type, whereas in the DCP model, it is constant across all exploited types. For this reason, the optimal menu may contain multiple exploitative contracts in the DIP model, but no more than one in the DCP model.

Another important difference between the DIP and DCP model lies in the externality that sophisticated types exert on naive types. In the DIP model, the optimal menu always includes the first-best contract for the fully sophisticated type. The reason is that the indirect utility from this contract is zero for all types, hence, its inclusion in the menu does not affect the IC constraints of any type.

This property is not universally valid in the DCP model. To see why, consider the following example. Suppose that u and v are as given in Example 3.3: $v(a) = a(1-2a)$ and $u(a) = a(1-a)$. In this case, u lies above v. In the context of the DCP model, this means that the agent is over-optimistic about the value of his second-period action. Suppose that the optimal menu includes a non-exploitative contract, which induces a real action a^v. By the definition of exploitative contracts, $t(a^v) \leq v(a^v)$. Therefore, $t(a^v) < u(a^v)$, such that any type with $\theta > 0$ obtains strictly positive indirect utility from this contract. This affects the IC constraints of all types who choose an exploitative contract.

This is a standard informational-rent effect familiar from price-discrimination models. In the DIP model, the informational-rent effect holds only among the exploited types. In the DCP model, it may hold across the entire type space. Therefore, the usual distortions that arise in price-discrimination models may also emerge in the DCP model. In particular, sophisticated types may be crowded out, even when there is a surplus in the interaction with the fully sophisticated type.

The reason for this difference lies in the way different types evaluate rigid contracts in the two models. A rigid contract induces $a^v = a^u$. In the DIP model, every agent would agree that the indirect utility from this contract is $u(a^v) - t(a^v)$, regardless of his type. Thus, if a fully sophisticated agent selects a rigid contract with $u(a^v) = t(a^v)$, there is no informational rent to higher types. In contrast, in the DCP model, agents with different θ *do differ* in their evaluation of a rigid contract. Therefore, if the fully sophisticated type accepts a rigid contract, this may create an informational externality for higher types.

There are some configurations of u and v, for which it is possible to remove this externality. For instance, consider the specification of Example 3.3: $u(a) = \frac{1}{2} - \left|a - \frac{1}{2}\right|$ and $v(a) = \frac{1}{2}a$. That is, the principal knows that the agent's utility will be increasing in the quantity he consumes, whereas the agent believes that with probability θ, he prefers intermediate quantity levels. In this example, $u(a) > v(a)$ at some a, while v lies above u at $\arg\max_a v(a) = 1$. This means that the principal can include the following contract in his menu: $t(1) = \max_a v(a) = \frac{1}{2}$, $t(a^*) = u(a^*)$, and $t(\cdot) = \infty$ otherwise. This contract generates the $\theta = 0$ first-best revenue for the principal, and yields zero indirect utility for every type. In particular, the relatively sophisticated types would accept this contract without affecting the IC constraints of the relatively naive types. Thus, as in the general case in the DIP model, sophisticated types do not exert any externality on the relatively naive, exploited types. However, in contrast to the DIP model, the non-exploitative contract offered to the sophisticated types may not be rigid. This property allows the following simple characterization.

Proposition 8 *Assume that $u(a) > v(a)$ for some a and $v(a) \geq u(a)$ at $a = \arg\max v$. The* *optimal menu of contracts consists of the pair of contracts, $t_1(\cdot)$ and $t_2(\cdot)$, satisfying*

$$t_1(a) = \begin{cases} v(a) & if \quad a = \arg\max_a v(a) \\ u(a) & if \quad a = a^* \\ \infty & if \quad other \end{cases}$$

$$t_2(a) = \begin{cases} \underline{\theta}\Delta^* + v(a) & if \quad a = \arg\max_a v(a) \\ \underline{\theta}u(a) + (1-\underline{\theta})v(a) & if \quad a = a^* \\ \infty & if \quad other \end{cases}$$

such that all types $\theta < \underline{\theta}$ choose $t_1(\cdot)$ and all types $\theta > \underline{\theta}$ choose $t_2(\cdot)$, where $\underline{\theta}$ solves $\psi(\underline{\theta}) = 0$.

The optimal menu contains exactly two contracts. Both contracts induce the same real action, $\arg\max_a v(a)$, and the same imaginary action. They differ only in the transfers. Thus, the optimal menu looks like a pair of non-linear pricing schedules. Note that the sophisticated types do not choose rigid contracts. Indeed, the flexibility of the non-exploitative contracts enables the principal to remove the informational externality that sophisticated types exert on naive types.

For instance, let us return to the specification of u and v given in Example 3.3. Recall that $F(\theta) = \theta$. The optimal menu consists of two contracts: $t_1(\frac{1}{2}) = t_1(1) = \frac{1}{2}$, and $t(\cdot) = \infty$ otherwise; and $t_2(\frac{1}{2}) = \frac{3}{8}$, $t_2(1) = \frac{5}{8}$, and $t(\cdot) = \infty$ otherwise. All types $\theta < \frac{1}{2}$ choose $t_1(\cdot)$ and all types $\theta > \frac{1}{2}$ choose $t_2(\cdot)$. The result that $t_2(\cdot)$ is identical to the exploitative contract in Example 3.3 is purely accidental (this follows from the fact that $\max_a u(a) = \max_a v(a)$).

To conclude this section, let us restate the main differences between the DIP and DCP model. The element of speculative trade is subtly different in the two models. In the DIP model, the "surplus" from speculative trade may vary across types, whereas in the DCP model, it is constant across all exploited types. In the DIP model, the optimal menu may contain multiple exploitative contracts, depending on the magnitude of the conflict of interests between the two selves. In the DCP model, there is at most one exploitative contract. In the DIP model, the unexploited types never exert any externality on the exploited types. In the DCP, this need not be the case.

5 Discussion

5.1 Incorrect priors versus objective risk

A fundamental assumption in our model is that the principal and the agent hold different prior beliefs regarding the agent's future preferences. Their disagreement gives rise to speculative trade, in the form of what we referred to as "exploitative contracts". Moreover, it results in discrimination between types on account of their different 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 2, 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 v with probability $1-\theta$. As in the model of Section 2, θ is the agent's type, and the principal cannot observe it. However, in contrast to the model of Section 2, 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 $a = \arg\max_a u(a)$, 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:

$$\max_{a^u, a^v, t^u, t^v} \theta t^u + (1-\theta)t^v$$

subject to the constraints:

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

The solution to this maximization problem is $a^u = a^v = \arg\max u_a(a)$, and $t^u = t^v = \max u$. 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, we can see that the prior disagreement between the principal and the agent concerning the agent's changing tastes is crucial for the effects analyzed in this paper. A similar no-discrimination result can also be obtained when agents are time-consistent.

5.2 Competition

Our paper studies the optimal design of contracts for diversely naive agents in the context of monopolistic markets. This is in the tradition of the literature on contract design. Some of the real-life examples discussed in the paper do indeed involve monopolistic markets. The cable industry is often characterized by regional monopolies. Publishing houses usually have exclusive dealership relations with authors, which gives them monopoly power over fans of "their" authors. However, other examples (cell phones, credit cards) pertain to markets with some amount of competition. Therefore, it is interesting to examine how our analysis is affected by the introduction of competition into the model.

A natural extension of our model would study a simultaneous-move game between two firms, where each firm offers a menu of contracts. Each agent chooses a contract from the union of the firms' menus. Full analysis of this game lies beyond the scope of our paper, as it involves competition in a complex strategy space, namely menus of non-linear pricing schemes. We conjecture that in Nash equilibrium, the exploitative effect will vanish: firms will make zero profits. However, the structure of equilibrium contracts will be "non-standard", and continue to rely on the distinction between "real" and "imaginary" actions.

5.3 Imperfectly verifiable actions

The model of Section 2 imposes no restrictions on the set of possible contracts. In other words, the principal can offer the agent any commitment device. In some applications, this assumption is unrealistic. Consider, for example, a situation studied by DellaVigna and Malmendier (2003). The principal operates a health club. Let a denote the annual number of visits to the club. Let $u(a) = a$ and $v(a) = a(1 - a)$. That is, u always lies above v. The interpretation is that in period 1, the agent would like to visit the club as many times as possible. Yet, at period 2, he becomes lazy and prefers a lower attendance rate. Let $F(\theta) = \theta$.

Facing a fully sophisticated agent ($\theta = 0$), the principal's optimal contract sets $t(1) = 1$ and $t(\cdot) = \infty$ otherwise. That is, the principal offers the sophisticated type a perfect commitment device which enables him to exercise according to his first-period preferences. Within the model of Section 2, the optimal menu includes this contract, and it is selected by all types $\theta < \frac{1}{2}$.

In situations such as loan repayment or cell-phone consumption (see the examples of Section 2), it is reasonable to assume that the principal can carefully monitor the agent's consumption, and that he can supply such commitment devices to the sophisticated types.

However, in the context of the health club example, such contracts seem unrealistic.[7] Instead, it makes sense to restrict the domain of feasible contracts to flat-rate membership contracts (see DellaVigna and Malmendier (2003)). That is, the principal is restricted to contracts of the form: $t(a) = T$ for every $a \leq \bar{a}$ and $t(a) = \infty$ for every $a > \bar{a}$. Hence, the menu of contracts offered by the principal consists of a set of action-transfer pairs $\{(\bar{a}_\theta, T_\theta)\}_{\theta \in [0,1]}$.

Note that this restriction implies that both the imaginary action and the real action for type θ would lie below the threshold action specified in that type's contract. This means that the principal applies the same transfer for the imaginary and real actions, i.e., for every θ, $t(a^u_\theta) = t(a^v_\theta) \equiv T_\theta$. Therefore, *all types would end up choosing the same real action,* $\arg\max_a v(a)$.

For the above specification of u and v, it can be shown that the optimal menu consists of two contracts: $(\frac{1}{2}, \frac{1}{2})$ and $(1, \frac{3}{4})$, such that all types $\theta < \frac{1}{2}$ choose the former contract and all types $\theta > \frac{1}{2}$ choose the latter. Note that while the optimal menu looks like an ordinary non-linear pricing schedule, all types end up choosing $\arg\max_a v(a) = \frac{1}{2}$ in the second period. This result is qualitatively similar to the empirical findings of DellaVigna and Malmendier (2003). The non-linear pricing schedule screens agents according to their degree of naiveté: relatively naive types choose the high-intensity contract, whereas relatively sophisticated types choose the low-intensity contract.

5.4 Alternative notions of partial naiveté

In our framework, an agent's degree of naiveté is measured by the probability that he assigns to the event that his preferences will not change. An alternative measure of naiveté is introduced in O'Donoghue and Rabin (2001) for (β, δ) model of time-inconsistency (this is the measure of naiveté used in DellaVigna and Malmendier (2004)). 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 naiveté. Note that this notion, in contrast to ours, combines the two "behavioral" features of the agent's preferences: the misconception of his future tastes and his time-inconsistency. Consequently, it is impossible to isolate the separate effects of each of these features.

Lowenstein, O'Donoghue and Rabin (2003) (henceforth, LOR) propose a generalization of the above measure of partial naiveté. According to this alternative measure, a partially naive agent in our set-up holds the following belief in the first period: with probability one,

[7] The health club can in principle monitor the agent's number of visits (say, because the agent needs to swipe his membership card upon entry). However, if the agent is fined for low attendance, it will be easy for him to manipulate this monitoring device. More stringent devices seem unreasonably draconian in this context.

his second period utility function will be 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 naiveté. 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. Thus, according to our notion, a partially naive agent is uncertain whether or not his preferences will change, but he knows exactly what his future preferences might be.

We illustrate the difference between the two measures in the context of our model. Given some contract $t(a)$, what is the indirect utility of an agent, who is partially naive according to each of the definitions? According to our notion of naiveté, the indirect utility is equal to

$$\theta\left[u\left(a^{u}\right)-t\left(a^{u}\right)\right]+(1-\theta)\left[u\left(a^{v}\right)-t\left(a^{v}\right)\right]$$

where $a^u \equiv \arg\max_a[u(a) - t(a)]$ and $a^v \equiv \arg\max_a[v(a) - t(a)]$. In contrast, an agent who is partially naive in the LOR sense, believes that his second period utility function is given by some convex combination of u and v, $w(a, \alpha) \equiv \alpha u(a) + (1-\alpha)\, v(a)$ with $\alpha \in [0, 1]$. The indirect utility of such an agent from a contract $t(a)$ is given by

$$u[\arg\max_{a}(w\left(a, \alpha\right)-t\left(a\right))]$$

Note that the LOR notion of partial naiveté is similar to ours, in the sense that it separates between the agent's misconception of his future tastes and other features of his preferences (such as possible time-inconsistency). However, the focus of LOR is different than ours. In particular, they do not investigate the question of optimal mechanism-design when agents' private types are given by their degree of naiveté

We offer a way to reconcile the different notions of partial naiveté. The LOR approach may be viewed as a reduced form of a model in which the agent's bias results from incorrect probabilities assigned to some underlying states of the world. For instance, a consumer may underestimate the magnitude of his future satisfaction from some product because he does not anticipate certain contingencies in which his future self will derive pleasure from the product. Consider an academic who is about to purchase a website designing software. He is currently is interested in the software for purely professional purposes. He knows that his future self will be interested in non-professional applications of the software. However, he fails to anticipate some of these applications - e.g., posting family photos on his website.

References

[1] Amador, M., I. Werning and G-M Angeletos (2004), "Commitment vs. Flexibility", Mimeo, Standford GSB.

[2] Benabou, R. and J. Tirole (2002), "Self-Confidence and Personal Motivation", *Quarterly Journal of Economics* **117**, 871-915.

[3] Buunk, B.P., R.L. Collins, S.E. Taylor, N.W. Van Yperen and G.A. Dakof (1990), "The Affective Consequence of Social Comparison: Either Direction Has its Ups and Downs", *Journal of Personality and Social Psychology* **59**, 1238-1249.

[4] DellaVigna, S. and U. Malmendier (2003), "Overestimating Self-Control: Evidence from the Health Club Industry", Mimeo, University of California at Berkeley.

[5] DellaVigna, S. and U. Malmendier (2004), "Contract Design and Self-Control: Theory and Evidence", *Quarterly Journal of Economics* **119**, 353-402.

[6] Esteban, S., E. Miyagawa and M. Shum (2003), "Nonlinear Pricing with Self-Control Preferences", Mimeo, Columbia University.

[7] Fang, H. and G. Moscarini (2003), "Morale Hazard", *Journal of Monetary Economics,* forthcoming.

[8] Gilpatric, S. (2003), "Present-Biased Preferences, Self-Awareness and Shirking in a Principal-Agent Setting", University of Tennessee, mimeo.

[9] Gruber, J. and B. Koszegi (2001), "Is Addiction 'Rational?' Theory and Evidence" *Quarterly Journal of Economics* **116**, 1261-1305.

[10] Gul, F. and W. Pesendorfer (2001), "Temptation and Self-Control", *Econometrica* **69**, 1403-1435.

[11] Gul, F. and W. Pesendorfer (2004), "Harmful Addiction", *Review of Economic Studies*, forthcoming.

[12] Helgeson V.S. and S.E. Taylor (1993), "Social Comparisons and Adjustment Among Cardiac Patients", *Journal of Applied Social Psychology* **23**, 1171-1195.

[13] Kahneman, D. (1992), "Reference Points, Anchors, Norms, and Mixed Feelings", *Organizational Behavior and Human Decision Processes* **51**, 296-312.

[14] Kahneman, D., J. Knetsch, and R. Thaler (1991), "The Endowment Effect, Loss Aversion, and Status Quo Bias", *Journal of Economic Perspectives* **5**, 193-206.

[15] Krishna, V. (2002), *Auction Theory*, Academic Press, San Diego.

[16] Laibson, D. and L. Yariv (2004), "Safety in Markets: An Impossibility Theorem for Dutch Books", Mimeo, UCLA.

[17] Lowenstein, G., D. Read and R. Baumeister (Eds.) (2002), *Time and Decision: Economic and Psychological Perspectives on Intertemporal Choice*, New York: Russel Sage Foundation Press.

[18] Lowenstein, G., T. O'Donoghue and M. Rabin (2003), "Projection Bias in Predicting Future Utility", *Quarterly Journal of Economics* **118**, 1209-1248.

[19] O'Donoghue, T. and M. Rabin (1999a), "Addiction and Self-Control" in *Addiction: Entries and Exists*, J. Elster (ed.). New York: Russel Sage Foundation Press.

[20] O'Donoghue, T. and M. Rabin (1999b), "Incentives for Procrastinators", *Quarterly Journal of Economics* **114**, 103-124.

[21] O'Donoghue, T. and M. Rabin (1999c), "Doing It Now or Later", *American Economic Review* **89**, 103-124.

[22] O'Donoghue, T. and M. Rabin (2001), "Choice and Procrastination", *Quarterly Journal of Economics* **116**, 121-160.

[23] Piccione, M. and A. Rubinstein (2003), "Modeling the Economic Interaction of Agents with Diverse Abilities to Recognize Equilibrium Patterns", *Journal of European Economic Association* **1**, 212-223.

[24] Rubinstein, A. (1993), "On Price Recognition and Computational Complexity in a Monopolistic Model", *Journal of Political Economy* **101**, 473-484.

[25] Sarafidis, Y. (2004), "Inter-temporal Price Discrimination with Time Inconsistent Consumers", Mimeo, INSEAD.

[26] Shafir, E. (1993), "Choosing vs. Rejecting: Why Some Options Are Both Better and Worse than Others", *Memory & Cognition*, **21**, 546-556.

[27] Spiegler, R. (2003), "The Market for Quacks", Mimeo, Tel-Aviv University.

[28] Spiegler, R. (2004), "Competition over Agents with Boundedly Rational Expectations", Mimeo, Tel-Aviv University.

[29] Tversky, A. and D. Kahneman (1991), "Loss Aversion in Riskless Choice: A Reference-Dependent Model", *Quarterly Journal of Economics*, 1039-1061.

[30] Van den Steen, E. (2001), "Essays on the Managerial Implications of Differing Priors", PhD Dissertation, Stanford Business School.

[31] Weinstein, N.D. (1980), "Unrealistic Optimism about Future Life Events", *Journal of Personality and Social Psychology* **39**, 806-820.

[32] Weeks et al. (1998), "Relationship Between Cancer Patients' Predictions of Prognosis and Their Treatment Preferences", *Journal of the American Medical Association* **279**, 1709-1714.

[33] Yildiz, M. (2003), "Bargaining without a common prior – An immediate agreement theorem", *Econometrica* **71**, 793-811.

Appendix: Proofs

Proof of Proposition 1. If the menu contains a non-exploitative contract, then there must be a unique such contract, satisfying $t(a) = \max_a u(a)$ for $a = \arg\max_a u(a)$ and $t(\cdot) = \infty$ otherwise.

If there are no exploitative contracts, then clearly there will be one non-exploitative contract. Suppose that the menu contains an exploitative contract. Suppose that there is a family $\{t_\theta\}$ of exploitative contracts that generate a revenue of less than $\max_a u(a)$. Let t_θ be the exploitative contract that generates the least amount of revenue. The principal can deviate by replacing this contract with the above non-exploitative contract. Because the non-exploitative contract yields zero utility for every type, the IR_θ and $IC_{\phi,\theta}$ constraints will continue to be satisfied for every $\phi \neq \theta$. Finally, if IC_θ is violated, then type θ will deviate to a contract that yields a higher revenue, by the definition of t_θ.

To complete the proof we use the following lemma:

Lemma 1 *Any exploitative contract that satisfies IR_θ also satisfies IR_ϕ with a strict inequality for every $\phi > \theta$.*

Proof. If t_θ is exploitative, then $u(a^v_\theta) - t(a^v_\theta) < 0$. If this contract satisfies IR_θ, then $u(a^u_\theta) - t(a^u_\theta) > 0$. For any $\phi > \theta$, the weight on the latter term increases at the expense of the former term, such that $U(\phi, \theta) > U(\theta, \theta)$. By $IC_{\phi,\theta}$, $U(\phi, \phi) \geq U(\phi, \theta)$. Therefore, $U(\phi, \phi) > 0$. ■

By Lemma 1, if type θ chooses an exploitative contract, then every type $\phi > \theta$ satisfies IR_ϕ unbindingly. Therefore, such a type must strictly prefer his contract to the single non-exploitative contract, which generates zero utility to every type. It follows that the type space $[0, 1]$ can be partitioned into two intervals, such that high types choose an exploitative contract, and low types choose the single non-exploitative contract. □

Proof of Remark 1. Assume the contrary - i.e., $U(\underline{\theta}, \underline{\theta}) > 0$. By Lemma 1, $U(\underline{\theta}, \phi) > 0$. Suppose that the principal deviates by modifying all the exploitative contracts as follows: for every $\phi \geq \underline{\theta}$, $t(a^u_\phi)$ and $t(a^v_\phi)$ are both reduced by some arbitrarily small ε. This modification leaves all the IR, IC, UR and VR constraints intact, and generates a higher revenue, a contradiction.

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

Lemma 2 $v(a^v_\theta) - u(a^v_\theta) > v(a^u_\theta) - u(a^u_\theta)$ *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, $v(a^v_\theta) - t(a^v_\theta) \geq v(a^u_\theta) - t(a^u_\theta)$. Substituting the two inequalities into $U(\theta, \theta)$, we obtain $u(a^v_\theta) - t(a^v_\theta) \geq 0$, in contradiction to our assumption that t_θ is exploitative. ■

Lemma 3 *For every exploitative contract* t_θ*, we can set w.l.o.g.* $a^u_\theta = \arg\max_a (u(a) - v(a))$.

Proof. Denote $a^* \equiv \arg\max_a (u(a) - v(a))$. Lemma 2 guarantees that $a^* \neq a_\theta$. Suppose that $a^u_\theta \neq a^*$. The principal can modify the original contract to one that replaces a^u_θ with a^* (i.e., impose an infinitely large fine on playing 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, IR_θ and the $IC_{\theta,\cdot}$ constraints are preserved. The only thing that remains to be verified is that VR_θ is satisfied. That is, picking a_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 $v(a^*) - u(a^*) < v(a^u_\theta) - u(a^u_\theta)$, NR_θ continues to be satisfied. ■

Lemma 4 *$t(a^u_\theta)$ is non-increasing in θ in the range $\theta > \underline{\theta}$.*

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

$$\begin{aligned}
\frac{\theta}{1-\theta}\left[u\left(a^u_\theta\right) - t_\theta(a^u_\theta)\right] + \left[u\left(a^v_\theta\right) - t_\theta(a^v_\theta)\right] - \frac{\theta}{1-\theta}\left[u\left(a^u_\phi\right) - t_\phi(a^u_\phi)\right] - \left[u\left(a^v_\phi\right) - t_\phi(a^v_\phi)\right] &\geq 0 \\
\frac{\phi}{1-\phi}\left[u\left(a^u_\phi\right) - t_\phi(a^u_\phi)\right] + \left[u\left(a^v_\phi\right) - t_\phi(a^v_\phi)\right] - \frac{\phi}{1-\phi}\left[u\left(a^u_\theta\right) - t_\theta(a^u_\theta)\right] - \left[u\left(a^v_\theta\right) - t_\theta(a^v_\theta)\right] &\geq 0
\end{aligned}$$

(For $\theta = 1$ we need not divide the inequality by $1-\theta$.)

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

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

Hence, $t_\theta(a^u_\theta) \geq t_\phi(a^u_\phi)$ if and only if $\theta \leq \phi$. ■

The following Lemma is instrumental in characterizing the change in $t_\theta(a^v_\theta)$ with respect to the type θ.

Lemma 5 *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_2,θ_1} is satisfied, then $U(\theta_3, \theta_2) \geq U(\theta_3, \theta_1)$.*

Proof. type θ_i prefers t_{θ_2} to t_{θ_1} if $U(\theta_2, \theta_i) \geq U(\theta_1, \theta_i)$, where:

$$\begin{aligned}
U(\theta_2, \theta_i) &= \theta_i\left[u\left(a^u_{\theta_2}\right) - t_{\theta_2}(a^u_{\theta_2})\right] + (1-\theta_i)\left[u\left(a^v_{\theta_2}\right) - t_{\theta_2}(a^v_{\theta_2})\right] \\
U(\theta_1, \theta_i) &= \theta_i\left[u\left(a^u_{\theta_1}\right) - t_{\theta_1}(a^u_{\theta_1})\right] + (1-\theta_i)\left[u\left(a^v_{\theta_1}\right) - t_{\theta_1}(a^v_{\theta_1})\right]
\end{aligned}$$

Rewriting:

$$\theta_i\left\{\left[(u_2 - t_2) - (u_2 - t_2)\right] - \left[(u'_1 - t'_1) - (u_1 - t_1)\right]\right\} \geq (u_1 - t_1) - (u_2 - t_2) \qquad (13)$$

where u_i and t_i are abbreviated notation for $u(a^v_{\theta_i})$ and $t_{\theta_i}(a^v_{\theta_i})$, and u'_i and t'_i are abbreviated notation for $u(a^u_{\theta_i})$ and $t_{\theta_i}(a^u_{\theta_i})$.

By Lemma 3, $u'_1 = u'_2$, and by Proposition 2, $t'_2 \leq t'_1$. Therefore, $u'_2 - t'_2 \geq u'_1 - t'_1$. If the R.H.S of the inequality (13) is negative, then t_{θ_1} dominates t_{θ_1}, and IC_{θ_1,θ_2} fails to hold. It follows that L.H.S of (13) is non-negative, and therefore, non-decreasing in θ_i. ■

Using this lemma we establish the following result.

Lemma 6 *$t(a^v_\theta)$ is non-decreasing in θ.*

Proof. Assume not. Suppose that $t(a^v_\theta)$ attains a global maximum at some $\theta < 1$. Modify the menu by omitting all t_ϕ for $\phi > \theta$. Given that t_θ satisfies IR_θ, it follows from the proof of Lemma 1 that $U(\theta, \phi) > 0$. 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 strictly increases the expected revenue for the principal. Let $\bar{\theta}$ be the highest type below θ for which $t(\cdot)$ attains a local maximum. If no such $\bar{\theta}$ exists, the proof is complete. If there exists a $\bar{\theta}$, omit all contracts t_ϕ for $\phi \in (\bar{\theta}, \hat{\theta}]$ where $\hat{\theta} \in (\bar{\theta}, \theta]$ satisfies $t\left(a_{\hat{\theta}}\right) = t\left(a_{\bar{\theta}}\right)$. By the same argument as above, no type ϕ would want to deviate to a contract t_ω with $\omega < \bar{\theta}$. If any type ϕ deviates to a contract t_ω with $\omega > \hat{\theta}$, the principal's expected revenue will only increase. Again, this modification leads to a strictly higher expected revenue for the seller. Continuing in this fashion 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. Define $m(\theta) \equiv t_\theta\left(a^v_\theta\right) - u\left(a^v_\theta\right)$. 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 $\theta \geq \underline{\theta}$,

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

I.e., $V(\theta)$ is a maximum of a family of affine functions, and hence it is convex on $[\underline{\theta}, 1]$.[8]

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,

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

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

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

This means that by setting δ_θ, the slack in VR_θ, to zero for all $\theta \geq \underline{\theta}$, the principal raises his revenue without violating IR_θ or any of the incentive compatibility constraints. □

Proof of Proposition 4. *Proof of Step 1.* Because there is no link between the different types (the incentive compatibility constraints have already been incorporated into the objective function), we proceed via point-by-point optimization.

Proof of Step 2. Substituting (9) into (8), we obtain the following expression for the principal's objective function,

$$F(\underline{\theta}) \max_a u(a) + \int_{\underline{\theta}}^{1} \max_a \left[\psi(x) \cdot (\Delta^* + v(a) - u(a)) + u(a)\right] \cdot f(x)dx \tag{14}$$

Denote the expression $\max_a \left[\psi(x) \cdot (\Delta^* + v(a) - u(a)) + u(a)\right]$ by $\psi(x)q(x) + u(a_x^v)$. From (14) it follows that if $\psi(x)q(x) + u(a_x^v) < \max_a u(a)$ for some x, then the principal is better off assigning x a non-exploitative contract. We claim that in the optimal menu it must be the case that if type θ is exploited, then $\psi(x)q(x) + u(a_x^v)$ is increasing in x for all types higher than or equal to θ.

To see this, recall that by assumption, $\psi(x)$ is an increasing function. By Observation 2, $q(x) \geq 0$ for all $x \geq \underline{\theta}$. In addition, by Proposition 1, if type θ is exploited, then so is type $\phi > \theta$. It follows that $\psi(\phi)q(\theta) + u(a_\theta^v) > \psi(\theta)q(\theta) + u(a_\theta^v)$. Therefore, the optimal menu must satisfy that $\psi(\phi)q(\phi) + u(a_\phi^v) > \psi(\theta)q(\theta) + u(a_\theta^v)$.

The above observation has the following implication. If $\psi(\theta)q(\theta) + u(a_\theta^v) < \max_a u(a)$ for all θ, then no exploitative contract should be offered, i.e., $\underline{\theta} = 1$. If $\psi(\theta)q(\theta) + u(a_\theta^v) = \max_a u(a)$ for some set of types , then $\underline{\theta}$ should be set equal to the highest type in this set, provided $\psi(x)q(x) + u(a_x^v)$ is increasing 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 $\psi(\theta)q(\theta) + u(a^v_\theta)$ cannot lie above $\max_a u(a)$ for all θ.

Proof of Step 3. By part (ii) of Proposition 4, $VR_{\underline{\theta}}$ is binding. By Remark 1, $IR_{\underline{\theta}}$ is binding. The combination of these two binding constraints allow us to solve for $t'_{\underline{\theta}}(a^*)$ and $t_{\underline{\theta}}\left(a^v_{\underline{\theta}}\right)$.

Proof of Step 4. To solve for the two unknowns $t_\theta(a^v_\theta)$ and $t_\theta(a^*)$, we use the following pair of equations. The first equation is given by part (ii) of Proposition 4, which states that VR_θ is binding. The second equation is given by the two alternative formulations of $U(\theta, \theta)$: (1) and (5).

This completes the proof of Proposition 4. □

Proof of Proposition 5. Let us first prove the sufficiency part. Suppose that (12) holds, and yet the menu does not include an exploitative contract. Then the principal extracts $\max_a u(a)$ from all types, using a single contract with $a = \arg\max_a u(a)$ (and infinite fines on all other actions).

Now, add to this menu a contract with $a^u = \arg\max_a(u(a) - v(a))$ and $a^v = \max_a v(a)$. Note that it cannot be the case that $a^v = a^u$, because otherwise, the inequality (12) would be violated. Let us set $t(a^v)$ and $t(a^u)$ such that $v(a^v) - t(a^v) = v(a^u) - t(a^u)$. In this way, every type that chooses this new contract will choose the action a^v at the second period.

We now show that we can set $t(a^v)$ to be larger than $\max_a u(a)$, such that some types will choose the new contract. The only thing we need to show is that for some types θ, the following inequality holds:

$$\theta \cdot [u(a^u) - t(a^u)] + (1 - \theta) \cdot [u(a^v) - t(a^v)] > 0 \tag{15}$$

Note that this is the IR and IC constraint for those types who choose the new contract. Because $t(a^u) = t(a^v) + v(a^u) - v(a^v)$ we may rewrite (15) as follows:

$$t(a^v) < u(a^v) + \theta \cdot [u(a^u) - v(a^u) + v(a^v) - u(a^v)] \tag{16}$$

For θ sufficiently close to one, we can find a $t(a^v) > \max_a u(a)$ which satisfies (16) whenever (12) holds.

Let us turn to the necessary part. Suppose that (12) is violated, and yet the menu contains an exploitative contract. By Lemma 3, we can set $a^u = \arg\max_a(u(a) - v(a))$ for such a contract. Also, the condition $v(a^v) - t(a^v) \geq v(a^u) - t(a^u)$ must hold. By the definition of exploitative contracts, $t(a^v) > u(a^v)$. Finally, the IR condition must hold for types who choose exploitative contracts:

$$\theta \cdot [u(a^u) - t(a^u)] + (1-\theta) \cdot [u(a^v) - t(a^v)] \geq 0$$

Now it can be shown that if (12) does not hold, then there exists no type θ for whom all these conditions are satisfied. □

Proof of Proposition 6. Assume, contrary to the statement of the proposition, that there exists a pair of types $\phi > \theta$ such that type θ chooses an exploitative contract $t_\theta(\cdot)$, while type ϕ chooses a non-exploitative contract $t_\phi(\cdot)$. By $IC_{\theta,\phi}$,

$$\theta \left[u\left(a_\theta^u\right) - t_\theta(a_\theta^u)\right] + (1-\theta)\left[v\left(a_\theta^v\right) - t_\theta(a_\theta^v)\right] \geq \theta\left[u\left(a_\phi^u\right) - t_\phi(a_\phi^u)\right] + (1-\theta)\left[v\left(a_\phi^v\right) - t_\phi(a_\phi^v)\right]$$

By the definition of an exploitative contract, $v\left(a_\theta^v\right) - t_\theta(a_\theta^v) < 0$ while $v\left(a_\phi^v\right) - t_\phi(a_\phi^v) \geq 0$. Hence, for the above inequality to hold, it must be the case that $u\left(a_\theta^u\right) - t_\theta(a_\theta^u) > u\left(a_\phi^u\right) - t_\phi(a_\phi^u)$. Therefore, we obtain:

$$\theta\left\{\left[u\left(a_\theta^u\right) - t_\theta(a_\theta^u)\right] - \left[u\left(a_\phi^u\right) - t_\phi(a_\phi^u)\right] - \left[v\left(a_\theta^v\right) - t_\theta(a_\theta^v)\right] + \left[v\left(a_\phi^v\right) - t_\phi(a_\phi^v)\right]\right\}$$
$$+\left[v\left(a_\theta^v\right) - t_\theta(a_\theta^v)\right] - \left[v\left(a_\phi^v\right) - t_\phi(a_\phi^v)\right] \geq 0$$

Since the argument multiplying θ is strictly positive, the above expression continues to be strictly positive if we replace θ with ϕ. But this contradicts $IC_{\phi,\theta}$. □

Proof of Proposition 7. The proof consists of a series of lemmas.

Lemma 7 *The optimal menu contains at most one non-exploitative contract.*

Proof. Suppose first that v lies above u. Then, the principal can attain the first-best outcome without exploiting agents, by offering a single contract, $t(a) = v(a)$ for every a. This contract yields zero indirect utility to every type, and therefore, all types will accept it. In the second period, they will be willing to choose $\arg\max_a v(a)$.

Now suppose that $u(a) > v(a)$ for some a. Suppose that the menu contains more than one non-exploitative contract. Let t_θ denote the non-exploitative contract that generates the maximum revenue among the non-exploitative contracts. Let us distinguish between two cases.

Case I. Suppose that $v(a_\theta^v) > u(a_\theta^v)$. In this case, the principal can deviate by eliminating all contracts that earn him less than t_θ, and at the same time modifying t_θ into $\hat{t}_\theta$ such that $\hat{t}_\theta(a_\theta^v) = v(a_\theta^v)$ and $\hat{t}_\theta(a^*) = u(a^*)$. Note that the UR_θ and VR_θ constraints continue to hold, such that a^* and a_θ^v are the imaginary and real actions induces by $\hat{t}_\theta$. Because t_θ is non-exploitative, $\hat{t}_\theta(a_\theta^v) \geq t_\theta(a_\theta^v)$. Note that $\hat{t}_\theta$ yields zero utility for any type. Therefore, all types who previously chose the eliminated contracts will choose some contract in the new menu. Moreover, no type who originally chose a more profitable contract will now choose $\hat{t}_\theta$. Therefore, the deviation cannot reduce the principal's profits.

Case II. Suppose that $u(a_\theta^v) > v(a_\theta^v)$. In this case, t_θ yields non-negative utility to any type. Therefore, the principal can deviate by eliminating all contracts that earn less than t_θ, and he will not lose customers.■

Lemma 8 *for every exploitative contract t_θ, we can set w.l.o.g. $a_\theta^u = a^* \equiv \arg\max_a(u(a) - v(a))$.*

Proof. Let $(t_\theta)_\theta$ be an optimal menu of contracts for some $v(\cdot)$ and $u(\cdot)$ such that all types higher than some type $\underline{\theta}$ choose an exploitative contract. Consider some type $\theta \geq \underline{\theta}$ for whom $a_\theta^u \neq a^*$. If $a_\theta \neq a^*$, then we may use the proof of Lemma 3.

Now assume that $a_\theta = a^*$. By the definition of a_θ^u,

$$u\left(a_\theta^u\right) - t_\theta\left(a_\theta^u\right) \geq u\left(a^*\right) - t_\theta\left(a^*\right) \tag{17}$$

By VR_θ,

$$v\left(a^*\right) - t_\theta\left(a^*\right) \geq v\left(a_\theta^u\right) - t_\theta\left(a_\theta^u\right)$$

Adding the two inequalities, we obtain

$$v\left(a^*\right) - u\left(a^*\right) \geq v\left(a_\theta^u\right) - u\left(a_\theta^u\right)$$

This implies, by the definition of a^*, that $v\left(a^*\right) - u\left(a^*\right) = v\left(a_\theta^u\right) - u\left(a_\theta^u\right)$. This in turn implies, by (17), that

$$u\left(a_\theta^u\right) - t_\theta\left(a_\theta^u\right) = u\left(a^*\right) - t_\theta\left(a^*\right) \tag{18}$$

Suppose that the principal modifies the original contract offered to type θ as follows. If the agent chooses a^* he is asked to pay $t_\theta\left(a^*\right)$. If he chooses any other action, he pays an infinitely large fine. From (18) it follows that the above modification to the original contract does not alter any of the constraints. Therefore, it is possible to induce the same outcome with $a_\theta^u = a^*$.

Lemma 9 *The optimal menu contains at most one exploitative contract.*

Proof. Rewrite $U(\theta,\theta)$ as follows:

$$U(\theta,\theta)=\theta\left[u\left(a_\theta^u\right)-t_\theta\left(a_\theta^u\right)-v\left(a_\theta^v\right)+t_\theta\left(a_\theta^v\right)\right]+\left[v\left(a_\theta^v\right)-t_\theta\left(a_\theta^v\right)\right]$$

By the lemma above, w.l.o.g. we can let $a_\theta^u=a^*$ for all θ who choose an exploitative contract. Define $\Delta^*\equiv u\left(a^*\right)-v\left(a^*\right)$ as before. The following notation will be useful:

- $d_\theta\equiv v\left(a^*\right)-t_\theta\left(a^*\right)$
- $\delta_\theta\equiv v\left(a_\theta^v\right)-t_\theta\left(a_\theta^v\right)-d_\theta$
- $q\left(\theta\right)\equiv\left[u\left(a_\theta^u\right)-t_\theta\left(a_\theta^u\right)\right]-\left[v\left(a_\theta^v\right)-t_\theta\left(a_\theta^v\right)\right]$

As in the DIP model, $q(\theta)\geq 0$ for every type θ who chooses an exploitative contract. To see why, note that by the definition of an exploitative contract, $v\left(a_\theta^v\right)-t_\theta\left(a_\theta^v\right)<0$. Hence, for IR_θ to hold it must be the case that $u\left(a_\theta^u\right)-t_\theta\left(a_\theta^u\right)>0$.

It follows that for any type θ who chooses an exploitative contract we can write:

$$\begin{aligned}
q\left(\theta\right) &= u\left(a^*\right)-v\left(a_\theta^v\right)+t_\theta\left(a_\theta^v\right)-t_\theta\left(a^*\right)\\
&= u\left(a^*\right)-v\left(a_\theta^v\right)+v\left(a_\theta^v\right)-d_\theta-\delta_\theta-v\left(a^*\right)+d_\theta\\
&= u\left(a^*\right)-v\left(a^*\right)-\delta_\theta\\
&= \Delta^*-\delta_\theta\geq 0
\end{aligned}$$

By the same reasoning that led to Proposition 3, any contract that satisfies type θ's IC constraints has the property that

$$\begin{aligned}
U\left(\theta,\theta\right) &= U(\underline{\theta},\underline{\theta})+\int_{\underline{\theta}}^{\theta}q\left(\theta\right)dx\\
&= U(\underline{\theta},\underline{\theta})+\int_{\underline{\theta}}^{\theta}\left(\Delta^*-\delta_\theta\right)dx\\
&= U(\underline{\theta},\underline{\theta})+\int_{\underline{\theta}}^{\theta}\Delta^*dx-\int_{\underline{\theta}}^{\theta}\delta_\theta dx
\end{aligned}$$

It follows that just as in the DIP case, $\delta_\theta = 0$ for all $\theta \geq \underline{\theta}$. Hence, any incentive compatible contract belonging to the optimal menu must satisfy

$$U(\theta, \theta) = (\theta - \underline{\theta}) \Delta^* \tag{19}$$

Therefore, the contract chosen by exploited types is independent of the type. ■

Lemma 10 *For every $\theta > \underline{\theta}$, $a_\theta^v = \arg\max_a v(a)$.*

Proof. The previous lemmas established that the optimal menu includes at most one exploitative contract $t(\cdot)$ and at most one non-exploitative contract $\hat{t}(\cdot)$. In addition, proposition 6 established that there exists a lowest exploited type $\underline{\theta}$. By $IC_{\underline{\theta},\phi}$ for $\phi < \theta$,

$$\underline{\theta}\left[u\left(a^u\right) - t(a^u)\right] + (1 - \underline{\theta})\left[v\left(a^v\right) - t(a^v)\right] = \underline{\theta}\left[u\left(b^u\right) - \hat{t}(b^u)\right] + (1 - \underline{\theta})\left[v\left(b^v\right) - \hat{t}(b^v)\right]$$

where a^u and a^v are the imaginary and real actions induced by $t(\cdot)$, and b^u and b^v are the imaginary and real actions induced by $\hat{t}(\cdot)$. Using $VR_{\underline{\theta}}$ we may express $t(a^u)$ as a function of $t(a^v)$, $v(a^v)$ and $v(a^u)$. The principal's objective is to maximize

$$F\left(\underline{\theta}\right) t\left(b^v\right) + \left[1 - F\left(\underline{\theta}\right)\right] t\left(a^v\right)$$

Substituting $t(b^v)$ from the above equation into this objective function, one obtains that the objective function increases with $v(a^v)$. □

Proof of Proposition 8. We have established that there exists $\underline{\theta}$, such that all types $\theta < \underline{\theta}$ choose the same non-exploitative contract, and all types $\theta > \underline{\theta}$ choose the same exploitative contract. Let us first characterize the non-exploitative contract. Given the conditions on u and v, the principal can offer the following contract: $t(\arg\max_a v(a)) = \max_a v(a)$, $t(a^*) = u(a^*)$. By assumption, v lies above u at $\arg\max v$ and u lies above v at a^*. Therefore, for every type θ who chooses this contract, $a_\theta^v = \arg\max_a v(a)$ and $a_\theta^u = a^*$. Note that this contract yields zero indirect utility for every type. Moreover, because it generates a revenue of $\max_a v(a)$ from types who choose it, it is an optimal non-exploitative contract.

Denote the exploitative contract by t. Let a^u and a^v denote the imaginary and real actions induced by the contract. We have already established that $a^u = a^*$ and $a^v = \arg\max_a v(a)$. For every exploited type θ, the IC constraint (which coincides with IR_θ because the non-exploitative contract yields zero indirect utility for all types) and the VR constraints are:

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

Because $t(\cdot)$ is exploitative, $v(a^{v}) - t(a^{v}) < 0$ and therefore, $u\left(a^{u}\right) - t(a^{u}) > 0$. It follows that the L.H.S of the first inequality is increasing in θ.

The two constraints are binding at $\underline{\theta}$: otherwise, the principal can profitably deviate by increasing $t(a^{v})$ and adjusting $t(a^{u})$ accordingly, such that the two constraints continue to hold at $\underline{\theta}$, and therefore, for all $\theta \geq \underline{\theta}$. Using the identities $a^{u} = a^{*}$ and $a^{v} = \arg\max_{a} v(a)$, we can solve for $t(a^{v})$ and $t(a^{u})$:

$$\begin{aligned} t\left(\arg\max_{a} v(a)\right) &= \underline{\theta}\Delta^{*} + \max_{a} v(a) \\ t(a^{*}) &= \underline{\theta} u\left(a^{*}\right) + (1-\underline{\theta})\, v(a^{*}) \end{aligned}$$

It remains to find the value of $\underline{\theta}$:

$$\underline{\theta} = \arg\max_{\theta\in[0,1]} \left\{ F\left(\theta\right) \left[\max_{a} v\left(a\right) \right] + \left[1 - F\left(\theta\right)\right] \left[\theta\Delta^{*} + \max_{a} v\left(a\right) \right] \right\} = \arg\max_{\theta\in[0,1]} \theta \left[1 - F\left(\theta\right)\right]$$

Therefore, $\underline{\theta}$ is given by $\psi(\underline{\theta}) = 0$. $\square$