Cash or Credit? On the Superiority of Flexible Contracts

Dmytro Kylymnyuk
University of Geneva

Abstract

This paper studies an adverse-selection model in which a time-inconsistent buyer has a choice to pay for the good in one of the two future periods. We show that the flexible contract is optimal even if the buyer is fully aware of his time-preferences and correctly predicts his future behavior. Such a contract gives the seller an extra screening device and not only increases the seller’s profit but also improves the efficiency of trade. The self-ignorance (naïveté) of the buyer has an ambiguous effect on the seller’s profit: it makes the buyer pay a higher price but, at the same time, it reduces the screening power of the flexible contract and makes discrimination more costly.

KEYWORDS: time-inconsistency, flexible contracts, credit

1 Introduction

A flexible contract allows the agent to make or modify choices after the contract is concluded. The phenomenon is widely observed. A plane ticket is sold at different prices depending on whether or not it can be reimbursed if the buyer does not use it. A borrower who takes a loan from a bank may usually postpone the repayment at an additional cost. A bank offers a client to select between different payment cards and the client will decide which services to use after he has selected the card. In a flexible contract, the agent cannot fully commit to a particular plan of actions at the moment of concluding an agreement with another party.

Why do we observe such contracts? One reason is given by Courty and Li (2000) who study a model of price discrimination in which consumers have only imperfect knowledge about their own valuation and
discover its exact value in the future. The authors show that sequential screening mechanisms, such as a menu of refund contracts, allow the monopoly to increase its profit by extracting more surplus from the consumers who face greater uncertainty.

Heidhues and Köszegi (2010) propose another explanation which is related to time-inconsistency and self-ignorance, or naïveté: some people make systematic mistakes about their own behavior in the future. For instance, a borrower who initially intends to pay back in a short time may change his mind later without any apparent reason, postpone the payment and finally pay higher interests. When the borrower underestimates his own propensity to delay, he may accept a contract which he would otherwise decline. This argument requires not only that the individual’s preferences change over time but also that he is, at least to some extent, ignorant of this change. Such individuals are not only time-inconsistent but also partially naïve rather than fully sophisticated, and for this reason they incorrectly foresee their own behavior.

The existing literature on contracting under time-inconsistency, which includes Della Vigna and Malmendier (2004), Köszegi (2005), Eliaz and Spiegler (2006), demonstrates how one of the contracting parties (principal) can take advantage of the self-ignorance of the other party (agent). However, this literature does not explain why similar contractual arrangements may exist when the time-inconsistent agent is not naïve and cannot be deceived by the principal.

The primary goal of this paper is, first, to demonstrate that flexible contracts may be optimal even if the agent is perfectly aware of his time-inconsistency, correctly predicts his future choices and cannot be attracted by an offer which he will then regret; and, second, to explain that contrary to what is implied by the results in other papers, the principal may actually loose rather than gain from the agent’s naïveté. We show that a time-inconsistent agent may enjoy a lower price when his naïveté increases. Analyzing the individual decision-making framework, O’Donogue and Rabin (1999) demonstrate that time-inconsistent people may be better-off when they are naïve than when they are sophisticated. We show that this type of self-ignorance may be beneficial also in the principal-agent relationship.

We consider a contract between a monopolistic seller and a time-inconsistent buyer. The latter buys the good in period one and has a choice to pay for it in one of the two future periods: in cash (period two) or on credit (period three). Selecting a contract, the buyer makes two decisions separated in time: about the quantity in period one and about the price in period two. The situation may be illustrated by a credit card
user who at the end of a specified period must either pay the balance or revolve it at a higher interest rate. We argue that when the consumption precedes the payment, those with lower time-consistency not only tend to pay later but may also attribute a higher value to the good at the moment of buying. To extract the surplus from the inconsistent buyer with a high valuation, the seller offers a relatively cheap credit option for the consistent-type buyer with a low valuation. The inconsistent type would not like to select a cheap-credit contract because in that case he would over-use the credit option and end up paying a higher price. Flexibility of payment gives the seller an additional instrument to screen the buyers of different types, which not only increases the seller’s profit but also improves the efficiency of trade. In fact, flexibility substitutes for the use of the traditional discrimination instrument which implies downward quantity distortions.

The role of flexibility in our model is conceptually different from that in Heidhues and Köszegi (2010). Rather than being a decoy device, as it is for naïve agents in their paper, it becomes a screening device and allows the principal to reveal the information about the agent type. Instead of containing a trap for a naïve agent, it constitutes a threat for a sophisticated agent; instead of attracting the naïve agent to accept the contract in which he would pay more, it discourages the sophisticated agent from accepting the contract in which he would otherwise pay less. This allows us to argue that the principal may benefit from time-inconsistency of the agent even when the latter is perfectly sophisticated and cannot be deceived.

We also extend our original model to the case in which the agent is not only time-inconsistent but also partially naïve. The effect of the buyer’s naïveté on the seller’s profit is ambiguous. On the one hand, the naïve buyer tends to postpone the payment and pay more. (This positive effect is in line with the results in the above-mentioned literature.) From the other hand, buyer’s naïveté makes screening more complicated: when the buyer underestimates his propensity to pay more and later, it is more difficult to dissuade him from accepting the contract designed for another type. The positive effect, however, does not depend on the degree of naïveté but rather on time-inconsistency. If the buyer is naïve and not completely sophisticated, he buys more and pays a relatively higher price, but neither the quantity he buys nor the difference between the price he expects to pay and the price he actually pays change when he become more or less naïve. At the same time, the negative effect depends on the degree of naïveté: the seller has to leave a higher rent to the buyer when the latter becomes more naïve. This implies that even though the seller benefits from the buyer’s naïveté, he would nonetheless
prefer to deal with a buyer who is sufficiently sophisticated.

The rest of the paper is organized as follows. In Section 2, we explain why a time-inconsistent buyer may be willing to pay a relatively higher price and we show how the seller can use a flexible contract to extract more rent from such a buyer. In Section 3, we consider the situation in which the buyer is not only time-inconsistent but also partially naïve and compare the results to the case of a perfectly sophisticated buyer. Conclusions are given in Section 4.

2 Contracting with perfectly sophisticated time-inconsistent agents

2.1 Present-biased preferences: time-inconsistency or willingness to pay

Consider an individual who obtains and consumes \( q \) units of the good in period one and has to pay for it either \( t \) in period two (in cash) or \( T \) in period three (on credit). His valuation for the good is given by the function \( S(q) \) such that \( S' > 0, S'' < 0 \) and \( S(0) = 0 \). The individual discounts future payoffs with a hyperbolic parameter \( \beta \), such that \( 0 < \beta < 1 \); his intertemporal utilities in period one and two are

\[
S(q) - \beta t - \beta T
\]

\[
-t - \beta T. \quad (1)
\]

Hyperbolic discounting represents the time-preferences which are present-biased. Such preferences lead to time-inconsistency: in later periods, the individual may take actions which are not optimal from the perspective of earlier periods. In our case, if the choice between the two payments were made in period one, the individual would prefer to pay in period two whenever \( t \leq T \). However, in period two this is optimal only if \( t \leq \beta T \). Consequently, if \( \beta T < t < T \), there is a conflict between the period-one intention and the period-two action. This conflict increases when \( \beta \) becomes lower: the period-one preferences between \( t \) and \( T \) are

\[1\]These is an particular example of the \((\delta, \beta)\)-preferences analysed in Laibson (1997). To make the notation less cumbersome, we assume without loss of generality that the standard exponential discounting parameter \( \delta \) equals one. All the results in this paper are qualitatively similar for a non-trivial discounting parameter, \( \delta < 1 \).
independent of $\beta$; in period two, however, the individual with a lower $\beta$ may wish to postpone the payment even when the difference between $T$ and $t$ is sufficiently large. A low-$\beta$ individual is, therefore, more inclined to inefficient procrastination than an individual with a higher $\beta$.

Another consequence of present-biased preferences is that even if the individual could commit himself to paying in a given period (or alternatively, if there were only two periods and everyone should pay in period two), the maximal price that he would accept to pay increases when $\beta$ goes down.

The parameter $\beta$ plays a double role: that of discounting and that of time inconsistency. As a result, the individual with a lower $\beta$ has a higher willingness to pay for the good and is also more likely to pay later. This positive relation between valuation for the good and propensity to postpone the payment is a natural consequences of present-biased preferences if the consumption of the good precedes the payment. However, for the purpose of our analysis, it may be more convenient to have a model which allows us to disentangle the two roles of $\beta$.

### 2.2 Two-parameter model

Consider the following representation of the intertemporal preferences in period one and two:

$$vS(q) - t - T$$  \hspace{1cm} (3)

$$-t - \beta T.$$  \hspace{1cm} (4)

The parameter $v$ stands for the valuation that the individual attributes to the good and thus entirely determines his willingness to pay in period one. The parameter $\beta$ accounts for time-inconsistency and determines the willingness to procrastinate. When $v = 1/\beta$, the two representations, (1)-(2) and (3)-(4) are equivalent. More generally, $v$ could be higher (lower) than $1/\beta$ if the time interval between periods one and two is longer (shorter) than that between periods two and three. Even more generally, the parameter $v$ may represent both the discounting of future payments and some intrinsic valuation for the good.

We analyze a principal-agent relation in which the buyer (agent) has the time preferences described above. The buyer can be of two types: low $(\underline{v}, \underline{\beta})$ and high $(\overline{v}, \overline{\beta})$, such that $\underline{v} < \overline{v}$, and $\underline{\beta} > \overline{\beta}$. The higher-valuation type is also more inconsistent. The probability that the agent
is of low type is given by \( p \) and the probability that he is of high type equals \( 1 - p \).

The good is produced at a constant marginal cost \( c \) by a monopolistic seller (principal), who is indifferent about the intertemporal allocation of transfers and maximizes

\[
t + T - cq. \tag{5}
\]

Alternatively, one can think about a monopolistic bank or another credit organization which lends money to the agent who then buys the good on a competitive market.

Note that from the viewpoint of period one, there is no difference between cash and credit for both parties: for a given amount of payment, both the principal and the agent are indifferent about the period in which this payment is made.

To begin with, we assume that there is only one period in which the agent pays. This is actually equivalent to the framework in which the agent has the choice between the two periods but can commit himself to pay in a specific period at the moment of signing the contract.

### 2.3 Rigid contract

Consider the situation in which the agent buys \( q \) in period one and pays \( t \) in period two. If the principal observes the agent’s type, he can make him a take-it-or-leave-it offer with the efficient quantities \( q^* \) and \( \bar{q}^* \) given by \( vS'(q^*) = c \) and \( \bar{v}S'(\bar{q}^*) = c \), and the corresponding transfers are \( \tilde{t} = vS(q^*) \) and \( \tilde{\bar{t}} = \bar{v}S(\bar{q}^*) \). This contract allows the principal to implement the quantity that maximizes the surplus from trade and appropriate the entire surplus without leaving a positive rent to the agent.

If the agent’s type is private information, the efficient contract cannot be implemented because the low-type (\( \beta \)) agent would like to mimic the high-type. Let \((q, t)\) and \((\bar{q}, \bar{t})\) be the quantity and price that the principal offers to the low and high type respectively. The problem that the principal solves is:

\[
\max_{q,t,\bar{q},\bar{t}} \quad p(t - cq) + (1 - p)(\bar{t} - c\bar{q}) \tag{6}
\]

\[
vS(q) - t \geq 0 \tag{7}
\]

\[
\bar{v}S(\bar{q}) - \bar{t} \geq 0 \tag{8}
\]
where (7), (8) are individual rationality and (9), (10) incentive compatibility constraints.

This is the standard second-degree price discrimination problem: the only binding constraints are (7) and (10). Substituting for $t$ and $\bar{t}$ in (6) and differentiating with respect to $q$ and $\bar{q}$, we obtain

$$vS(q) - t \geq vS(\bar{q}) - \bar{t}$$

(9)

$$\bar{v}S(\bar{q}) - \bar{t} \geq \bar{v}S(q) - \bar{t}.$$  

(10)

The optimal contract has the usual characteristics of the principal-agent relation under adverse-selection: a downward distortion of the low-type quantity allows the principal to decrease the rent left to the high type. Inefficient quantities increases the expected profit of the principal; however, this profit is still lower than that under complete information.

### 2.4 Flexible contract

We are going to show now that a more elaborate structure of contract allows the principal to extract more or sometimes the whole surplus from the agent and decrease or eliminate the efficiency losses associated with asymmetric information. Let us consider a contract that consists of a quantity and a pair of transfers $(q, t, T)$ and allows the agent to pay either $t$ in period two or $T$ in period three. Instead of a usual quantity-transfer menu, the principal offers a menu of menus or a menu of flexible contracts. The contract gives the possibility to postpone the payment until period three and the difference $T - t$ can be interpreted as a credit price or a penalty for not paying in time. It is crucial that the choice between $t$ and $T$ is made in period two and not in period one.

In this simple framework, it is sufficient to consider the two contracts which have the following form: $(\bar{q}, \bar{t})$ and $(q, t, T)$; only the contract designed for the low type includes a credit option. The idea is to select the value of $\bar{t}$ and $T$ in such a way that if the contract $(q, t, T)$ is chosen, the low type will pay in cash and the high type will pay on credit. This requires $\bar{t} \leq T$ and $\bar{t} \geq T$, that is

$$\frac{t}{\beta} \leq T \leq \frac{\bar{t}}{\beta}.$$  

(13)

See, for example, Laffont and Martimort (2002), Chapter 2.
When (13) holds, the incentive-compatibility constraint (10) is modified as follows:

\[ \bar{v}S(\bar{q}) - \bar{t} \geq \bar{v}S(q) - \bar{T}, \]  

(14)

where \( \bar{t} \) is replaced by \( \bar{T} \). Since \( \bar{T} \geq \bar{t} \), the constraint is now more likely to be relaxed. This solution to this problem is summarized in the following proposition.

**Proposition 1:** The optimal low-type transfers are such that \( \bar{t} = vS(q) \), \( \bar{T} = \frac{vS(q)}{\beta} \) and the low type pays in cash.

Moreover, if \( \bar{\beta} < \bar{v}/\bar{v} \), the optimal contract implies the efficient quantities \( q = q^* \), \( \bar{q} = q^* \), and leaves no rent to the agent, that is \( \bar{t}_1 = \bar{v}S(\bar{q}) \). Otherwise, the low-type quantity is given by

\[ [\bar{v} - \frac{p}{1-p}(\bar{v} - \frac{v}{\beta})]S'(q) = c \]  

(15)

and the high-type agent enjoys a positive rent, that is \( \bar{t}_1 = \bar{v}S(\bar{q}) - (\bar{v} - \frac{v}{\beta})S(\bar{q}) \).

Comparing (15) to (12), one can see that the flexible contract introduces less distortions and hence is more efficient. Moreover, when the high-type consistency is sufficiently low relative to the ratio of the two types' valuations, the distortions disappear and everything happens as if the principal observed the type of the agent. The flexible contract increases the total surplus from trade under asymmetric information and also allows the principal to appropriate a larger part of this surplus.

To see more clearly the intuition behind this result, we rewrite the right-hand side of the constraint (14) as

\[ (\bar{v} - \bar{v})S(\bar{q}) - (\frac{\bar{v}}{\beta} - \bar{v})S(\bar{q}). \]  

(16)

The first term is the usual *information rent* that the principal has to leave to the high-type agent and the second term is the *time-inconsistency cost* which the high type will pay in terms of a higher price if he accepts the contract of the low type. If the difference between \( \bar{T} \) and \( \bar{t} \) can be made sufficiently high, which happens when \( \bar{\beta} \) is low enough, the time-inconsistency cost exceeds the information rent and the high type would not mimic the low type. When this occurs, no distortion is necessary to extract the rent. Even when the time-inconsistency cost does not completely eliminate the information rent, it still makes it lower. Rent-extraction becomes less important for the principal, who does not need to distort the low-type quantity as much as in the rigid contract.
Remark 1: The value of $\beta$ does not affect the results; all that matters is that the type with low valuation is more consistent than the high-valuation type. The infinitely small difference between the two $\beta$'s introduces a discontinuous increase in efficiency. For example, we can assume without loss of generality that the low type is perfectly consistent, that is $\beta = 1$.

Remark 2: Note that if $v = 1/\beta$, the quantities are efficient. If the difference in the willingness to pay for the good is entirely due to the discounting of the future payments, as it is in the hyperbolic discounting model, then the optimal contract is always efficient.

One interesting feature of this model is that the standard revelation principle, which establishes the equivalence between direct and indirect mechanisms for implementation, does not apply here. Indeed, instead of a simple direct revelation mechanism with full commitment (rigid contract), the principal prefers to use a more complicated two-period procedure in which the agent takes two decisions separated in time. Even though the agent completely reveals his type in period zero by selecting between the two contracts, the principal gives him another choice to make in period one.

The flexible contract facilitates revelation of private information because it gives the principal an additional instrument to screen the agent. The time-inconsistent agent has a conflict between his present and future preferences. The principal makes use of this conflict by offering the agent to take decisions in two different periods. Instead of a standard quantity-pair menu, the principal offers the agent a menu of menus without the possibility to commit to a specific choice in the future.

However, this additional screening instrument comes at the price of two constraints in (13), which are nothing other than the incentive-compatibility constraints for the two types in period one. It is the second of these constraints which may become binding. When this happens, the principal cannot further increase $T$ and has to leave some rent to the agent.

Another feature of the optimal flexible contract is that it may include options which are never used in equilibrium. In fact, the credit option $T$ is never used and everyone pays in period two. It can be shown that the efficiency of the flexible contract and the principal's profit actually decrease if the agent may have to use the credit option (for example, because of liquidity constraints in period two), but does not know in period one whether this is going to happen (for example, because of uncertainty about his period-two income).

The credit option is never used because even though the agent is time-inconsistent and cannot control his future actions, he can still correctly
anticipate them, that is the agent is sophisticated rather than naïve. Therefore, the principal cannot make the agent accept the contract in which the latter will end up paying more than he expected. However, even if the agent is perfectly sophisticated, the flexible contract may still be useful as a screening device. Instead of containing a trap for a naïve agent, it constitutes a threat for a sophisticated agent; instead of attracting the naïve agent to accept the contract in which he would pay more, it discourages the sophisticated agent from accepting the contract in which he would otherwise pay less.

3 Contracting with partially naïve time-inconsistent agents

To clarify the exposition, we assume that the low-valuation type is fully consistent, that is \( \beta = 1 \), and that the high type is not only inconsistent but also partially naïve: at the moment he signs a contract (period one), he underestimates his time-inconsistency, that is overestimates the value of his own \( \beta \). Let \( \tilde{\beta} \) stand for the perceived value of \( \beta \), such that \( \beta < \tilde{\beta} \leq 1 \). (Until now, we considered the case in which \( \beta = \tilde{\beta} \), that is the agent was perfectly sophisticated.) In period one, the naïve agent thinks that in period two he will behave as if his discounting factor were \( \tilde{\beta} \) rather than \( \beta \). The naïveté increases with \( \tilde{\beta} \) and the agent is totally naïve when \( \tilde{\beta} = 1 \).

We consider the contracts of the following form: \((\eta, t, T)\) and \((q, t, T)\). The main difference with the previous cases is that now the credit option \( T \) is included in each type’s contract. The principal may take advantage of the agent’s naïveté by offering him a contract in which the agent will pay more (and later) than what he has expected at the moment of accepting the contract. For this, \( t \) and \( T \) must satisfy two conditions. First, accepting the contract the agent has to think that in period one he would prefer to pay \( t \) rather than wait until period two and pay \( T \), i.e. \( t \leq \tilde{\beta}T \). Second, in period one the agent has to postpone the payment, i.e. \( t \geq \beta T \). Since the principal would like to make the agent pay as much as possible, that is to maximize \( T \), only the second constraint is binding: \( t = \beta T \).

The principal’s problem takes the following form:

\[
\max_{q, t} \left( p(t - cq) + (1 - p)(\frac{t}{\beta} - cq) \right) \\
\nu S(q) - t \geq 0
\]
The two contracts are designed in such a way that the inconsistent type pays on credit whatever the contract that he selects. The crucial difference between the two contracts is that accepting his contract the inconsistent types wrongly thinks that he will pay in cash; at the same time, he understand correctly that he will pay on credit if he selects the contract designed for the consistent type. When the inconsistent type is partially naïve, the optimal contract combines the elements of threat, which makes sense for a sophisticated agent, and of trap, which is useful when the agent is naïve.

For the same reason as before, the constraint (18) is binding: no rent is left for the low-valuation type. The relevant question is which of the two remaining constraints is binding and which is slack.

**Proposition 2:** The optimal high-type quantity is given by $\bar{v}S'(\bar{q}) = \bar{\beta}c$. and the optimal low-type transfers are $\bar{t} = vS(q), \bar{T} = \frac{vS(q)}{\bar{\beta}}$.

Moreover, if $\tilde{\beta} < \frac{v}{\bar{\nu}}$, the constraint (20) is slack and the contract implies the efficient low-type quantity, $q = q^*$, and $\bar{t} = vS(q), \bar{T} = \frac{vS(q)}{\bar{\beta}}$. Otherwise, the constraint (19) is slack and the low-type quantity is given by

$$[\nu + \frac{(1-p)}{p\beta}(\frac{v}{\beta} - \nu)]S'(q) = c$$

and $\bar{t} = vS(q) - (\nu - \frac{v}{\beta})S(q), \bar{T} = \frac{\bar{t}}{\beta}$.

Comparing Proposition 2 to Proposition 1, we can see the effect of time-inconsistency and naïveté on the quantities and transfers. As was shown above, when the agent is sophisticated, time-inconsistency tends to decrease quantity distortions and increase efficiency in flexible contracts. The effect of naïveté is quite the opposite: quantity distortions increase (upward for the inconsistent type and downward for the consistent type), when the degree of naïveté becomes more important. Moreover, when the agent is sufficiently naïve ($\tilde{\beta}$ is sufficiently high), a flexible contract leads to even higher distortions than a rigid contract. For example, if the agent is completely naïve, $q$ is given by
\[ [v + \frac{(1-p)}{p\beta} (v - \overline{v})]S'(q) = c, \]  
(22)

which is lower than the corresponding quantity in the rigid contract. Since the naïve type pays a relatively higher price than that paid by a sophisticated type ($\overline{\beta}$ instead of $\beta$), the rent-extraction becomes relatively more important for the principal, who is ready to make a larger sacrifice in terms of efficiency and distort the consistent-type quantity even more. Note also that even though time-inconsistency decreases distortions when the agent is sophisticated, it actually introduces additional distortions and leads to even higher inefficiency when the agent is naïve.

Another interesting characteristic is that the inconsistent agent may benefit from being more naïve. As $\beta$ goes down, the agent pays a lower price but buys exactly the same quantity $q$. The agent’s gains correspond to the principal’s losses. It is more difficult to use a credit option as a screening device when the inconsistent type becomes more naïve and does not understand that he would pay more if he accepts the consistent-type’s contract. The seller has to leave a larger rent to the inconsistent agent who is more naïve.

4 Conclusion

Although the existing literature on contracting with time-inconsistent agents explains how the principal can use flexible contracts to exploit the agent’s naïveté, it does not answer the question why such contracts may be useful when the agent is sophisticated and cannot be deceived. This paper analyses the case in which the time-inconsistent buyers may also attribute a higher valuation for the good and shows that flexible contract is optimal even if the buyer is completely sophisticated. Such a contract gives the principal an additional screening device and not only increases the principal’s profit but also decreases the quantity distortions and improves the efficiency of trade.

Moreover, the use of this screening device relies on the buyer’s sophistication. If the agent is not only time-inconsistent but also naïve, the principal has to leave him a higher rent in order to reveal his type as the degree of naïveté increases. The principal prefers to deal with the agent who is partially naïve but sufficiently sophisticated. This implies that self-ignorance may be beneficial and that time-inconsistent buyers may gain from being more naïve.
5 References


