WORKING PAPERS

Finite Horizon Holdup and How to Cross the River

Simon Martin
Karl H. Schlag

September 2017

Working Paper No: 1706

DEPARTMENT OF ECONOMICS

UNIVERSITY OF VIENNA

Finite Horizon Holdup and How to Cross the River

Simon Martin* Karl H. Schlag**

September 18, 2017

Abstract

When should one pay the ferryman? When should one pay for delivery of a good if there are no institutions or these are too costly to enforce contracts? We suggest to break up the transaction into many small rounds of investment and payment. We show that the efficient investment can be implemented in an ε-subgame perfect equilibrium for any given ε if there are sufficiently many rounds of investment. This shows that when the horizon is finite, the holdup problem that emerges from backwards induction is not robust. Equilibria with stable and robust strategies require more periods.

Keywords: holdup problem, ε-subgame perfect equilibrium, finite horizon, enforcement without contracts, gradualism

JEL Codes: D23, C72, L14

*University of Vienna. Email: simon.martin@univie.ac.at

**University of Vienna. Email: karl.schlag@univie.ac.at

1 Introduction

"Don't pay the ferryman,
Don't even fix a price,
Don't pay the ferryman,
Until he gets you to the other side"

Chris de Burgh, 1982

Chris de Burgh's 1982 pop song 'Don't pay the ferryman' is a reference to the ferryman Charon from Greek mythology, who took the deceased from one side of the river Styx to the other side, in exchange for a small fee (Nardo, 2002). When should one pay the ferryman? According to Chris de Burgh, not until he gets you to the other side, since he doesn't have an incentive to continue once you paid. However, also this clearly cannot be the solution to the problem as the passenger has no incentive to pay the ferryman once he has reached his destination.

A broad variety of economic problems, termed holdup problems, can be interpreted as a very similar situation. One agent, say, the seller, has to make an upfront investment in order to produce or deliver something valuable to another agent, the buyer. Who should move first? If the irrevocable investment is made first, the buyer has no incentive to pay anymore - he is already at the other side of the Styx, so why pay? Alternatively, the payment could be made first. But then why should the seller still make her investment? Holdup emerges. The Nash equilibrium prediction is that no trade will take place.

The existing solutions in economic research fall into three categories. First, institutions and enforceable contracts may solve the problem and recover the possibility for trade. Second, repeated interaction and reputational concerns may create sufficiently strong incentives to build up a lasting relationship (see Shapiro (1982) for an early theoretical consideration and Dulleck et al. (2011) and Palfrey and Prisbrey (1996) for more recent experiments). Third, if the duration of the relationship between buyer and seller is uncertain and investments and payments are split up into smaller pieces, then both sides may have an incentive to continue investing and paying until the game terminates randomly (Pitchford and Snyder, 2004).

These solutions each have their own drawbacks. Institutions are costly and the investment information need not be verifiable[1]. Shadow markets definitely lack the legal

[1]Note that in environments where costly institutions exist, it is actually better *not* to split up payments and investments. Appealing to the institution is no longer optimal in case of deviation if the relative gains are small, thus effectively eliminating the threat of subsequent enforcement.

enforceability aspect. With growing decentralized markets there may not be sufficient reputational concerns. Most trade is organized with known transaction and delivery dates.

Apart from Pitchford and Snyder (2004), the idea of splitting up the project is already informally suggested by Dixit and Nalebuff (1993) in the context of shadow markets. In their case, splitting serves as a tool to limit possible future losses if the business partner turns out not to be trustworthy, whereas in our environment the holdup problem emerges even absent any incomplete information about the opponent's type. Splitting up the project is also frequently observed in real life. The key for a house or a car is handed over only once the payment was made. This is an optimal strategy within our model. In job order contracting, the entire contract is split up into many subcontracts, and each is paid for upon completion. These contracts frequently involve equal sized parts, which is not a solution in our context unless each of these parts are smaller than ε.

In this paper, we propose a novel solution to the holdup problem that is particularly simple to implement and moreover provides insights about when existing investment schedules are optimal or not. We approach the problem by splitting up the total investment into smaller parts and using the solution concept of ε-subgame perfect equilibria (Mailath et al., 2005) instead of the standard subgame perfect equilibrium. In a subgame perfect equilibrium (SPE) no player can obtain strictly more in any subgame by deviating. The concept of an ε-SPE makes a minor adjustment. There is a constant ε, which is typically small, such that no player can obtain at least ε more in any subgame by deviating. Intuitively, why bother about arbitrarily small gains (figuratively often referred to as peanuts)? The ε threshold can capture deliberation costs, costs of embarrassment costs due to deviating from the suggested plan and unmodeled uncertainty about how others react to own deviation.

Just replacing SPE by ε-SPE does not solve the holdup problem as ε is assumed to be small, but gains from deviation, should trade be valuable, are expected to be large. We split up total investment and payment into smaller pieces; as we now show, that alone would not suffice either.

Given the finite horizon, the holdup problem originates in the final period. Lack of future consequences make costly actions (such as investment or payment) in the last period non optimal. Using the backwards induction logic, costly actions in earlier periods can also not be sustained and the holdup problem occurs. However, the first step in this logic, the behavior in the final period, is not necessarily compelling if the gains of deviation are only small. This is the intuition behind our approach. We split up the total investment into many small ones in order to make that last investment small and

invoke an equilibrium concept in which possibilities to only minorly improve payoffs are not followed.

Total investment x as well as the payment p are split up into sequences of T smaller investments $(x_t)_{t=1}^T$ and payments $(p_t)_{t=1}^T$. We then look for sequences such that in each round the seller wishes to continue to invest and the buyer wishes to continue to pay when facing the alternative of terminating the relationship. However this would not solve the problem if we use SPE as solution concept, as the original problem would identify itself in the final period. This would then unravel back to the first.

When adopting ε-SPE instead as solution concept the problem is solved for any $\varepsilon > 0$. We eliminate the hold problem where it originates, namely in the final period. By making the final payment small enough, the buyes does not wish to deviate, thus providing incentives for earlier investments and payments. We address the following questions: Under which conditions does an equilibrium exist? What is the optimal investment and payment schedule from the buyer's or seller's point of view? Which shape do they have?

In the economic problems we consider, there is a single investment and transaction possibility between a seller and a buyer, which may be broken down into smaller pieces. In particular, we rule out the possibility of future interaction and therefore any reputational concerns. Moreover, we assume that no institutions enforcing any contractual arrangements exist, based on the idea that efforts are either not observable, not verifiable, or enforcement is simply too costly.

The buyer obtains the irrevocable right to exclusively use or consume whatever the seller produced up to period t. We allow for a general production technology $f : \mathbb{R}_+ \to \mathbb{R}_+$ and normalize it such that the social surplus $f(x)-x$ is maximized at $x^* = 1$. For instance, the example with the ferryman can be modeled using the production function f such that $f(x) = 0$ for $x < 1$, and $f(x) = 1 + s$ for $x \geq 1$ where s is some parameter with $s > 0$.

In an ε-SPE of the game, each player has to be willing to invest or pay the specified amount in each period t. We show that for any admissible payoff and any $\varepsilon > 0$, an ε-SPE exists provided T is sufficiently large. For the ferryman example, only two periods are needed: A large initial investment and payment, followed by a small additional investment and payment. The buyer does not have an incentive to deviate after the first investment, because it does not yet have any value for him. Neither player deviates from the last investment, respectively payment, because gains from deviation are very small and we assume that players don't run after peanuts.

We also present the sequence of investments and payments that implements this ε-SPE. Moreover, we show that for production functions with increasing social surplus the

investments and payments are decreasing over time. Constant investments in each period work as long as there is a small payment in the last period, but require more periods than the investment plan we describe. The reason is that in that case the fact that the outside option becomes increasingly attractive for the buyer is not accounted for.

On top of ε-SPE, we introduce the stricter notion of terminal ε-SPE, in which players are willing to forego small gains only in the last period. All our main results still hold, showing that solving the holdup problem in the last period solves it altogether. Terminal ε-SPE are robust to best responses, but this comes at the expense of requiring much more periods to be implemented for production functions with increasing surplus. For the ferryman technology, there is no such trade-off because of the steep increase in social surplus at the end of the project, deterring the buyer from deviation even if high payments are needed. Thus, even in a terminal ε-SPE never more than 3 periods are necessary.

The ability to achieve cooperation in finite horizons by considering only deviations that ensure a minimal additional payoff was highlighted by Mailath et al. (2005) in the context of a Prisoners' dilemma. In their paper, the ε can be taken small if supergame payoffs are defined as average stage game payoffs. The strategic choice to split up an investment into small parts to lower incentives to deviate was informally suggested by Dixit and Nalebuff (1993). Our analysis shows that typically only the last investments and payments need to be small. For other applications of ε-optimality see e.g. Radner et al. (1980), Baye and Morgan (2004), Barlo and Dalkiran (2009) and Milgrom (2010). Moreover, maximin and minimax regret often select ε-optimal solutions, e.g the robust policies identified in Bergemann and Schlag (2011) are ε-optimal.

In terms of the application, the closest paper to ours is by Pitchford and Snyder (2004) who show how one can get around the holdup problem by splitting up the entire investment into many small investments when the buyer seller relationship is indefinite. In their paper the relationship is assumed in each round to continue with a sufficiently large probability, e.g., because project continuation is "conditional on past experience" (p. 89). Conversely, our model also allows for efficient investment in the more realistic scenario where the termination date is known. Although the mechanisms for ensuring investment are very different, our paper shares with theirs the decreasing investment. However, in Pitchford and Snyder (2004) the investments converge to 0, whereas our smallest investment is ε.

Our modelling assumption that investments are non-negative and thus irrevocable carries some resemblance to the framework of Lockwood and Thomas (2002). Two players infinitely interact and each player's payoff decreases in her own action but increases in

the action of the other player. Since investments in their model are irrevocable, the worst credible threat is to stop cooperating in the future. Players are incentivized by the perspective of future cooperation of the other player. Conversely, we show that the efficient outcome can be attained if we allow that players are willing to give up small payoffs relative to optimality. Moreover, we don't need to split up the project into arbitrarily many parts. For the ferryman example, only two periods are enough for any $\varepsilon > 0$.

Admati and Perry (1991) overcome the impossibility of socially desirable investments into a joint project by making costs contingent on project completion. In this environment, costs are not effectively sunk, because they will be recovered entirely if the project does not complete. Such an environment requires enforceable contracts or an institution that ensures recovery in case of breakup, which may be both costly to enact and to call upon if needed. Conversely, our model does not require contracts or institutions of any kind. The structure of the game allows implementation of the efficient outcome once we split up the entire project into smaller parts.

We proceed as follows. We lay out our main model in Section 2. The main results are in Section 3. Various numerical examples are shown in Section 4. Section 5 provides a discussion of our results, and Section 6 concludes.

2 Model

Consider an interaction between a buyer and a seller that lasts a given number of T periods. In each period t the seller can make an investment x_t and then the buyer can make a payment p_t. We denote the period t time units before the end as period t, e.g., the last payment is p_1. Both investments and payments are restricted to be positive or zero. We denote the sequence of investments and payments as $(x_t)_{t=1}^T$ and $(p_t)_{t=1}^T$.

Investments by the seller provide value to the buyer according to a production function $f : \mathbb{R}_+ \to \mathbb{R}_+$, where x is the sum of investments and $f(x)$ is weakly increasing in x. We assume that $f(0) = 0$ and that the surplus $f(x) - x$ is uniquely maximized at $x^* = 1$.

Two types of production functions are of special interest. (i) The maximal surplus exceeds the value of any smaller investment. Formally, $f(1) - 1 \geq f(x)$ for all $x < 1$. We refer to this production function as the ferryman technology as it generalizes the situation underlying the boat ride where the passage has no value until the passenger reaches the other side. Any investment less than the efficient one is worse for the buyer than paying the efficient investment. (ii) The willingness of the buyer to pay for each additional investment is larger than the cost for the seller for this investment. Formally,

$f(y) - y$ is weakly increasing in y. Production functions of this type arise naturally in many environments.

We assume throughout that the buyer owns the property rights of any investment made by the seller. In particular, in any period the value of the good to the buyer is given by the sum of past investments.

The timing of the game is as follows. The game consists of at most T periods. In each period the seller moves first and either chooses an investment or chooses to break up the relationship. If the seller chooses to invest then the buyer observes this investment and either chooses a payment for the seller or chooses to break up the relationship. This ends the period. If within this period one of the two players has chosen to break up then the game ends. If neither of the two parties has chosen to break up the relationship and there have been less than T periods then the next period begins with rules as described above.

Payoffs are determined as follows. The cost of investment is born by the seller at the beginning of the period. The benefits of the investment for the seller are realized at the end of a period. Whenever there is a break up or all T periods have passed then each player gets an outside option normalized to 0.

Specifically, if either the seller or the buyer breaks up in period t then payoffs of the two players are given by $u_s = 0$ and

$$u_b = f\left(\sum_{i=t-1}^{T-t} x_i\right).$$

If neither of the players breaks up then, from the perspective of the beginning of period t and assuming that the total sum of investments equals 1, the seller's payoff is given by

$$u_s = \sum_{i=1}^{t}(p_i - x_i) \tag{1}$$

and the buyer's payoff is given by

$$u_b = f(1) - \sum_{i=1}^{t} p_i. \tag{2}$$

Note that according to the unique subgame perfect equilibrium no investments will be made. After the last investment was made, the buyer doesn't have an incentive to make the final payment p_1 if $p_1 > 0$, since the project is already finished. Anticipating no final payment, the seller of course doesn't want to make the final investment if $x_1 > 0$. By backwards induction, we find that investments have to be 0 in all periods.

As we will show in the next section, employing ε-SPE instead radically changes our insights.

2.1 Equilibrium concept

Our solution concept is ε-SPE (Mailath et al., 2005), which applies ε-Nash equilibrium to all subgames. A player only deviates if he gains more than a given threshold ε. Formally, the concept is defined as follows:

Definition 1. *Denoting as A_i the set of actions for player i and as σ^*_{-i} the strategy profile of the other players, a strategy profile σ^* is an* ***ε-Nash equilibrium*** *if $u_i(a_i, \sigma^*_{-i}) \leq u_i(\sigma^*) + \varepsilon$, $\forall a_i \in A_i$, $\forall i$.*

Definition 2. *A strategy profile σ^* is an* ***ε-subgame perfect equilibrium*** *if it induces an ε-Nash equilibrium in every subgame.*

Possible reasons why one might not want to deviate when the gains are small or negligible include: i) deliberation costs, ii) costs from being embarrassed (or alternatively, extra utility from staying at the proposed equilibrium) and iii) existence of possibility to prevent small deviations (e.g., small retaliation is available).

In the following we will limit attention to equilibria in which any deviation is punished by discontinuing the relationship. There is no loss of generality in this assumption when searching for outcomes that can be sustained in some ε-SPE. In order to check whether a given sequence of investments and payments can be supported in an ε-SPE, we need to make sure that neither player has an incentive to deviate in any period.

We define $\sum_{k=i}^{j} h_k = 0$ for $i > j$. In order to guarantee that the seller does not have an incentive to deviate from an investment x_t, it has to hold that:

$$-x_t + p_t + \sum_{i=1}^{t-1}(p_i - x_i) \geq -\varepsilon. \tag{3}$$

Conversely, for the buyer we require that

$$f(1) - p_t - \sum_{i=1}^{t-1} p_i \geq f\left(1 - \sum_{i=1}^{t-1} x_i\right) - \varepsilon. \tag{4}$$

As a robustness check for our analysis, we also introduce the stricter notion of terminal ε-SPE, in which only the buyer is willing to give up peanuts in his very last move. We discuss additional motivation for this alternative solution concept in Section 5. Importantly, this shows that only the buyer needs to be willing to forego small gains in order to solve the finite horizon holdup problem; the seller still plays strictly best responses in each move.

Definition 3. *A strategy profile σ^* is a* ***terminal ε-subgame perfect equilibrium*** *if it induces a Nash equilibrium in every subgame except the last subgame, and an ε-Nash equilibrium in the last subgame.*

Therefore, in a terminal ε-SPE, the seller does not have an incentive to deviate from an investment x_t as long as it holds that

$$-x_t + p_t + \sum_{i=1}^{t-1}(p_i - x_i) \geq 0. \tag{5}$$

Conversely, for the buyer we require that in each period $t > 1$ it holds that

$$f(1) - p_t - \sum_{i=1}^{t-1} p_i \geq f\left(1 - \sum_{i=1}^{t-1} x_i\right) \tag{6}$$

and in period 1 it holds that

$$\begin{aligned} f(1) - p_1 &\geq f(1) - \varepsilon \\ p_1 &\leq \varepsilon. \end{aligned} \tag{7}$$

Given these definitions, we are now ready to describe equilibria in this game.

3 Efficient Equilibria

We first characterize the set of feasible equilibrium payoffs.

Proposition 1. *(i) In any (terminal-)ε-SPE, the sum of payoffs is bounded from above by $f(1) - 1$. (ii) In any ε-SPE, each player's equilibrium payoffs are within the interval $[-\varepsilon, f(1) - 1 + \varepsilon]$. (iii) In any terminal ε-SPE, each player's equilibrium payoffs are within the interval $[0, f(1) - 1]$.*

Proof. (i) Prices are only transfers in this model and thus exactly cancel out in the sum of payoffs. Formally we have that

$$\begin{aligned} u_s + u_b = \sum_{i=1}^{T}(p_i - x_i) + f\left(\sum_{i=1}^{T} x_i\right) - \sum_{i=1}^{T} p_i = \\ f\left(\sum_{i=1}^{T} x_i\right) - \sum_{i=1}^{T} x_i \leq f(1) - 1 \end{aligned}$$

where the last inequality follows from the assumption that $f(x) - x$ is maximized at $x = 1$.

(ii) This follows immediately from the no-deviation constraints at period T and the maximal sum of payoffs.

(iii) This follows immediately from the no-deviation constraints at period T and the maximal sum of payoffs. □

Equipped with this knowledge about feasibility, we now discuss conditions for equilibrium existence under the ferryman technology, and then move production functions where social surplus increases with investments.

3.1 Ferryman technology

As introduced above, we consider production functions where $f(1)-1 \geq f(x)$ for all $x < 1$ in this subsection. These include the one underlying the boat ride story. We first give an intuitive overview of the results and afterwards we provide a more formal treatment.

Contrary to production functions with increasing social surplus, in the ferryman technology there is no trade-off between robustness and the number of periods required: Both the ε-SPE and the terminal ε-SPE can be implemented with at most 3 periods. Due to the steep increase in surplus at $x = 1$, the buyer does not have an incentive to deviate from paying throughout the project.

Only 2 periods are necessary if the buyer's payoff is sufficiently high; in that case, his first payment p_2 is sufficiently low to deter deviation to non-payment. As the buyer's payoff decreases, p_2 increases until eventually he would deviate. In that case, a warm-up period with some initial payment but zero investments is necessary.

Note that the ferryman example with $f(x) = (1+s) \cdot \mathbb{1}_{x \geq 1}$ is a special case where two periods are always enough. Even getting close to the other side of the river is of no value to the passenger if he cannot swim. Thus neither player has an incentive to deviate from the large first investment, respectively payment.

Taking up the example given in the introduction, a house or a car without keys is useless to the buyer. While passing over the keys is almost costless for the seller, receiving the keys for the house or car is extremely valuable for the buyer. This stark asymmetry at the end of the project makes the small but final transaction very powerful. The buyer is willing to make even substantial initial payments, anticipating large future gains for relatively small payments. Consequently, only very few periods are needed.

Formally, we have the following result:

Proposition 2. *Assume $f(1) - 1 \geq f(x)$ for all $x < 1$ and $\varepsilon > 0$.*

(i) Let $w_1 = f(1 - 2\varepsilon) - \varepsilon$. There exists an ε-SPE with efficient investments and $u_b = w$ if $w \in [w_1, f(1) - 1 + \varepsilon]$ and there are least 2 periods, or if $w \in [-\varepsilon, w_1)$ and there are at least 3 periods. In the corresponding ε-SPE, investments are $x_1 = 2\varepsilon$ and $x_2 = 1 - 2\varepsilon$, $p_1 = \varepsilon$, and the remaining payments are such that the buyer does not deviate and $u_b = w$ is obtained.

(ii) Let $w_2 = f(1-\varepsilon)$. There exists a terminal ε-SPE with efficient investments and $u_b = w$ if $w \in [w_2, f(1)-1]$ and there are at least 2 periods, or if $w \in [0, w_2)$ and there are at least 3 periods. In the corresponding ε-SPE, investments are $x_1 = \varepsilon$ and $x_2 = 1-\varepsilon$, $p_1 = \varepsilon$, and the remaining payments are such that the buyer does not deviate and $u_b = w$ is obtained.

(iii) The above claims cannot be established using less periods.

Proof. (i) We establish existence under 3, respectively 2, periods. For more then 3, respectively 2, periods, existence follows by setting investments and payments to 0 in all additional periods. In this case, no player has an incentive to deviate by construction.

Fix $x_1 = 2\varepsilon$ and $p_1 = \varepsilon$, which is the maximal payment and investment that deters deviation from either player in period 1 in an ε-SPE. In order to reach a total investment of 1, we set $x_2 = 1-2\varepsilon$. The buyer does not have an incentive to deviate on the payment of p_2 as long as it holds that

$$\begin{aligned} f(1) - p_1 - p_2 &\geq f(1-2\varepsilon) - \varepsilon \\ p_2 &\leq \overline{p_2} := f(1) - f(1-2\varepsilon). \end{aligned}$$

The payment p_2 also needs to be such that a final payoff of $u_b = w$ is obtained, i.e.:

$$\begin{aligned} f(1) - p_1 - p_2 &= w \\ p_2 &= f(1) - \varepsilon - w. \end{aligned}$$

so deviation is deterred as long as $p_2 \leq \overline{p_2}$ or alternatively

$$\begin{aligned} f(1) - \varepsilon - w &\leq f(1) - f(1-2\varepsilon) \\ w &\geq f(1-2\varepsilon) - \varepsilon = w_1 \end{aligned}$$

and therefore in that case 2 periods suffice. Otherwise we need to set $p_2 = \overline{p_2}$ in order to deter deviation, and make the remaining payment in an additional period 3 where $p_3 = f(1-2\varepsilon) - \varepsilon - w$, which is always positive and deters deviation by construction.

(ii) We establish existence under 3, respectively 2, periods. For more then 3, respectively 2, periods, existence follows by setting investments and payments to 0 in all additional periods. In this case, no player has an incentive to deviate by construction.

Fix $x_1 = \varepsilon$ and $p_1 = \varepsilon$, which is the maximal payment and investment that deters deviation from either player in period 1 in a terminal ε-SPE. In order to reach a total investment of 1, we set $x_2 = 1-\varepsilon$. The buyer does not have an incentive to deviate on

the payment of p_2 as long as it holds that

$$f(1) - p_1 - p_2 \geq f(1-\varepsilon)$$
$$p_2 \leq \hat{p_2} := f(1) - f(1-\varepsilon) - \varepsilon.$$

The payment p_2 also needs to be such that a final payoff of $u_b = w$ is obtained, i.e.:

$$f(1) - p_1 - p_2 = w$$
$$p_2 = f(1) - \varepsilon - w.$$

so deviation is deterred as long as $p_2 \leq \hat{p_2}$ or alternatively

$$f(1) - \varepsilon - w \leq f(1) - f(1-\varepsilon) - \varepsilon$$
$$w \geq f(1-\varepsilon) = w_2$$

and therefore in that case 2 periods suffice. Otherwise we need to set $p_2 = \hat{p_2}$ in order to deter deviation, and make the remaining payment in an additional period 3 where $p_3 = f(1-\varepsilon) - w$, which is always positive and deters deviation by construction.

(iii) The proof for (i) and (ii) shows that 2 periods are not enough when w is sufficiently high. Thus we only need to show that also 1 period is never enough. In that case, we need to have that $x_1 = 1$. A necessary condition for non-deviation of the seller is that $p_1 \geq 1 - \varepsilon$. However, in that case the buyer would deviate to non-payment since $-(1-\varepsilon) < -\varepsilon$. Therefore, 1 period cannot suffice to get a total investment of 1. □

3.2 Production functions with increasing surplus

A broad class of production functions does not exhibit such substantial surplus gains at the very end of the project, but rather surplus is generated along throughout. In this subsection we consider production functions in which the surplus is increasing. We first give an intuitive overview of the results and afterwards we provide a more formal treatment.

Similar to the ferryman technology, the investment and payment schedule that requires the minimal number of periods features a very small final payment and, importantly, a pay-as-you-go schedule during the project; the seller is exactly compensated for his investment in each period. This is the maximal amount the buyer is willing to pay in each period and therefore minimizes the number of periods required. Investments decrease over time, because the outside option of breaking up the relationship because increasingly attractive for the buyer as the project value increases. Therefore, constant investments

in each period require considerably more periods. The binding constraint remains the penultimate period, so we have to add periods until that final payment is sufficiently small to deter deviation to non-payment by the buyer. Thus we cannot take advantage of otherwise decreasing payments, prolonging the project duration.

Contrary to ferryman technology and as will be illustrated in numerical examples below, for production functions with increasing surplus there is trade-off between robustness and the number of periods required: terminal ε-SPE require more periods than ε-SPE. In a terminal ε-SPE, only smaller amounts can be paid (and thus also invested) in order to deter deviation, and therefore the project implementation takes longer.

Formally, we have the following results:

Proposition 3. *Let $f(y)-y$ be weakly increasing in y, $\varepsilon > 0$ and $w \in [-\varepsilon, f(1)-1+\varepsilon]$. Then there exists a $T \in \mathbb{N}$ such that*

i) the following sequence of investments and payments constitute an ε-SPE with efficient investments and payoffs $u_b = w$: $x_1 = 2\varepsilon$, $p_1 = \varepsilon$,

$$p_t = x_t = f\left(1 - \sum_{i=1}^{t-2} x_i\right) - f\left(1 - \sum_{i=1}^{t-1} x_i\right),$$

for $2 \leq t \leq T-1$, $x_T = 1 - \sum_{i=1}^{T-1} x_i$ and

$$(p_T, p_{T+1}) = \begin{cases} \left(f(1-\sum_{i=1}^{T-2} x_i) - f(1-\sum_{i=1}^{T-1} x_i), f(1) - \sum_{t=1}^{T} p_t - w\right) \\ \qquad \text{if } w \leq f(1-\sum_{i=1}^{T-1} x_i) - \varepsilon \\ \left(f(1) - \sum_{t=1}^{T-1} x_t + \varepsilon - w, 0\right) \text{ otherwise} \end{cases}$$

and immediate termination whenever anyone deviates;

ii) for any $T' < T$, a ε-SPE with $\sum_{i=1}^{T'} x_i = 1$ and payoffs $u_b = w$ does not exist.

Proof. i) The structure of the proof is as follows: show how T is determined; that the seller doesn't have an incentive to deviate in any period; that the buyer doesn't have an incentive to deviate in any period; and finally that indeed the buyer obtains $u_b = w$.

We define a sequence z_t as follows: $z_1 = 2\varepsilon$ and $z_t = f(1 - \sum_{i=1}^{t-2} z_i) - f(1 - \sum_{i=1}^{t-1} z_i)$ for $t \geq 2$. It follows readily that $z_t \geq 0$ for all t. Note that z_t is increasing in t since the assumption that $f(y) - y$ is increasing implies that

$$f\left(1 - \sum_{i=1}^{t-1} z_i\right) - \left(1 - \sum_{i=1}^{t-1} z_i\right) \geq f\left(1 - \sum_{i=1}^{t} z_i\right) - \left(1 - \sum_{i=1}^{t} z_i\right)$$

$$f\left(1 - \sum_{i=1}^{t-1} z_i\right) - f\left(1 - \sum_{i=1}^{t} z_i\right) \geq \sum_{i=1}^{t} z_i - \sum_{i=1}^{t-1} z_i$$

$$z_t \geq z_{t-1}.$$

Since 1 is the unique maximizer of $f(y) - y$, $z_2 > 0$. Therefore, there exists a unique K such that

$$\sum_{i=1}^{K-1} z_i < 1 \leq \sum_{i=1}^{K} z_i.$$

Setting $T = K$ uniquely identifies T. Note that, by construction of z_t, we have that $x_t = z_t$ for all $t \leq T-1$, and $x_T = 1 - \sum_{i=1}^{T-1} x_i = 1 - \sum_{i=1}^{T-1} z_i$.

We now consider seller optimality in each period. Note that, by definition of x_t and p_t for $t \leq T-1$, it holds that

$$\sum_{i=1}^{t} x_i = \sum_{i=1}^{t} p_i + \varepsilon.$$

and it also holds that

$$\sum_{i=1}^{t} x_i = f(1) - f\left(1 - \sum_{i=1}^{t-1} x_i\right) + 2\varepsilon.$$

The seller's continuation payoff in period t is given by

$$u_s^t = \sum_{i=1}^{t} (p_i - x_i)$$

and therefore the seller doesn't have an incentive to deviate as long as it holds that

$$\sum_{i=1}^{t} (p_i - x_i) \geq -\varepsilon$$

which is satisfied in all periods $t \leq T-1$ by definition of x_t and p_t. We are thus left with showing that the seller also doesn't have an incentive to deviate in period T, i.e., that it holds that

$$x_T = 1 - \sum_{i=1}^{T-1} x_i \leq p_T.$$

Define $\bar{w} = f\left(1 - \sum_{i=1}^{T-1} x_i\right) - \varepsilon$, i.e., as the cutoff point of w such that for $w \leq \bar{w}$, the remaining payments are split across p_T and p_{T+1} and for $w \geq \bar{w}$, the entire payment is made in p_T and $p_{T+1} = 0$.

Then for the case where $w \leq \bar{w}$, we need to show that

$$1 - \sum_{i=1}^{T-1} x_i \leq f\left(1 - \sum_{i=1}^{T-2} x_i\right) - f\left(1 - \sum_{i=1}^{T-1} x_i\right)$$

which holds since

$$1-\sum_{i=1}^{T-1}x_i = 1-\left(\sum_{i=1}^{T}z_i - z_{T-1}\right) \leq f\left(1-\sum_{i=1}^{T-2}x_i\right) - f\left(1-\sum_{i=1}^{T-1}x_i\right)$$

$$1-\sum_{i=1}^{T}z_i + f\left(1-\sum_{i=1}^{T-2}x_i\right) - f\left(1-\sum_{i=1}^{T-1}x_i\right) \leq f\left(1-\sum_{i=1}^{T-2}x_i\right) - f\left(1-\sum_{i=1}^{T-1}x_i\right)$$

$$1 \leq \sum_{i=1}^{T}z_i = \sum_{i=1}^{K}z_i$$

which holds by definition of K.

For the case where $w > \bar{w}$, we need to show that

$$1-\sum_{i=1}^{T-1}x_i \leq f(1) - \sum_{i=1}^{T-1}x_i + \varepsilon - w$$

which holds since $w \leq f(1) - 1 + \varepsilon$. Thus the seller never has an incentive to deviate.

We now show that also the buyer never has an incentive to deviate. Generally, this holds as long as his continuation payoff

$$f(1) - \sum_{i=1}^{t}p_i$$

exceeds his outside option

$$f\left(1-\sum_{i=1}^{t-1}x_i\right) - \varepsilon.$$

By definition of p_t and x_t, this holds for all $t \leq T-1$. For the case where $w \leq \bar{w}$, we are left with showing that

$$f(1) - \sum_{i=1}^{T}p_i \geq f\left(1-\sum_{i=1}^{T-1}x_i\right) - \varepsilon$$

which holds since this can be rewritten as

$$f(1) - \sum_{i=1}^{T-1}x_i + \varepsilon - p_T =$$

$$f(1) - \left(f(1) - f\left(1-\sum_{i=1}^{T-2}x_i\right)\right) + \varepsilon - \left(f\left(1-\sum_{i=1}^{T-2}\right) - f\left(1-\sum_{i=1}^{T-1}\right)\right) =$$

$$f\left(1-\sum_{i=1}^{T-1}x_i\right) + \varepsilon \geq f\left(1-\sum_{i=1}^{T-1}x_i\right) - \varepsilon$$

$$2\varepsilon \geq 0.$$

For the case where $w > \bar{w}$, the condition can be rewritten as

$$\begin{aligned} f(1) - \sum_{i=1}^{T} p_i = \\ f(1) - \sum_{i=1}^{T-1} p_i - p_T = \\ f(1) - \sum_{i=1}^{T-1} p_i - \left(f(1) - \sum_{i=1}^{T-1} p_i - w \right) = \\ w \geq f\left(1 - \sum_{i=1}^{T-1} x_i\right) - \varepsilon \end{aligned}$$

which holds whenever $w > \bar{w} = f\left(1 - \sum_{i=1}^{T-1} x_i\right) - \varepsilon$. Thus also the buyer doesn't have an incentive to deviate.

Finally, by definition of p_t and in particular p_T and p_{T+1}, we always have that $u_b = f(1) - \sum_{i=1}^{T+1} p_i = w$.

ii) The sequence of investments and payments is such that both the seller and buyer constraints are binding in each period. Suppose, by means of contradiction, that for some $T' < T$, a ε-SPE with efficient investments and $u_b = w$ exists. This requires that either the seller or the buyer constraint is not binding in some period t. Assume that both constraints are still binding in all previous periods. Suppose that the buyer constraint is not binding in period t such that $p'_t < p_t$. This also reduces the maximal investment the seller is willing to make to $x'_t = x_t - (p_t - p'_t)$. This, in turn, reduces also the maximal payment in period $t+1$, i.e., it holds that $p'_{t+1} \leq p_{t+1}$ where

$$p'_{t+1} = f(1) - f\left(1 - \sum_{i=1}^{t-1} x_i - x'_t\right) + \varepsilon + \sum_{i=1}^{t-1} p_i + p'_t$$

and

$$p_{t+1} = f(1) - f\left(1 - \sum_{i=1}^{t-1} x_i - x_t\right) + \varepsilon + \sum_{i=1}^{t-1} p_i + p_t$$

since the inequality simplifies to

$$\begin{aligned} f\left(1 - \sum_{i=1}^{t-1} x_i - x_t\right) + p'_t \leq f\left(1 - \sum_{i=1}^{t-1} x_i - x'_t\right) + p_t \\ f\left(1 - \sum_{i=1}^{t-1} x_i - x_t\right) \leq f\left(1 - \sum_{i=1}^{t-1} x_i - x_t + p_t - p'_t\right) + p_t - p'_t \end{aligned}$$

which holds by the assumption that $f(y) - y$ is increasing.

Therefore, also $x'_{t+1} \leq x_{t+1}$ etc. The same holds if we start with assuming that first the seller constraint is not binding.

Hence, in period T we have that

$$x'_T = 1 - \sum_{i=1}^{T} x'_i \geq x_T = 1 - \sum_{i=1}^{T} x_i.$$

The hypothesis that still $u_b = w$ is attained requires that $\sum_{i=1}^{T} p'_i = \sum_{i=1}^{T} p_i$. Ensuring that the buyer does not have an incentive to deviate in period T requires that

$$f(1) - \sum_{i=1}^{T} p_i \geq f(x_T) - \varepsilon.$$

As the left-hand side is identical both in the sequence of investments as payments with binding constraints and without binding constraints, but on the right-hand side $x'_T \geq x_T$, we cannot have that this inequality is satisfied for non-binding constraints if it was not also already satisfied under binding constraints. Therefore, we cannot have a ε-SPE with $T' < T$.

□

Remark: The investments in Proposition 3 are increasing in t for $t \in [T-1, 1]$, i.e., decreasing over time.

Besides illustrating the requirement of a minimal number of periods, Proposition 3 also shows the sequence of investments and payments such that an ε-SPE exists. Key to proving the result is the assumption that $f(y) - y$ is increasing, which guarantees that as we keep adding periods with non-zero investments, eventually we the sum of investments exceeds 1.

We now show that our main results also hold for terminal ε-SPE.

Proposition 4. *Let $f(y) - y$ be weakly increasing, $\varepsilon > 0$ and $w \in [0, f(1) - 1]$. Then there exists a $T \in \mathbb{N}$ such that*

i) for any $T' \geq T$, a terminal ε-SPE with efficient investments and payoffs $u_b = w$ exists and

ii) for any $T' < T$, a terminal ε-SPE with efficient investments and payoffs $u_b = w$ does not exist.

Proof. See appendix. □

4 Examples

We now present various numerical examples, both for ε-SPE and terminal ε-SPE. In particular, in Table 1 we show the minimal number of periods required such that the respective equilibrium in which the buyer obtains all the surplus exists, given ε and a production function f. The ferryman technology always requires just 2 periods (see Proposition 2).

The number of periods required decreases in ε. As player's interpretation of 'how small' peanuts are which are not worth running after, final payments and investments can be larger, and therefore also subsequent investments and payments. Terminal ε-SPE require more periods than ε-SPE, since investments increase much slower because of the requirement of strict optimality in all periods except the first. The difference is particularly striking for the concave production function.

Highway procurement is a typical example where job order contracting is frequently employed. In the terminology employed in this paper, the project may be thought of as a production function with increasing social surplus; e.g. the utility to the buyer may be linear in the investment. So why is the standard job order contracting model with equal parts not a solution to the holdup problem? Unless payments and investments in the final period are small, the usual market unraveling logic applies.

Constant payments may work, however, once an extra period with small investments and payments is added. In Table 1, we also show how many periods an investment and payment schedule with constant investments (in all periods except the final period) would take in an ε-SPE. In order to compute the values in Table 1, we still fix $x_1 = 2\varepsilon$ and $p_1 = \varepsilon$, so in order to reach a total investment of 1 we need that $x_t = p_t = \frac{1-2\varepsilon}{T-1}$ for all $t > 1$. We already know from the construction of payments in Proposition 3 that the maximal admissible payment in period 2 is $f(1) - f(1-2\varepsilon)$, so we find the required T by solving

$$\frac{1-2\varepsilon}{T-1} = f(1) - f(1-2\varepsilon)$$

for T and taking the ceiling thereof. In the ε-SPE with minimal required periods, investments are decreasing (Proposition 3), whereas with constant investments they obviously are not. Thus, projects take much longer to implement.

In Figure 1, we present the investments in each period of an ε-SPE for $f(x) = 2\sqrt{x}$ and different values of ε. The smaller ε, the longer it takes until substantial investments are undertaken, contributing to a longer total duration.

		$\min T$ for ...		
	ε	$2\sqrt{x}$	$1.1 \cdot \min(x, 1)$	$1.1 \cdot \mathbb{1}_{x \geq 1}$
ε-SPE	.01	18	19	2
	.05	7	8	2
	.1	4	5	2
Terminal ε-SPE	.01	393	50	2
	.05	74	30	2
	.1	35	26	2
Constant investments ε-SPE	.01	50	46	2
	.05	10	10	2
	.1	5	5	2

Table 1: Smallest T given ε for various production functions.

5 Discussion

Do we need that *both* players willing to forego peanuts in *each* move, as proposed by the ε-SPE? As we have already shown, the stricter notion of terminal ε-SPE suffices, in which only *one* player behaves ε-optimal only in the *terminal* move.

Moreover, when considering ε-SPE, we should consider some notion of doubt about the type of the opponent. When the buyer makes the first payment in the ferryman example, he faces considerable risk: If his beliefs about the seller being an ε-type are wrong, the seller may not make the final investment and the buyer makes a huge loss. Thus, we introduce the notion of *stability* in case no player obtains a substantial loss in utility if his opponent subsequently deviates to some (strictly) better response.

Definition 4. *A strategy profile σ^* is ε-**stable to better responses** if it holds that $u_i(\sigma_i^*, \sigma_{-i}) \geq u_i(\sigma^*) - \varepsilon$ whenever $u_j\left(\sigma_j, \sigma_{-j}^*\right) > u_j(\sigma^*)$ for $j \neq i$.*

If a player correctly anticipates that the opponent will deviate, she might also want to adjust her own action. Thus, the concept of *robustness* is used when a player gets no substantial gain by knowing their opponents will deviate to a (strictly) better response.

Definition 5. *A strategy profile σ^* is ε-**robust to better responses** if it holds that $u_i(\sigma_i^*, \sigma_{-i}) \geq \max_{\sigma_i \in \Delta A_i} u_i(\sigma_i, \sigma_{-i}) - \varepsilon$ if $u_j\left(\sigma_j, \sigma_{-j}^*\right) > u_j(\sigma^*)$ for $j \neq i$.*

Given these definitions, we are now ready to state this section's main result: ε-SPE may be neither stable nor robust, but our terminal ε-SPE is both stable and robust.

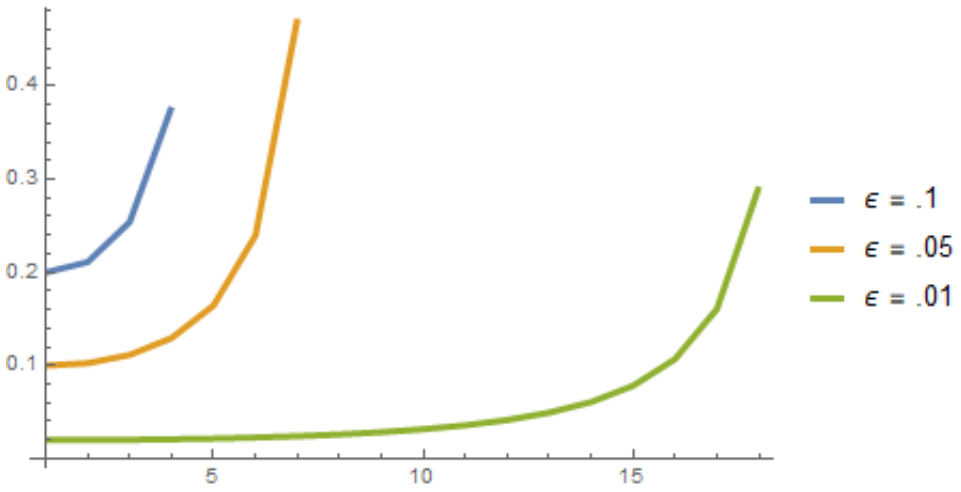

Figure 1: Investments in ε-SPE for $f(x) = 2\sqrt{x}$, given ε, over time. Note that period 1 is the last period.

Proposition 5. *(i) ε-SPE may be neither robust nor stable.*

(ii) The strategy profiles employed in the terminal ε-SPE in Proposition 2(ii) and Proposition 4 are ε-stable to better responses and ε-robust to better responses.

Proof. (i) E.g., for the ferryman technology with target payoffs $u_b = f(1)-1+\varepsilon$, the buyer has a better response of $p_1' = 0$ instead of $p_1 = \varepsilon$, in which case the seller's continuation payoff after making the final investment $x_1 = 2\varepsilon$ is -2ε. If the seller knew that the buyer would subsequently deviate to paying nothing, he would also invest 0 in the last period for a continuation payoff of 0. Therefore, this violates robustness since

$$-2\varepsilon \leq 0 - \varepsilon = -\varepsilon.$$

Also the seller has a better response instead of making the final investment of $x_1 = 2\varepsilon$, namely deviating to $x_1' = 0$. But then the first payment $p_2 = 1 - 2\varepsilon$ violates stability since

$$f(1-2\varepsilon) - 1 + 2\varepsilon = -1 + 2\varepsilon < f(1) - 1. \tag{8}$$

(ii) We need to show that if a strategy profile σ^* is a terminal ε-SPE in Proposition 2(ii) or Proposition 4, than it is also a) ε-stable and b) ε-robust.

a) Stability vacuously holds in all subgames except the penultimate, in which the seller is worried that the buyer deviates from the equilibrium payment of $p_1 = \varepsilon$ to the better response $p_1' = 0$. In this case, the seller's continuation payoff is $-\varepsilon$, whereas in equilibrium it would be $-\varepsilon + x_1 = 0$. Therefore, stability is satisfied since

$$-\varepsilon \geq 0 - \varepsilon.$$

b) Analogously, robustness holds in all subgames except the penultimate. If the seller knew that the buyer would subsequently deviate, he would deviate to $x_1' = 0$ himself. Therefore also robustness holds since

$$0 \geq 0 - \varepsilon.$$

□

Proposition 5 together with the numerical examples in Section 4 highlight an important trade-off: One may get robustness and stability, but only the expense of more sub-periods.

Finally, we may also think of ε as a temptation to deviate. In Proposition 3, all constraints are always binding, which implies that players would strictly prefer to stop the relationship whenever possible if ε would be 0. This temptation is eliminated in the terminal ε-SPE.

6 Conclusion

The lyrics that make the charm of Chris de Burgh's song identify the problem: when to pay the ferryman. According to our model, the recommendation in the song to only pay at the other side does not solve the problem. The ferryman would not be willing to take a passenger that follows this recommendation. In this paper we enable the passage, provided the ferryman is willing to replace the optimal with something that is almost optimal. Our suggestion is to pay the ferryman almost everything when almost at the other side, and to pay the rest upon arrival.

The story of the ferryman is more than a myth. The difficulty to find the right timing is apparent anytime there is trade. Two payments are needed for the passage. When the technology is concave more periods of payment are necessary. However, payments should not be equally distributed unless one is willing to have many periods. The final rounds are characterized by small investments and payments, as there is only little joint interaction in the future to reward and punish current behavior. Investments in earlier rounds are larger as the payoffs generated in the future can be used to incentivize to continue. An important ingredient is that players do not deviate from the plan unless they get substantially more. Terminal ε-SPE only uses this variation in the last round, at the expense of increasing the number of periods needed, possibly dramatically. In fact, this paper uncovers that the holdup problem is not robust. It only arises because we impose preferences where each player deviates whenever she can get a strictly higher payoff, regardless of how small this improvement is.

Admittedly, it is not clear that investments can always be split up arbitrarily and at no cost. We present predictions that involve the fewest possible investments, implicitly reflecting costs in splitting the total investment. Our paper can be used to understand which investments can be made when there are only limited possibilities for splitting the total investment. Similarly, a cost of splitting the investment can easily be incorporated explicitly.

We think that experiments are needed to understand how the possibility to split investment and payment will be used to mitigate holdup and to enable cooperation. In reality, uncertainty about preferences of the opponent also often plays a role. Bargaining under uncertainty has been considered by Kartal (2016). It would be interesting to investigate the holdup problem under incomplete information.

References

Admati, A. R. and Perry, M. (1991). Joint projects without commitment. *The Review of Economic Studies*, 58(2):259–276.

Barlo, M. and Dalkiran, N. A. (2009). Epsilon-nash implementation. *Economics Letters*, 102(1):36–38.

Baye, M. R. and Morgan, J. (2004). Price dispersion in the lab and on the internet: Theory and evidence. *RAND Journal of Economics*, pages 449–466.

Bergemann, D. and Schlag, K. (2011). Robust monopoly pricing. *Journal of Economic Theory*, 146(6):2527–2543.

Dixit, A. K. and Nalebuff, B. J. (1993). *Thinking strategically: The competitive edge in business, politics, and everyday life*. WW Norton & Company.

Dulleck, U., Kerschbamer, R., and Sutter, M. (2011). The economics of credence goods: An experiment on the role of liability, verifiability, reputation, and competition. *The American Economic Review*, 101(2):526–555.

Kartal, M. (2016). Honest equilibria in reputation games: The role of time preferences. *Unpublished paper*.

Lockwood, B. and Thomas, J. P. (2002). Gradualism and irreversibility. *The Review of Economic Studies*, 69(2):339–356.

Mailath, G. J., Postlewaite, A., and Samuelson, L. (2005). Contemporaneous perfect epsilon-equilibria. *Games and Economic Behavior*, 53(1):126–140.

Milgrom, P. (2010). Simplified mechanisms with an application to sponsored-search auctions. *Games and Economic Behavior*, 70(1):62–70.

Nardo, D. (2002). *The Greenhaven Encyclopedias Of - Greek and Roman Mythology*. Greenhaven Press.

Palfrey, T. R. and Prisbrey, J. E. (1996). Altuism, reputation and noise in linear public goods experiments. *Journal of Public Economics*, 61(3):409–427.

Pitchford, R. and Snyder, C. M. (2004). A solution to the hold-up problem involving gradual investment. *Journal of Economic Theory*, 114(1):88–103.

Radner, R. et al. (1980). Collusive behavior in noncooperative epsilon-equilibria of oligopolies with long but finite lives. *Journal of Economic Theory*, 22(2):136–154.

Shapiro, C. (1982). Consumer information, product quality, and seller reputation. *The Bell Journal of Economics*, pages 20–35.

Appendix

Proof of Proposition 4. (i) We first show that for $T' = T$, the following sequence of investments and payments constitute a terminal ε-SPE with payoffs $u_b = w$: $x_1 = p_1 = \varepsilon$, $x_2 = p_2 = f(1) - f(1-\varepsilon) - \varepsilon$ and

$$p_t = x_t = f\left(1 - \sum_{i=1}^{t-2} x_i\right) - f\left(1 - \sum_{i=1}^{t-1} x_i\right),$$

for $2 < t \leq T-1$, $x_T = 1 - \sum_{i=1}^{T-1} x_i$ and

$$(p_T, p_{T+1}) = \begin{cases} \left(f\left(1 - \sum_{i=1}^{T-2} x_i\right) - f\left(1 - \sum_{i=1}^{T-1} x_i\right), f(1) - \sum_{t=0}^{T} p_t - w\right) \\ \qquad \text{if } f(1) - \sum_{t=0}^{T-1} p_t - w \geq f\left(1 - \sum_{i=1}^{T-2} x_i\right) - f\left(1 - \sum_{i=1}^{T-1} x_i\right) \\ \left(f(1) - \sum_{t=0}^{T-1} p_t - w, 0\right) \text{ otherwise} \end{cases}$$

and immediate termination whenever anyone deviates.

The remainder of the proof is analogous to the proof of Proposition 3 and therefore omitted.

(ii) The remainder of the proof is analogous to the proof of Proposition 3 and therefore omitted.

□