

Martin, Simon; Schlag, Karl

Conference Paper

Finite Horizon Holdup and How to Cross the River

Beiträge zur Jahrestagung des Vereins für Socialpolitik 2017: Alternative Geld- und Finanzarchitekturen - Session: Economic Theory, No. C03-V2

Provided in Cooperation with:

Verein für Socialpolitik / German Economic Association

Suggested Citation: Martin, Simon; Schlag, Karl (2017) : Finite Horizon Holdup and How to Cross the River, Beiträge zur Jahrestagung des Vereins für Socialpolitik 2017: Alternative Geld- und Finanzarchitekturen - Session: Economic Theory, No. C03-V2, ZBW - Deutsche Zentralbibliothek für Wirtschaftswissenschaften, Leibniz-Informationszentrum Wirtschaft, Kiel, Hamburg

This Version is available at:

<https://hdl.handle.net/10419/168136>

Standard-Nutzungsbedingungen:

Die Dokumente auf EconStor dürfen zu eigenen wissenschaftlichen Zwecken und zum Privatgebrauch gespeichert und kopiert werden.

Sie dürfen die Dokumente nicht für öffentliche oder kommerzielle Zwecke vervielfältigen, öffentlich ausstellen, öffentlich zugänglich machen, vertreiben oder anderweitig nutzen.

Sofern die Verfasser die Dokumente unter Open-Content-Lizenzen (insbesondere CC-Lizenzen) zur Verfügung gestellt haben sollten, gelten abweichend von diesen Nutzungsbedingungen die in der dort genannten Lizenz gewährten Nutzungsrechte.

Terms of use:

Documents in EconStor may be saved and copied for your personal and scholarly purposes.

You are not to copy documents for public or commercial purposes, to exhibit the documents publicly, to make them publicly available on the internet, or to distribute or otherwise use the documents in public.

If the documents have been made available under an Open Content Licence (especially Creative Commons Licences), you may exercise further usage rights as specified in the indicated licence.

Finite Horizon Holdup and How to Cross the River

Simon Martin*

Karl H. Schlag**

February 14, 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 the invest technology is concave and there are sufficiently many rounds of investment. This shows that the holdup problem that emerges from backwards induction in a finite horizon is not robust.

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.

A broad variety of economic problems, termed holdup problems, can be interpreted as a very similar situation. One agent, say, the seller (she), has to make an upfront investment in order to produce or deliver something valuable to another agent, the buyer (he). Who should move first? If the irrevocable investment is made first, the seller has no incentive to pay anymore - he is already at the side of the Styx, so why pay? Alternatively, the payment could be made first. But then why should the seller still make her investment? Indeed, the unique subgame-perfect equilibrium of games along these lines is that no trade takes place.

The existing solutions to these problems 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. Third, if the game's duration is uncertain, both sides may have an incentive to continue investing and paying until the game terminates randomly (Pitchford and Snyder, 2004).

However, institutions may too costly or may not always be available, e.g., the investment is not observable by the court. Some economic interactions are not continuously repeated with large probability, yet trade emerges. Finally, most trade is organized with known transaction and delivery dates.

In this paper, we propose a novel solution to finite horizon holdup problems by splitting up the total investment into smaller chunks and using the solution concept of ε -subgame perfect equilibria (Mailath et al., 2005) instead of the standard subgame perfect equilibrium. A subgame perfect equilibrium (SPE) identifies situations in which each player can

obtain strictly more by deviating. The concept of an ε -SPE makes a minor adjustment. There is a constant ε , which is typically small, such that each player can obtain by deviating at most ε more than her equilibrium payoff. Intuitively, why bother about arbitrarily small gains? The ε threshold can capture deliberation costs, embarrassment costs and unmodeled uncertainty about play of the other.

The finite horizon holdup problem finds its origin 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 in many small ones in order to make that last investment small and postulate 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 continues to invest and the buyer continues to pay. 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 adapting ε -SPE instead as solution concept the problem is solved for any $\varepsilon > 0$. By making both the final investment and payment small enough, no side has an incentive to deviate, thus also allowing previous investments to take place. We also 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 relationship between a seller and a buyer, which may however 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 exists, 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 $f : \mathbb{R}_+ \rightarrow \mathbb{R}_+$ 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 $x = 0$, 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's incentive compatibility constraints for continuation have to hold in each period t . Our model specification accommodates both discounting at a factor $\delta \in (0, 1]$ and obtaining an outside payoff at any time. For simplicity, we first present a set of results for the limiting case in which $\delta = 1$. We show that an ε -SPE exists in which the buyer obtains the entire social surplus plus an extra ε if ε is sufficiently large. We also present the sequence of investments and payments that implements this ε -SPE. In particular, we show that if the production function f is concave, then the investments and payments are decreasing (except for a possibly larger first and final investment).

Besides the ferryman example, there are a wide range of possible applications of this principle. The key for a house or a car is handed over only once it was paid for. In job order contracting, the entire contract is split up into many subcontracts, and each is paid for upon completion. While these contracts frequently involve equal sized parts, our analysis shows that this need not be the case. Gaining more importance recently due to the rise of the Internet and online retailing, payments may be made on delivery. Shadow markets definitely lack the legal enforceability aspect, so may well be characterized by our framework. More broadly, all transactions involving either an up-front payment (tailor, car mechanic, car purchase, car rental, ...), security deposit (long-term apartment rental, booking hotel, ...) or registration fee (conference, job application - sunk costs of applying, ...) can be thought of along the lines of our idea.

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. Note that 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). Note that our analysis shows that typically only the last investments and payments need to be small.

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. As in formal models underlying the folk theorem, in their paper the relationship is assumed in each round to continue with a sufficiently large probability. There is no fixed date for delivery of a final product nor room for infrequent relationships with various different market partners. Although the mechanisms for ensuring investment are very different, our paper shares with theirs the decreasing investment. Note also that the efficient investment can

be supported in our model with a known termination date, while in Pitchford and Snyder (2004) it is only attained in the limit as the continuation probability tends to 1.

We lay out our main model in section 2. The special case without discounting, the main results and a few examples are described in section 3. Finally, we discuss our results and conclude in section 4.

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 .

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

Two types of production functions are of special interest. (i) The willingness of the buyer to pay for each additional investment is larger than the cost for the seller for this investment. Formally, $f'(x) \geq 1$ for all $x < 1$. (ii) The maximal surplus exceeds the value of any smaller investment. Formally, $f(1) - 1 \geq \lim_{x \rightarrow 1^-} f(x)$. We refer to this one as the ferryman technology as the passage has no value until the passenger reaches the other side.

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.

Compensating the seller for his previous effort and to keep him motivated for future effort, the buyer may choose to pay a certain, possibly negative, amount p_t in each period. We denote the sequence of investments and payments as $(x_t)_{t=1}^T$ and $(p_t)_{t=1}^T$.

Both the seller and the buyer discount payoffs between periods using a discount factor δ where $\delta \in (0, 1]$. Their outside options, which they may obtain anytime if they break up the relationship while investments are still ongoing or at the end of the game otherwise, are given by $w_s \geq 0$, $w_b \geq 0$ for the seller and the buyer respectively.

The timing of the game is as follows. In each period t with $1 \leq t \leq T$, the seller moves first and decides between investing x_t or breaking up the relationship and thereby obtaining the outside option w_s immediately. The cost of the investment is beared by the seller at the beginning of the period. The benefits of the investment for the buyer are realized at the end of the period. Future investments and payments are discounted at a

rate δ by both players.

Reflecting the idea of holdup problems, the buyer observes whether the seller invested or not. She then faces choice of paying p_t or breaking up the relationship. In that case, she obtains both the outside option w_b as well as everything that was produced up to that point, i.e., $f(\sum_{i=1}^t x_i)$.

The game ends after T periods if no player chose the outside option before.

Assuming the game continues until period T , the seller's equilibrium payoffs at the beginning of period t are given by

$$u_s^t(x_1, x_2, \dots, x_T, p_1, p_2, \dots, p_T) = \delta \sum_{i=t}^T \delta^{i-t} p_i - \sum_{i=t}^T \delta^{i-t} x_i + \delta^{T-t+1} w_s \quad (1)$$

and for the buyer

$$u_b^t(x_1, x_2, \dots, x_T, p_1, p_2, \dots, p_T) = -\delta \sum_{i=t}^T \delta^{i-t} p_i + \delta^{T-t+1} (f(1) + w_b). \quad (2)$$

Note that the unique subgame perfect equilibrium is that no investments take place at all. After the last investment was made, the buyer doesn't have an incentive to make the final payment p_T if $p_T > 0$, since the project is already finished. Anticipating no final payment, the seller of doesn't want to make the final investment x_T . 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

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 . Denoting as A_i the set of actions for player i and as σ_{-i}^* the equilibrium strategy profile of the other player, a strategy profile σ^* is a ε -Nash equilibrium if $u_i(a_i, \sigma_{-i}^*) \leq u_i(\sigma^*) + \varepsilon$, $\forall a_i \in A_i$, $\forall i$. A strategy profile σ^* is a ε -subgame perfect equilibrium if it is a ε -Nash equilibrium in every subgame.

Possible reasons why one might not want to deviate when the gains are small or negligible include: i) switching costs (as compared to the status quo), ii) deliberation costs, iii) costs from being embarrassed and iv) 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 are best for one of the two parties.

In order to check whether a given sequence of investments and payments constitute an ε -SPE, we need to make sure that the incentive compatibility constraints hold for both the seller and the buyer in every period.

We define $\sum_{k=i}^j h_k = 0$ for $i > j$.

In order for an investment x_t to be incentive compatible (IC) for the seller in period t , it has to hold that:

$$\delta p_t - x_t - \sum_{i=t+1}^T (\delta^{i-t} x_i - \delta^{i-t+1} p_i) + \delta^{T-t+1} w_s \geq -\varepsilon + w_s \quad (3)$$

which we can rewrite in terms of the minimum price p_t required s.t. the seller IC constraint is satisfied:

$$p_t \geq \hat{p}_t := c_s(t) - \sum_{i=t+1}^T \delta^{i-t} p_i$$

where $c_s(t)$ is independent of prices and defined as

$$c_s(t) := \frac{1}{\delta} (w_s - \varepsilon) - w_s \delta^{T-t} + \sum_{i=t}^T \delta^{i-t-1} x_i.$$

Since \hat{p}_t depends only on future investments and payments we can compute \hat{p}_T :

$$\hat{p}_T = c_s(T) = \frac{x_T - \varepsilon}{\delta} + \left(\frac{1}{\delta} - 1 \right) w_s.$$

Conversely, IC for the buyer requires that

$$-p_t - \sum_{i=t+1}^T \delta^{i-t} p_i + \delta^{T-t} (f(1) + w_b) \geq -\varepsilon + w_b + f\left(\sum_{i=1}^t x_i\right) \quad (4)$$

or alternatively

$$\begin{aligned} p_t &\leq \bar{p}_t := \varepsilon - \sum_{i=t+1}^T \delta^{i-t} p_i + \delta^{T-t} (f(1) + w_b) - w_b - f\left(\sum_{i=1}^t x_i\right) \\ &= c_b(t) - \sum_{i=t+1}^T \delta^{i-t} p_i \end{aligned}$$

where $c_b(t)$ is independent of prices and defined as

$$c_b(t) := \varepsilon + \delta^{T-t} f(1) - (1 - \delta^{T-t}) w_b - f\left(\sum_{i=1}^t x_i\right)$$

and therefore

$$\bar{p}_T = c_b(T) = \varepsilon.$$

We are now ready to state our first result:

Proposition 1. *An ε -SPE in which total investments equal to 1 exists if and only if $c_s(t) \leq c_b(t) \forall t$.*

Proof: A sequence of payments $\{p_i\}_{i=1}^T$ constitute an ε -SPE in which total investments equal to 1 if $\hat{p}_t \leq \bar{p}_t \forall t$. By definition of \hat{p}_t and \bar{p}_t , this holds if and only if $c_s(t) \leq c_b(t)$, irrespective of the sequence of payments. QED

Proposition 1 shows that for equilibrium existence, we need not be concerned about possible prices. Equilibrium existence is entirely determined by primitives of the model and the sequence of payments.

Note that 1 does not yet qualify possible prices; in particular, it doesn't rule out the possibility for negative prices. Nevertheless it will be useful in the following.

3 No discounting

Consider the extreme case in which $\delta = 1$. As should be clear from Proposition 1, a whole plethora of ε -SPE may exist. We thus focus on two extreme types of ε -SPE: one in which the buyer's most preferred payoffs are realized, and another in which the seller's most preferred payoffs are realized. Taking the production function $f(x)$ and the number of periods T as fixed, we first describe which payoffs the buyer (seller) receives in his most preferred case. We proceed by constructing an ε -SPE such that these payoffs emerge, and provide conditions on ε such that the sequence of investments and payments constitute an ε -SPE.

In this section, we assume $\delta = 1$ throughout. Note that $\delta = 1$ implies that the outside options w_b and w_s can be omitted in IC constraints, since they exactly cancel out.

3.1 Buyer's favorite

Given a production function f and the number of periods T , what is the highest payoff the buyer can possibly get in an ε -SPE? Mathematically speaking, we want to maximize the buyer's ex-ante payoff as given by equation (2) subject to the incentive compatibility constraints (equations 3 and 4) in each period t , by choosing $(x_t)_{t=1}^T$ and $(p_t)_{t=1}^T$.

In Proposition 2 we first characterize an upper and a lower bound for the buyer's and the seller's payoff, respectively. We show that given ε is sufficiently large, we can characterize a ε -SPE that attains these payoffs. As we show, in this ε -SPE the IC for both the seller and the buyer are binding in all periods, except for possibly the first period

with actual investments. If ε is sufficiently large, there may be several periods without any positive investments or payments before the project actually starts.

Proposition 2. *Let $\delta = 1$.*

- (a) *In any ε -SPE we have that $u_b \leq f(1) - 1 + \varepsilon + w_b =: \bar{u}_b$ and $u_s \geq -\varepsilon + w_s =: \underline{u}_s$.*
- (b) *Given T , there exists ε_T^b such that for any $\varepsilon > \varepsilon_T^b$ an ε -SPE with payoffs $u_b = \bar{u}_b$ and $u_s = \underline{u}_s$ exists, and for any $\varepsilon < \varepsilon_T^b$ such an ε -SPE does not exist. For any $\varepsilon > \varepsilon_T^b$, there exists a $T' \leq T$ such that the following investments and payments constitute an ε -SPE with payoffs $u_b = \bar{u}_b$ and $u_s = \underline{u}_s$:*

$x_T = 2\varepsilon$, $p_T = \varepsilon$ and

$$p_t = x_t = \begin{cases} f\left(1 - \sum_{i=t+2}^T x_i\right) - f\left(1 - \sum_{i=t+1}^T x_i\right) & \text{if } 2 + T - T' \leq t \leq T - 1 \\ 1 - \sum_{i=2}^T x_i & \text{if } t = T - T' + 1 \\ 0 & \text{if } t \leq T - T' \end{cases} . \quad (5)$$

Before proceeding with the proof, we introduce a useful lemma that later is used to define ε_T^b and describes an important property of ε_T^b .

Lemma 1. *Define $\bar{\varepsilon}$ as*

$$\bar{\varepsilon} = \inf \left\{ \varepsilon \text{ s.t. } f(1) - 1 \geq -2\varepsilon + f\left(1 - \sum_{i=2}^T x_i\right) \right\}, \quad (6)$$

where x_i is as defined by (5) once we set $T' = T$. Then $\bar{\varepsilon}$ is decreasing in T .

Proof. Showing that $\bar{\varepsilon}$ decreases in T is identical to showing that $2\varepsilon - f\left(1 - \sum_{i=2}^T x_i\right)$ is increasing in ε .

For $2 \leq k \leq T$ it holds that

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

We proceed by induction. Assume that we know that $2\varepsilon - f\left(1 - \sum_{i=k+1}^T x_i\right)$ is increasing in ε , then so is $\sum_{i=k}^T x_i$.

Thus an increase in ε decreases $\left(1 - \sum_{i=k+1}^T x_i\right)$ and therefore increases $-f\left(1 - \sum_{i=k+1}^T x_i\right)$. Consequently $2\varepsilon - f\left(1 - \sum_{i=k}^T x_i\right)$ is increasing in ε .

The start of the induction at $k = T$ is given by $2\varepsilon + f(1) - f(1 - 0)$. This is increasing in ε , as initially assumed, and thus the proof is completed. \square

As we will show in the following, $\bar{\varepsilon}$ as defined by (6) is the ε_T^b to be used in Proposition 2. For continuous production functions f , there is always some $\bar{\varepsilon}$ that solves (6) with equality, but for discontinuous functions we need to allow for jumps in ε and therefore (6) may never be solved with equality.

Proof of Proposition 2.

(a) Since $\delta = 1$, period t IC constraint for the seller reduces to

$$p_t - x_t - \sum_{i=t+1}^T (x_i - p_i) \geq -\varepsilon. \quad (7)$$

Buyer optimality demands that this holds with equality in every period, since higher payments decrease buyer's payoff and he strictly prefers not to pay more than seller IC requires. Moreover, given that $\delta = 1$, payments above the IC constraint relax earlier constraints by exactly the same amount, which the buyer can always compensate for by an equivalent payment in an earlier period.

At $t = 1$, making (7) bind is equivalent to

$$\sum_{i=1}^T p_i = \sum_{i=1}^T x_i - \varepsilon. \quad (8)$$

Thus we have that

$$\begin{aligned} u_b &= - \sum_{i=1}^T p_i + f\left(\sum_{i=1}^T x_i\right) + w_b \\ &= - \sum_{i=1}^T x_i + \varepsilon + f\left(\sum_{i=1}^T x_i\right) + w_b. \end{aligned}$$

Using the fact that $f(x) - x$ is maximized at $x = 1$, we have that in any ε -SPE

$$u_b \leq f(1) - 1 + \varepsilon + w_b = \bar{u}_b$$

and by (8) that

$$u_s = \sum_{i=1}^T (p_i - x_i) + w_s \geq -\varepsilon + w_s = \underline{u}_s.$$

(b) We first show that our sequence of investments and payments constitute a ε -SPE if $\varepsilon > \varepsilon_T^b$ where ε_T^b is defined by (6), i.e., that both the seller's and the buyer's IC constraints are satisfied in all periods.

Consider first the case where ε is such that $\varepsilon_{T-1}^b > \varepsilon > \varepsilon_T^b$, and in that case let $T' = T$.

For $\delta = 1$, seller's period t IC constraint reduces to

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

For rewriting buyer's IC constraint we can use the fact $\sum_{i=1}^T x_i = 1$ or equivalently $\sum_{i=1}^t x_i = 1 - \sum_{i=t+1}^T x_i$ and therefore

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

and hence

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

Thus in period T where $x_T = 2\varepsilon$ and $p_T = \varepsilon$ both IC constraints are binding.

Moreover, by construction of p_t and x_t , in all periods $2 \leq t \leq T-1$ both IC constraints are binding.

Finally, first period investment and payment $p_1 = x_1 = 1 - \sum_{i=2}^T x_i$ imply that seller's IC constraints is also binding in period 1, and therefore in all periods.

We thus are left with showing that these payments are also IC for the buyer in period 1:

$$-p_1 - \sum_{i=2}^T p_i + f(1) \geq -\varepsilon + f\left(1 - \sum_{i=2}^T x_i\right)$$

which can be expanded to

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

and therefore

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

Since also the investments depend on ε , the minimal ε (if it exists) that satisfies this is implicitly defined via

$$f(1) - 1 \geq -2\varepsilon_T^b + f\left(1 - \sum_{i=2}^T x_i\right) \quad (9)$$

because the right-hand side of (9) is decreasing in ε as shown in the proof of Lemma 1. If no ε solves (9) with equality, the IC constraint holds for $\varepsilon > \varepsilon_T^b$.

If $\varepsilon_{T-1}^b > \varepsilon > \varepsilon_T^b$ then the IC constraints are satisfied in all periods both for the seller and the buyer and thus an ε -SPE exists.

Finally, payoffs in this ε -SPE are given by

$$\begin{aligned} u_b &= -\sum_{i=1}^T p_i + f(1) + w_b \\ &= -x_1 - \sum_{i=2}^T p_i + f(1) + w_b \\ &= -(1 - \sum_{i=2}^T x_i) - \sum_{i=2}^T p_i + f(1) + w_b \\ &= -1 + \sum_{i=2}^T p_i + \varepsilon - \sum_{i=2}^T p_i + f(1) + w_b \\ &= f(1) - 1 + \varepsilon + w_b = \overline{u_b} \end{aligned}$$

and

$$u_s = \sum_{i=1}^T (p_i - x_i) + w_s = -\varepsilon + w_s = \underline{u_s}.$$

Consider now the case where $\varepsilon \geq \varepsilon_{T-1}^b$. In that case, let T' be the largest integer such that $\varepsilon \geq \varepsilon_{T'}^b$. This implies that $0 < T' < T$ as ε_T^b decreases in T by Lemma 1. Therefore, the subgame starting from period $T - T' + 1$ is identical to the previous case and therefore investments and payments are IC for both the seller and the buyer. Moreover, $p_t = x_t = 0$ in all periods $t \leq T - T'$ are IC for both the seller and the buyer given that future investments and payments are unaffected. Finally, payoffs are unaffected too, and therefore also for $\varepsilon \geq \varepsilon_{T-1}^b$ we always have a ε -SPE with payoffs $u_b = \overline{u_b}$ and $u_s = \underline{u_s}$.

We now show that $\varepsilon < \varepsilon_T^b$ an ε -SPE with $u_b = \overline{u_b}$ and $u_s = \underline{u_s}$ does not exist.

Note that the investments and payments are such that both the seller and the buyer IC constraints are binding in each period if $\varepsilon = \varepsilon_T^b$ (or in every period except the first

if $\varepsilon > \varepsilon_T^b$, but ε_T^b is still the minimal ε that solves (6) if such a solution exists). An investment and payment schedule such that the seller's IC constraint is not binding in every period t , i.e., for some t it holds that

$$p_t > \sum_{i=t+1}^T (x_i - p_i) + x_t - \varepsilon,$$

cannot deliver payoffs $u_b = \bar{u}_b$ with a smaller ε , because doing so reduces u_b from (2). If the reduction in u_b is compensated by an equivalent reduction in payments in some other period t' , payoffs and investments are simply shifted across periods, because for $\delta = 1$ all future payments enter with a coefficient of 1 into the IC constraints for the seller and buyer in equations (3) and (4) respectively.

Similarly, an investment and payment schedule such that the buyer's IC constraint is not binding in any period cannot deliver payoffs $u_b = \bar{u}_b$ with a smaller ε , because that would tighten the seller's IC constraint in previous periods and thus leave more investment for the first period, increasing the necessary ε to satisfy the period 1 buyer IC.

Thus there cannot be another sequence of investments and payments that attains \bar{u}_b with a lower ε , and consequently $\varepsilon < \varepsilon_T^b$ implies that an ε -SPE with $u_b = \bar{u}_b$ and $u_s = \underline{u}_s$ does not exist.

□

Investments and payments in Proposition 2 are obtained in a backwards-induction manner. In the final period T , the IC payment is bounded from above by ε . Given that, the IC investment is bounded from above by 2ε , and so on. Thus, given $p_T = \varepsilon$ and $x_T = 2\varepsilon$, the seller already foregoes the maximum ε relative to his optimal payoff. Consequently, payments have to be larger or equal to investments in all previous periods. As we move further towards the beginning, terms in the distant in the future cancel out and the neat representation from expression (8) remains. Anticipating future investments and payments on the equilibrium path, the seller invests the value the project will have after two periods from t , minus the value the project will have after one period from t . This makes both seller and buyer IC bind in all periods, except possibly the initial periods.

We now present some conditions that allow us to characterize the sequence of investments, and then we conclude the section with a few examples.

Note that if f is concave then $f'(x) \geq 1$ for all $x \in [0, 1)$. The reason is as follows. Given that $f(x) - x$ is maximized for $x = 1$ on $x \leq 1$, it needs to be the case that $f'(1) \geq 1$. By concavity we have that $f'(x)$ is decreasing and hence $f'(x) \geq f'(1) \geq 1$ for all $x \in [0, 1)$.

Proposition 3. *Let $\delta = 1$ and $f'(x) \geq \gamma$ for all $x \in [0, 1)$ and $\gamma > 0$. Then for the sequence of investments as specified in Proposition 2(b) it holds that $x_t \geq \gamma x_{t+1}$ for $2 + T - T' \leq t \leq T - 1$.*

Proof. $f'(x) \geq 1$ implies that $f(y) - f(x) \geq \gamma(y - x)$. Note that $x_T = 2\varepsilon$ and therefore $1 - \sum_{i=t'+1}^T x_i$ is strictly smaller than 1 for all t' with $2 + T - T' \leq t' \leq T - 1$. Thus for t such that $2 + T - T' \leq t \leq T - 1$ we have that

$$x_t = f\left(1 - \sum_{i=t+2}^T x_i\right) - f\left(1 - \sum_{i=t+1}^T x_i\right) \geq \gamma \left(\left(1 - \sum_{i=t+2}^T x_i\right) - \left(1 - \sum_{i=t+1}^T x_i\right) \right) = \gamma x_{t+1}.$$

□

Proposition 3 shows in particular that if $f'(x) \geq 1$ holds for all $x \in [0, 1)$, then $x_t \geq x_{t+1}$ and therefore investments are gradually decreasing over time, i.e., a form of gradualism. The reason is that as the project continues, refusing to pay for the last progress and consuming what was already built becomes more and more attractive for the buyer. Alternatively speaking, the seller's threat of not continuing to invest becomes less harmful the more the project has progressed already. Therefore, the seller has to gradually invest less and less such that the buyer is still willing to pay for a marginal improvement.

Another interpretation is that, if f is concave, then so is $f(x) - x$, which means that for each additional amount invested, the added value for the buyer is decreasing up to the efficient total investment of $x = 1$, suggesting gradually decreasing payments.

Moreover, if $\gamma > 1$ in Proposition 3, this shows that the investments decrease at least at a geometric rate. In particular they are not constant. Therefore, T need not be large even if ε is small and $\gamma > 1$. Finally, this shows that even parts need not be optimal.

In Proposition 2 we take T as given, and provide a necessary condition for ε -SPE existence in terms of ε . In Proposition 4 we show also the opposite holds: Given any ε , an ε -SPE with buyer-optimal payoffs exists if T is sufficiently large.

Proposition 4. *Let $\delta = 1$ and $f'(x) \geq 1$ for all $x \in [0, 1)$. Then for any $\varepsilon > 0$, there exists some T_ε such that for all $\hat{T} \geq T_\varepsilon$ an ε -SPE with $u_b = \overline{u}_b$ and $u_s = \underline{u}_s$ exists.*

Proof. We proof the statement by showing that the last term of (6) converges to 0 as T goes to infinity, which is identical to showing that $\sum_{i=2}^T x_i$ goes to 1 as T becomes sufficiently large.

By Proposition 3 we have that x_t is decreasing in t . For any T it holds that $x_{T-1} > 0$. Therefore, for any $\hat{T} \geq T_\varepsilon$ it holds that

$$\sum_{i=2}^{\hat{T}} x_i \geq \sum_{i=2}^{T_\varepsilon} x_i \geq (T_\varepsilon - 3)x_{T_\varepsilon-1} \geq 1$$

once T_ε satisfies $T_\varepsilon \geq 1/x_{T_\varepsilon-1} + 3$. □

Next we establish for the ferryman technology that only two periods are needed.

Proposition 5. *Assume that $f(1) - 1 \geq \lim_{x \rightarrow 1^-} f(x)$. For any $\varepsilon > 0$ we have that $T = 2$ and $\varepsilon_2^b = \varepsilon$ satisfies the statement of Proposition 2(b).*

Proof. The proof follows immediately from noticing that (6) is equivalent to $f(1) - 1 \geq f(1 - 2\varepsilon) - 2\varepsilon$. □

We now present a few examples. Table 1 shows ε_T^b for different $f(x)$, which directly pins down the minimum T required for buyer's maximum payoff (Table 2). We then present the sequence of investments and payments for some prominent values of ε at the minimum T (Figure 1).

We discuss four different production functions: a concave function $2\sqrt{x}$, a linear function $1.1 \cdot \max(x, 1)$, a continuous convex function $2 \cdot \max(x^2, 1)$, and the discontinuous convex function $1.1 \cdot \mathbb{1}_{x \geq 1}$, already known as ferryman example from the introduction.

As shown already by Lemma 1, ε_T^b decreases in T (see Table 1), irrespective of the production function chosen. The column with ε_T^b for the ferryman production function is omitted, because even for $T = 2$, any $\varepsilon > 0$ allows to implement $u_b = \overline{u}_b$.

In Table 2, we present the minimal T required to implement $u_b = \overline{u}_b$ and $u_s = \underline{u}_s$ in an ε -SPE, given $\varepsilon = 0.05$ or $\varepsilon = 0.1$. As ε_T^b decreases in T , equivalently the minimal T decreases as ε increases. As pointed out above, the extremely convex ferryman production function requires only $T = 2$ for any $\varepsilon > 0$.

Finally, in Figure 1 we show the sequence of investments that implements $u_b = \overline{u}_b$ for $\varepsilon = 0.1$ and $T = 4$ and $\varepsilon = 0.05$ and $T = 7$ for $f(x) = 2\sqrt{x}$. As shown by Proposition 3, x_t decreases for $2 \leq T \leq T - 1$. However, for $\varepsilon = 0.05$, the first investment is smaller than the second, and vice-versa for $\varepsilon = 0.1$. In both cases, the final payment is the smallest.

T	ε_T^b for $2\sqrt{x}$	ε_T^b for $1.1 \cdot \max(x, 1)$	ε_T^b for $2 \cdot \max(x^2, 1)$
2	0.232051	0.238095	0.109612
3	0.141322	0.151057	0.027964
4	0.097091	0.107735	0.007132
5	0.071553	0.081899	0.001799
6	0.055250	0.064804	0.000451
7	0.044117	0.052703	0.000113
8	0.036134	0.043722	0.000028

Table 1: ε_T^b given T for various production functions.

ε	$\min T$ for $2\sqrt{x}$	$\min T$ for $1.1 \cdot \max(x, 1)$	$\min T$ for $2 \cdot \max(x^2, 1)$	$\min T$ for $1.1 \cdot \mathbb{1}_{x \geq 1}$
0.05	7	8	3	2
0.1	4	5	3	2

Table 2: Smallest T given ε for various production functions to attain buyer's favorite.

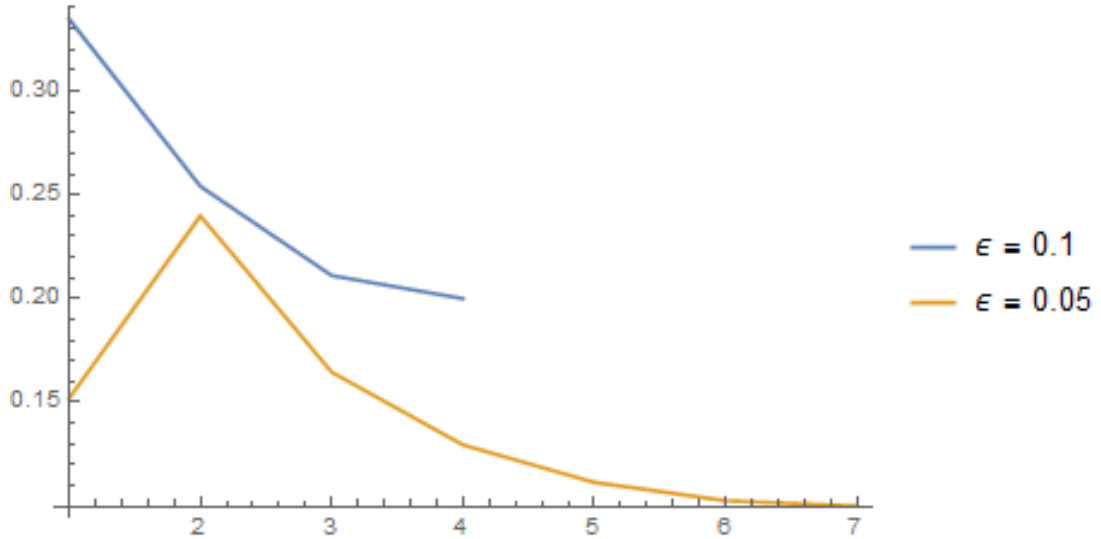


Figure 1: Investments for $f(x) = 2\sqrt{x}$.

4 Conclusion

We find that it is much easier to get paid for relation specific investments than standard theory predicts. One only has to split up the interaction into parts and drop the assumption that all marginally profitable deviations will be undergone. Typically the investments will not be constant but instead will decrease geometrically. Accordingly, even if the minimal gain needed to trigger deviations is small the total number of investments will not be that large. In particular, SPE as a solution concept is not robust if one assumes that marginally profitable deviations are never followed. In a similar vein one can also obtain cooperation in a linear public goods game.

References

- Dixit, A. K. and Nalebuff, B. J. (1993). *Thinking strategically: The competitive edge in business, politics, and everyday life*. WW Norton & Company.
- Mailath, G. J., Postlewaite, A., and Samuelson, L. (2005). Contemporaneous perfect epsilon-equilibria. *Games and Economic Behavior*, 53(1):126–140.
- Nardo, D. (2002). *The Greenhaven Encyclopedias Of - Greek and Roman Mythology*. Greenhaven Press.
- 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.

Appendix