

Gretschko, Vitali; Wambach, Achim

**Conference Paper**

## Common Values and the Coase Conjecture: Inefficiencies in Frictionless Contract (Re-)Negotiation

Beiträge zur Jahrestagung des Vereins für Socialpolitik 2015: Ökonomische Entwicklung - Theorie und Politik - Session: Bargaining, No. E12-V2

**Provided in Cooperation with:**

Verein für Socialpolitik / German Economic Association

*Suggested Citation:* Gretschko, Vitali; Wambach, Achim (2015) : Common Values and the Coase Conjecture: Inefficiencies in Frictionless Contract (Re-)Negotiation, Beiträge zur Jahrestagung des Vereins für Socialpolitik 2015: Ökonomische Entwicklung - Theorie und Politik - Session: Bargaining, No. E12-V2, ZBW - Deutsche Zentralbibliothek für Wirtschaftswissenschaften, Leibniz-Informationszentrum Wirtschaft

This Version is available at:

<https://hdl.handle.net/10419/113064>

**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.*

# COMMON VALUES AND THE COASE CONJECTURE: INEFFICIENCIES IN FRICTIONLESS CONTRACT (RE-)NEGOTIATION

ABSTRACT. *We consider the contracting problem of a principal who faces an agent with private information and cannot commit to not renegotiate a chosen contract. To analyze this problem, we propose an infinite horizon negotiation protocol in which renegotiation is frictionless, executed without delay and there are no restrictions on how many times the contracts can be renegotiated. We provide a general characterization of renegotiation-proof outcomes and show that those outcomes are supported by a Perfect Bayesian Equilibrium of the negotiation game. The general characterization of renegotiation-proof outcomes provides a powerful and simple to use tool for finding such outcomes in specific environments. Thus, we proceed by applying the results to adverse selection environments with private and common values. We show that with private values and common values of the 'Spence' type only fully efficient and separating contracts can be renegotiation proof. However, with common values of the 'Rothschild-Stiglitz' type inefficient and (partial) pooling contracts may constitute renegotiation-proof outcomes.*

*JEL classification:* C73, C78, D82

*Keywords:* Principal-Agent models, renegotiation, Coase-conjecture

## 1. INTRODUCTION

The solution to the screening problem of a principal who is endowed with all the bargaining power and wishes to contract with a privately informed agent is well known. The principal proposes a menu of contracts that is designed such that the agent optimally chooses one of the contracts according to his type. Typically, the chosen contract is inefficient and the choice of the agent reveals information. In this case, both parties can benefit from immediate renegotiation after information is revealed. As such renegotiation will be anticipated by the agent, it may distort his ex-ante incentives to accept any given contract in the first place. Thus, the optimality of the solution to the screening problem crucially depends on the assumption that the chosen contract will not be renegotiated.

For the signaling problem, that is, when the *informed party* makes the contract offers, the effects of allowing renegotiation were analyzed by Beaudry and Poitevin (1993). In this article we extend their seminal analysis to the screening problem, that is, to the case that the *uninformed party* makes the contract offers. We introduce an extension of the one-shot screening game in which an infinite number of rounds of frictionless renegotiation are permitted before contracts are executed. Thus, we model the idea that the principal has no commitment power by assuming that there are no physical costs of renegotiation and that any signed contract can be renegotiated any number of times. A solution to the resulting negotiation game is hard to obtain. As the principal cannot commit not to exploit the information revealed during the game, information revelation by the agent is difficult

to ensure. Moreover, the number of potential negotiation rounds is unbounded. Hence, we cannot use backward induction as in Bester and Strausz (2004) to apply some kind of revelation principle.

Our contribution is threefold. Our first contribution is to find a Perfect Bayesian Equilibrium of the negotiation game. Our second contribution is to introduce a new methodology for analysis of contracting problems without commitment. At this we focus on potential outcomes of the negotiation game. An outcome of the game for a given pair of strategies of the principal and the agent is a contract and a belief of the principal that resulted after the game had ended. The set of renegotiation-proof outcomes is then a set of outcomes that is generated by equilibrium strategies for every possible history of the negotiation game. We state desirable properties of such a set of renegotiation-proof outcomes and thus establish which outcomes can be expected to arise in equilibrium. We then select the renegotiation-proof outcome that maximizes the pay-off of the principal and show that this outcome is supported by a Perfect Bayesian Equilibrium of the negotiation game. This novel approach not only facilitates the equilibrium analysis of the negotiation game but also provides a powerful and easy to use tool to analyze specific instances of the considered problem. Our third contribution is to apply the characterization of renegotiation-proof outcomes to contracting problems with private and common values. We show that while only fully separating and efficient outcomes can arise with private values and common values of 'Spence' type, inefficient and pooling outcomes can be implemented in equilibrium with common values of the 'Rothschild-Stiglitz' type. This surprising result is of great interest as we assume that renegotiation is frictionless. Thus, following a similar reasoning as in the famous Coase conjecture, efficient outcomes were to be expected.

The characterization of the set of renegotiation-proof outcomes is based on two simple properties. First, for every renegotiation-proof outcome there is no other renegotiation-proof outcome that would make the principal better off (*internal consistency*). Second, in any state of the negotiation game it is feasible to reach a renegotiation-proof outcome in a single round of further negotiations (*external consistency*). Both properties reflect sequential rationality of the principal: suppose the negotiation game reaches a renegotiation-proof outcome and the principal proposes new contracts that would make him and the agent better off. External consistency ensures that in a further round of negotiations she will renegotiate those contracts. Internal consistency implies that the resulting outcomes make her not better off than the original outcome. Thus, the proof that such

renegotiation-proof outcomes are supported by a Perfect Bayesian Equilibrium of the negotiation game is a direct consequence of external and internal consistency.<sup>1</sup>

One of the main advantages of a general characterization of renegotiation-proof outcomes is that it provides a powerful tool to analyze specific instances of the general problem. We use the general characterization to derive a simple but useful necessary condition for an outcome to be renegotiation-proof: based on the information revealed in a renegotiation-proof outcome there should not, irrespective of the agents type, exist a single (pooling) outcome that would make the principal and the agent better off. It is easy to see why this is necessary for renegotiation-proofness. If such an outcome would exist, the principal could just offer the corresponding contract and both types of the agent could accept without revealing any additional information. Thus, this pooling outcome would make both parties better off without changing the strategic incentives in the negotiation game. We use this simple insight to prove that with private values only fully separating and efficient outcomes can be renegotiation-proof. This is due to the fact that in any inefficient outcome, the indifference curve of the principal is either steeper or flatter than both indifference curves of the agent. Thus, for every inefficient outcome there exists a single outcome that would make the principal and both of the agents better off without revealing additional information.

For common values we have to distinguish two cases: common values 'Spence' type and common values 'Rothschild-Stiglitz' type. Common values 'Spence' type represent the situation that the ranking of the marginal trade-offs between types of the agent is the same for the agent and the principal. This situation corresponds, for example, to the education model in Spence (1973) where education is both marginally more productive and less costly for the high type. In this case, for any inefficient outcome, the indifference curves of the principal are either both steeper or both flatter than the indifference curves of the agent. Thus, the same logic as with private values applies and there exists a single contract that would make the principal and both of the agents better off. It follows that only fully separating and efficient outcomes can be renegotiation proof.

The situation changes dramatically if common values of the 'Rothschild-Stiglitz' type are considered. Common values of the 'Rothschild-Stiglitz' type represent the situation that the marginal trade-offs between the types of the agent are ranked differently for the agent and the principal. This situation corresponds, for example, to the insurance model by Rothschild and Stiglitz (1976) where insurance is marginally less costly but also marginally less valuable for the low risk type. In this case, the logic used above is not applicable as no single pooling contract exists that makes both of

---

<sup>1</sup>Even though the general reasoning is straightforward, many technical difficulties arise. For example, in our set-up, neither the one-shot deviation nor the revelation principle do hold. Thus, the construction of an equilibrium with renegotiation-proof outcomes is one of the main contributions of this article.

the parties better off. Moreover, it may be the case that all pairs of efficient outcomes would make the principal strictly worse off as compared to the initial situation. Thus, the set of all efficient outcomes fails the external consistency property. We proceed by constructing a set renegotiation-proof outcomes that results in inefficient (partial) pooling outcomes. Interestingly, inefficient outcomes can be sustained in equilibrium even if there exist pairs of efficient outcomes that would make the principal and the agent strictly better off.

**Relation to the Literature.** Most of the previous analysis of renegotiation typically took one of two approaches. Either renegotiation-proof outcomes were characterized axiomatically or renegotiation was limited to finite negotiation protocols. The first approach usually yields clear-cut results in complex settings and thus powerful tools for the analysis of specific problems. However, the lack of foundation as an equilibrium of a non-cooperative game may raise doubts.<sup>2</sup> The second approach allows for equilibrium analysis but still leaves the principal with a considerable amount of commitment power.<sup>3</sup> For example, in our frictionless setting, limiting the renegotiation to  $n$  opportunities would allow the principal to implement the full commitment outcome. She could simply pass on  $n - 1$  opportunities and then propose the optimal contracts. Our approach combines the clarity and power of an axiomatic approach with the equilibrium analysis of a very flexible, infinite horizon negotiation protocol.

Our set-up is closely related to Beaudry and Poitevin (1993) who study the effects of immediate and unlimited renegotiation in a general signaling model. That is, in contrast to our work, only the informed agent can make the contract offers. In this case, separating but inefficient contracts are sustained by the threat that if the agent who signed an inefficient contract proposes a new contract, a switch in beliefs takes place and the uninformed party will assume that the agent is of the undesired type. However, this can only work if it is the informed party who makes the offers. Thus, we complement Beaudry and Poitevin (1993) by extending the analysis to a screening model. That is, we assume that the uninformed party makes all the offers and show that this changes the outcomes dramatically.

More recently, Strulovici (2014) analyzes an infinite horizon negotiation protocol for a set-up with private values where the uninformed party makes all the offers. In contrast to our work, negotiation can break down, that is, renegotiation is not frictionless. However, Strulovici (2014) shows that if the friction disappears, efficient and fully separating contracts arise in any Perfect Bayesian Equilibrium of the negotiation game. Thus, his analysis complements our results for private values.

---

<sup>2</sup>For examples of such an approach see Asheim and Nilssen (1997) and Vartiainen (2013).

<sup>3</sup>For examples of such an approach see Fudenberg and Tirole (1990), Hart and Tirole (1988), Hörner and Samuelson (2011), Skreta (2006), or Skreta (2013).

In a similar set-up, Maestri (2012) uses a the refinement that in any subgame the principal induces the continuation equilibrium that maximizes her payoffs. As in Strulovici (2014), when frictions disappear, only efficient outcomes arise in equilibrium.

With respect to the characterization of the renegotiation-proof outcomes, our article is related to Asheim and Nilssen (1997) and Vartiainen (2013). Asheim and Nilssen (1997) consider a monopolistic insurance market and Vartiainen (2013) an auction without commitment. Both articles use an axiomatic approach to the characterization of renegotiation-proof outcomes that is similar to our approach. That is, both rely on similar properties as our internal and external consistency to characterize renegotiation-proof outcomes. In both cases, this approach proves to be very useful in deriving clear results for otherwise very complex problems. We extend their analysis by providing a foundation of renegotiation-proof contracts as an equilibrium outcome of a very general negotiation game and applying the results to settings not considered by those authors.

Krasa (1999) uses a slightly different, but also axiomatic, approach. He defines an outcome as unimprovable (renegotiation proof) if agents would not want to deviate from it either by changing the allocation or by revealing (additional) information. His characterization of renegotiation-proof outcomes yields that in an insurance monopoly with two types either agents reveal their information fully or not all. This results in either full separation of types or pooling. This is not the case for our model as in the “Rothschild-Stiglitz” case, partial pooling can be supported in equilibrium.

## 2. THE SETUP

A principal (she) and an agent (he) negotiate over a contract. A contract is a tuple  $\omega \in \mathbb{R}_+^2$ . Let  $\theta \in \{L, H\}$  denote the type of the agent. The type is private knowledge to the agent and the principal has a prior characterized by  $\mu_0 = Pr(\theta = H)$ . When a contract  $\omega$  is signed by an agent of type  $\theta$ , the utility of the principal amounts to  $v(\omega, \theta)$ . The utility of the agent is then  $u(\omega, \theta)$ . Both,  $v(\omega, \theta)$  and  $u(\omega, \theta)$  are assumed to be quasiconcave in  $\omega$ . Let  $v_i(\omega, \theta)$  and  $u_i(\omega, \theta)$  denote the partial derivative with respect to the  $i$ -th component of  $\omega$ . The functions  $u(\omega, \theta)$  satisfy the standard single-crossing condition, that is,

$$(1) \quad -\frac{u_2(\omega, L)}{u_1(\omega, L)} > -\frac{u_2(\omega, H)}{u_1(\omega, H)}.$$

The principal prefers smaller values of  $\omega_1$  and larger values of  $\omega_2$ , whereas the opposite is true for the agent, that is,

$$v_1(\omega, \theta) < 0, \quad v_2(\omega, \theta) > 0, \quad u_1(\omega, \theta) > 0, \quad \text{and} \quad u_2(\omega, \theta) < 0.$$

A contract is  $\theta$ -efficient if it is the cheapest contract providing an agent of type  $\theta$  with a given utility level. That is, the iso-utility curve of the principal is tangent to the iso-utility curve of the agent in any such contract. For each  $\theta$ , denote by  $\xi_\theta$  the set of all  $\theta$ -efficient contracts. Sometimes we will refer to  $\xi_\theta$  as the efficient contract curve. To rule out pathological cases, we assume that  $\xi_\theta$  is smooth and upward sloping.

Whenever  $v$  is independent of  $\theta$ , that is,  $v(\omega, \theta) = v(\omega)$ , we will refer to private values. In this case, due to the single-crossing property, the efficient contract curve  $\xi_L$  lies to the left of  $\xi_H$ . Whenever the utility of the principal explicitly depends on the type of the agent, we will refer to common values. For common values we will distinguish two cases: the 'Spence' case and the 'Rothschild-Stiglitz' case.

*'Spence' case:* common values 'Spence' type represent the situation that the ranking of the marginal trade-offs between types is the same for the agent and the principal, i.e.,

$$(2) \quad -\frac{v_2(\omega, L)}{v_1(\omega, L)} < -\frac{v_2(\omega, H)}{v_1(\omega, H)}.$$

This situation corresponds, for example, to the education model in Spence (1973), where education is both marginally more productive and less costly for the  $H$  type. It follows from equation (1) and equation (2) that the efficient contract curves  $\xi_\theta$  do not cross and  $\xi_L$  lies to the left of  $\xi_H$ .

*'Rothschild-Stiglitz' case:* common values 'Rothschild-Stiglitz' type represent situations where the marginal trade-offs are ranked differently for the informed and the uninformed player, i.e.,

$$-\frac{v_2(\omega, L)}{v_1(\omega, L)} > -\frac{v_2(\omega, H)}{v_1(\omega, H)}.$$

This situation corresponds, for example, to the insurance model by Rothschild and Stiglitz (1976), where insurance is marginally less costly but also marginally less valuable for the high type. In this generality, in the 'Rothschild-Stiglitz' case it may happen that the efficient contract curves cross and that the efficient contract curve of the  $L$  type lies to the right of the efficient contract curve of the  $H$  type. Such pathological cases complicate the analysis without adding additional insight. Thus, we assume that in the 'Rothschild-Stiglitz' case the efficient contract curve of the  $L$  type lies to the left of the efficient contract curve of the  $H$  type.<sup>4</sup>

**The negotiation game.** The negotiation game unfolds as follows. In  $t = 0$  the agent observes his type  $\theta$ . In each following round  $t \in \mathbb{N}_+$  the principal offers a menu  $\mathcal{M}_t \subset \mathbb{R}_+^2$  of contracts. The

<sup>4</sup>This greatly simplifies the the exposition of the the strategies of the agent in the proof of Proposition 1 ensuring that only one of the types of the agent uses a mixed strategy. That the efficient contract curve of the  $L$  type lies to the left of the efficient contract curve of the  $H$  type can be ensured by assuming, for example, that the utility functions of the principal and the agent are additively separable, i.e.,  $u(\omega, \theta) = f(\omega_2, \theta) + \lambda_\theta \omega_1$  and  $v(\omega, \theta) = g(\omega_2, \theta) - \omega_1$  with  $\lambda_\theta \in \mathbb{R}_+$  and that if  $-f_1(\omega_2, L)/\lambda_L = g_1(\omega_2, L)$  then  $-f_1(\omega_2, H)/\lambda_H < g_1(\omega_2, H)$ .

number of contracts  $|\mathcal{M}_t|$  is bounded by an arbitrary constant  $K \geq 2$  that is fixed throughout the game.<sup>5</sup> The agent chooses a contract in  $\mathcal{M}_t$  or decides to hold on to the contract he chose in round  $t - 1$ . We denote by  $\omega_t$  the contract that the agent chose in round  $t$  and by  $\omega_0$  the initial contract, that is, the normalized outside option of the agent. The game ends if at time  $t$  the principal does not propose new contracts, that is, if  $\mathcal{M}_t = \emptyset$ . In this case  $\omega_{t-1}$  is executed.

Denote a potential history realized before the principal moves in round  $t$  as

$$h^p(t) = \{(\mathcal{M}_1, \omega_1), (\mathcal{M}_2, \omega_2), \dots, (\mathcal{M}_{t-1}, \omega_{t-1})\}.$$

A potential history realized before the agent moves is

$$h^a(t) = \{(\mathcal{M}_1, \omega_1), (\mathcal{M}_2, \omega_2), \dots, (\mathcal{M}_{t-1}, \omega_{t-1}), \mathcal{M}_t\}.$$

At round 1 there is no relevant history for either of the players, so  $h^{\{p,a\}}(1) = \emptyset$ .<sup>6</sup> Denote by  $h_k(t)$  the restriction of  $h(t)$  to the first  $k$  rounds. Let  $\mathcal{H}^t$  be the set of all histories in round  $t$  and  $\mathcal{H}$  be the set of all terminal histories. That is,  $h(t) \in \mathcal{H}$  if  $\mathcal{M}_t = \emptyset$ . Moreover, all infinite histories are terminal histories. That is,  $h(\infty)$  is an infinite history if there exist a series of histories such that  $h(\infty) = \lim_{t \rightarrow \infty} h(t)$  and  $\mathcal{M}_t = \emptyset$  is not in  $h(t)$  for all  $t$ . For every  $h(t) \in \mathcal{H}$ , we define the payoff of the agent and the principal as  $u(h(t), \theta) = u(\omega_{t-1}, \theta)$  and  $v(h(t), \theta) = v(\omega_{t-1}, \theta)$  if  $t < \infty$  and as  $u(h(\infty), \theta) = v(h(\infty), \theta) = -\infty$ .

Surely, the best way to model renegotiation is not obvious. With our proposal we hope to capture the effects of unlimited, immediate and frictionless renegotiation where the uninformed party makes all the offers. Thus, we complement the work of Beaudry and Poitevin (1993) who model a similar renegotiation game for the case that the informed party makes all the offers. The main elements of the negotiation game are: first, the uninformed party makes both the initial offer and all following propositions of renegotiation. Second, after the agent chooses one of the offers there is at least one more round of offers and neither the principal nor the agent can commit not to renegotiate. Thus, the renegotiation process can potentially last for an infinite number of rounds. However, third, the negative pay-off at infinite histories prevents the principal from stalling the negotiation indefinitely. Fourth, only the final signed contract is pay-off relevant and there is no discounting in-between negotiation rounds. Hence, the focus is on the effects of renegotiation rather than long-term relationships.

---

<sup>5</sup>As contracts can be infinitely renegotiated there is no guarantee that proposing only two contracts is without loss of generality, that is, the results obtained by Bester and Strausz (2004) do not apply to the setting at hand.

<sup>6</sup>We will drop the superscript from  $h^p$  and  $h^a$  whenever we refer to both or whenever it is unambiguous whose history is used.



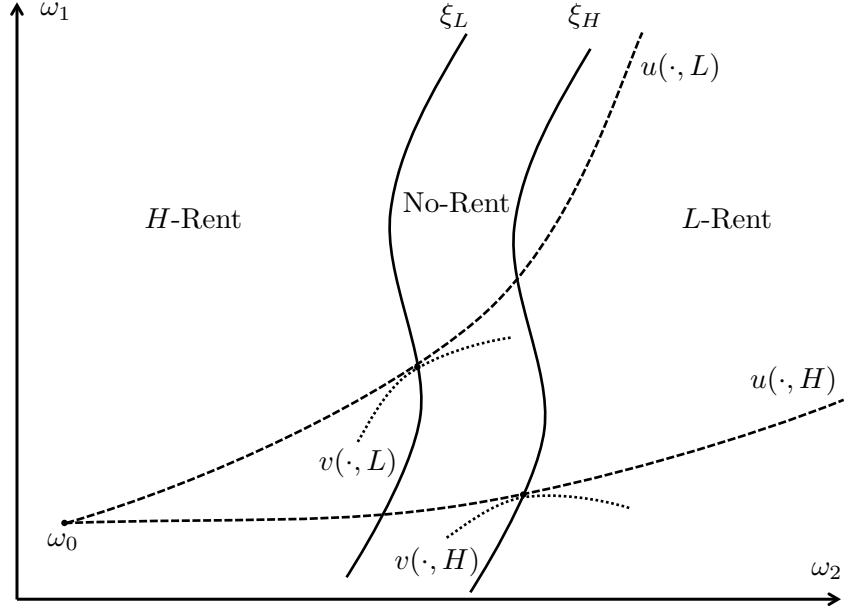


FIGURE 1. The setting.

We partition the contract space into three regions. We say that contracts that are left to  $\xi_L$  are in the 'H-Rent' configuration, contracts that are right to  $\xi_H$  are in the 'L-Rent' configuration, and contracts that are in the inner region between  $\xi_L$  and  $\xi_H$  are in the 'No-Rent' configuration. Essentially, the partition of the contract space reflects that in equilibrium once a contract has been signed that is, for example, in the 'H-Rent' configuration, further negotiation will leave the  $H$  type with a positive rent whereas the  $L$  type gains nothing from further negotiation.<sup>7</sup> Figure 1 depicts an illustration of the set-up.

**Strategies and beliefs.** Before we discuss the equilibria of the game it is useful to define strategies and beliefs of the principal and the agent. Denote by  $\mathcal{A}$  the set of all subsets of  $\mathbb{R}_+^2$  with at most  $K$  elements. A behavior strategy  $\sigma^p$  of the principal prescribes in each round  $t$  a distribution over contract menus  $\mathcal{M}_t \in \mathcal{A}$  conditional on the history  $h^p(t)$ .<sup>8</sup> That is,  $\sigma^p$  is a sequence of maps  $\sigma_t^p$  with

$$\sigma_t^p(h^p(t)) : \mathcal{H}^t \rightarrow \Delta(\mathcal{A}).$$

A behavior strategy  $\sigma^\theta$  of an agent of type  $\theta$  prescribes in each round  $t$  a probability distribution over contracts in  $\mathcal{M}_t \cup \{\omega_{t-1}\}$  conditional of the history  $h^a(t)$ . That is,  $\sigma^\theta$  is a sequence of maps

<sup>7</sup>The designation 'H-Rent', 'L-Rent', and 'No-Rent' configuration provides a very vivid definition and is taken from Strulovici (2014).

<sup>8</sup>We endow  $\mathcal{A}$  with the Borel sigma-algebra and denote by  $\Delta(\mathcal{A})$  the set of all probability measures over  $\mathcal{A}$ .

$\sigma_t^\theta$  with

$$\sigma_t^\theta(h^a(t)) : \mathcal{H}^t \rightarrow \Delta(\mathcal{M}_t \cup \{\omega_{t-1}\}).$$

At this,  $\Delta(\mathcal{M}_t \cup \{\omega_{t-1}\})$  denotes the set of all probability distributions over  $\mathcal{M}_t \cup \{\omega_{t-1}\}$ . A continuation strategy  $\sigma_+^{\{\theta,p\}}(t)$  is a truncated strategy. For example,  $\sigma_+^p(t) = \{\sigma_t^p, \sigma_{t+1}^p, \sigma_{t+2}^p, \dots\}$ .

The belief system of the principal is a sequence  $\{\mu_0, \mu_1, \dots\}$  where  $\mu_{t-1} \in [0, 1]$  are the beliefs held after a history  $h^p(t)$  that the agent is of type  $H$ .<sup>9</sup> The (second-order) belief system of the agent is a sequence  $\{\alpha_0, \alpha_1, \dots\}$  where  $\alpha_t$  is a probability measure on  $[0, 1]$  describing the belief of the agent over the belief of the principal. That is, for  $A \subset [0, 1]$ ,  $\alpha_t(A) = \Pr[\mu_t \in A]$ .

**Solution concept.** A Perfect Bayesian Equilibrium (PBE) is a strategy profile  $\sigma^p$  and  $\sigma^\theta$  and a belief system  $\{\mu_0, \mu_1, \dots\}$  and  $\{\alpha_0, \alpha_1, \dots\}$  that satisfy

- (i) For  $\theta \in \{L, H\}$ ,  $\sigma^\theta$  is a best reply to  $\sigma^p$  beginning at each history  $h(t)$ .
- (ii) Given  $\mu_t$ ,  $\sigma^p$  is a best reply to  $\sigma^\theta$  beginning at each history  $h(t)$ .
- (iii)  $\mu_t$  is derived from  $\mu_{t-1}$  given  $\sigma^\theta$  using Bayes' rule whenever possible
- (iv)  $\alpha_t(\{\mu_t\}) = 1$ .

It might seem rather unusual to explicitly include the second-order beliefs of the agent into the equilibrium definition. However, we will show in the subsequent sections that in the proposed equilibrium the strategy of the agent crucially depends on those beliefs. Thus, we call belief systems consistent given strategies  $\sigma^p$  and  $\sigma^\theta$  if the belief  $\mu_t$  of the principal results from bayesian updating with respect to these strategies and if the (second-order) belief of the agent is consistent, that is,  $\alpha_t(\{\mu_t\}) = 1$ . Whenever we discuss consistent beliefs, we will suppress the notation of the (second-order) belief of the agent and simply write  $\mu_t$  for both beliefs.

There are two more concepts that are useful to define before we turn to the solution of the game, that is, the state of the negotiation and the outcome function. A state of the negotiation in round  $t$  for a given pair of strategies and consistent belief system is  $C_t = (\omega_{t-1}, \mu_{t-1})$ . That is,  $\omega_{t-1}$  denotes the current signed contract and  $\mu_{t-1}$  the belief of the principal. The set of all states is denoted by  $\Gamma$  and  $\pi(C_t) = (1 - \mu_{t-1})v(\omega_{t-1}, L) + \mu_{t-1}v(\omega_{t-1}, H)$  denotes the expected utility of the principal in a given state. For a given history, pair of strategies and consistent belief system, the outcome function  $f(h^p(t), \sigma_+^p(t), \sigma_+^\theta(t)) \subset \Gamma$  gives the set of states after which the negotiation ends. That is,  $C = (\omega, \mu)$  is in  $f(h^p(t), \sigma_+^p(t), \sigma_+^\theta(t))$ , if there exists a  $t' \geq t$  such that a state  $C_{t'} = C$  is reached with positive probability starting from  $h^p(t)$  and  $\sigma^p(h^p(t')) = \emptyset$ . We call  $\Omega := \bigcup_{t \geq 0} f(h^p(t), \sigma_+^p(t), \sigma_+^\theta(t))$  the set of all (potential) outcomes of the negotiation and  $C = (\omega, \mu) \in \Omega$  an outcome of the negotiation.

<sup>9</sup>We slightly abuse notation as we suppress that different histories in period  $t$  might lead to a different posterior.

### 3. RENEGOTIATION-PROOF OUTCOMES

A solution to the proposed negotiation game is hard to obtain. As the principal cannot commit not to exploit the information revealed during the game, information revelation by the agent is difficult to ensure. Moreover, the number of potential negotiation rounds is unbounded. Hence, we cannot use backward induction as in Bester and Strausz (2004) to apply some kind of revelation principle. We sidestep this difficulties and analyze the properties of a set of outcomes  $\Omega$  that is generated by potential equilibrium strategies and derive necessary and sufficient conditions for  $\Omega$  such that the considered strategies indeed form an equilibrium. Such an approach has several advantages. First, having structured the set of potential outcomes greatly simplifies the construction of a PBE of the negotiation game. Second, and more importantly, our results imply that to characterize a PBE in a specific contracting setting it is sufficient to construct a set of outcomes  $\Omega$  that satisfies the defined conditions.

We start the analysis by proposing that potential equilibrium strategies of the principal and the agent in the negotiation game have two properties. First, both strategies are independent of previous play in the following sense: in any round  $t$  the strategy of the principal only depends on the current state  $C_t$  of the negotiation. The strategy of the agent only depends on the current signed contract, the proposal of the principal and the agents second order belief  $\alpha_t$ . Thus, we suppose that a potential equilibrium strategy of the principal can be written as  $\sigma_t^p(h^p(t)) = \sigma(C_{t-1})$  and a potential equilibrium strategy of the agent can be written as  $\sigma_t^\theta(h^a(t)) = \sigma(\mathcal{M}_t, \omega_t, \alpha_t)$ .<sup>10</sup> Second, in any round  $t$  and for any state  $C_{t-1}$  the strategy of the agent is such that for the principal it is optimal to either propose two contracts and end the game after one of those contracts is signed or end the game right away. That is, for the principal it is optimal to either propose  $\sigma^p(C_{t-1}) = \{\omega^L, \omega^H\}$  for some  $\omega^L, \omega^H \in \mathbb{R}_+^2$  and  $\sigma^p(C_t) = \emptyset$  or to propose  $\sigma^p(C_{t-1}) = \emptyset$  right away.<sup>11</sup>

The proposed strategies only depend on past play through the induced beliefs. Hence, we will drop the subscript  $t$  whenever it does not cause confusion. We are interested in the properties of the set of outcomes  $\Omega = \bigcup_{t \geq 0} f(h^p(t), \sigma_+^p(t), \sigma_+^\theta(t)) = \{C \in \Gamma; \sigma^p(C) = \emptyset\}$  generated by such strategies if the proposed strategies where to form an equilibrium. We will call  $C \in \Omega$  renegotiation-proof outcomes and start the characterization of renegotiation-proof outcomes by stating necessary conditions for such outcomes starting from a given state  $C$  with contract  $\omega$  and belief  $\mu$ .

---

<sup>10</sup>Such strategies capture the idea that bygones are bygones and thus there should be no reason to treat histories asymmetrically that start with the same contract and the same belief of the principal.

<sup>11</sup>Such strategies capture the idea that both the principal and the agent should be able to foresee the outcome of any subgame and thus without friction this outcome can be reached without delay.

**Definition 1** (*Feasibility*). We call  $C^L, C^H \in \Gamma$ , with  $C^\theta = (\omega^\theta, \mu^\theta)$ ,  $u(\omega^H, H) \geq u(\omega^H, L)$ , and  $u(\omega^L, L) \geq u(\omega^L, H)$ , feasible starting from  $C$  if the following conditions are satisfied:

- (i) (*Individual rationality of the agent*)  $u(\omega^\theta, \theta) \geq u(\omega, \theta)$ ,  $\theta \in \{L, H\}$
- (ii) (*Incentive compatibility*)  $u(\omega^H, H) > u(\omega^L, H) \Rightarrow \mu^L = 0$  and  $u(\omega^L, L) > u(\omega^H, L) \Rightarrow \mu^H = 1$
- (iii) (*Bayesian consistency*) There exists a  $p \in [0, 1]$  such that  $p\mu^L + (1-p)\mu^H = \mu$
- (iv) (*Individual rationality of the principal*)  $p\pi(C^L) + (1-p)\pi(C^H) \geq \pi(C)$

For  $C^H$  and  $C^L$  to be renegotiation-proof outcomes starting from  $C$  it is necessary that the agent is weakly better off compared to the initial situation in state  $C$  (requirement (i)). As the proposed strategy of the principal prescribes to end the negotiation at a renegotiation-proof outcome, the agent optimally chooses the contract that is most desirable for him. Thus, the principal has to take this into account when updating her belief (requirement (ii)). From the ex-ante point of view of the principal, the probability of reaching outcome  $C^L$  is some  $p \in [0, 1]$  (requirement (iii)). The principal is weakly better off if she agrees to negotiate towards  $C^L$  and  $C^H$  (requirement (iv)).

Let  $\mathcal{P}(\Gamma \times \Gamma)$  denote the power set of  $\Gamma \times \Gamma$  and  $X(C) : \Gamma \rightarrow \mathcal{P}(\Gamma \times \Gamma)$  denote the mapping of a state  $C$  to the set of all feasible outcomes starting from  $C$ .<sup>12</sup> Definition 1 merely states the essential necessary conditions. Therefore, the union over all feasible states will generally not yield a set of outcomes that is consistent with equilibrium play. Thus, the following definition introduces conditions that reflect that  $\Omega$  is generated by equilibrium strategies and narrows  $X(C)$  down to renegotiation-proof outcomes.

**Definition 2** (*Renegotiation-proofness*).  $\Omega \subset \Gamma$  is a set of *renegotiation-proof outcomes* if the following holds true.

- (i) (*Internal consistency*)  $C \in \Omega \Rightarrow \pi(C) = p\pi(C^L) + (1-p)\pi(C^H)$  for all  $(C^L, C^H) \in X(C)$  with  $C^\theta \in \Omega$ .
- (ii) (*External consistency*) For all  $C \in \Gamma$  there exist  $(C^L, C^H) \in X(C)$  with  $C^\theta \in \Omega$ .

External consistency is a consequence of the fact that  $\Omega$  is generated by strategies as described above. That is, in any state of the negotiation that is not a renegotiation-proof outcome the principal proposes two contracts and ends the negotiation afterwards. Internal consistency reflects that the proposed strategies should form an equilibrium. That is, it should not be feasible for the principal to deviate from an outcome in which the negotiation should end to another outcome after which the negotiation should end that would make him better off.

---

<sup>12</sup>Feasibility is defined for two potential outcomes. However, it is straightforward to extend feasibility to a single potential outcome  $C$  by setting  $C = C^L = C^H$ . In this case we will sometimes abuse notation and write  $C \in X(C)$ .

Internal and external consistency are a consequence of sequential rationality inherent in the equilibrium play of the principal in the negotiation game: suppose the negotiation reaches a renegotiation-proof outcome and the principal deviates from equilibrium play by proposing new feasible contracts that make him better off. External consistency ensures that if the principal follows his equilibrium strategy after the deviation, there exist contracts that make him even better off. Internal consistency then implies that the resulting contracts make her not better off than the original final outcome. We formalize this idea in the following section.

#### 4. SOLUTION OF THE NEGOTIATION GAME

To construct an equilibrium of the negotiation game we work our way backwards. We assume that a set  $\Omega$  that satisfies the conditions lined out in Definition 2 exists and then construct a PBE that generates  $\Omega$  as the set of outcomes. For this we need to define the strategy of the principal in a way such that indeed  $\Omega$  is generated as the set of final outcomes and that the principal behaves optimally. Thus, suppose an  $\Omega$  that satisfies the conditions of Definition 2 exists and define for each state  $C = (\omega, \mu)$

$$\begin{aligned} s(C) &= \arg \max_{(C^L, C^H)} p\pi(C^L) + (1-p)\pi(C^H) \\ \text{s.t. } & (C^L, C^H) \in X(C) \\ & C^\theta = (\omega^\theta, \mu^\theta) \in \Omega \\ & \mu = p\mu^L + (1-p)\mu^H \end{aligned}$$

as the optimal optimal outcomes in  $\Omega$  starting from state  $C$ .<sup>13</sup> In what follows we will slightly abuse notation and for  $s(C) = ((\omega^H, \mu^H), (\omega^L, \mu^L))$  define  $s^\theta(C) := \omega^\theta$ .

To ensure that the problem is well-behaved we make three assumptions. Imposing these assumptions is merely a matter of convenience as it greatly simplifies notation. In Section 5 we show that all of the assumptions are satisfied for all of the considered applications.<sup>14</sup>

**Assumption 1.** *For all  $C$  in  $\Gamma$ ,  $s(C)$  exists.*

**Assumption 2.** *For every  $C = (\omega, \mu)$  one of the following holds true*

- (i) *If  $\omega$  is in the 'H-Rent' configuration,  $u(s^L(C), L) = u(\omega, L)$  and  $u(s^H(C), H) \geq u(\omega, H)$ .*
- (ii) *If  $\omega$  is in the 'L-Rent' configuration,  $u(s^H(C), H) = u(\omega, H)$  and  $u(s^L(C), L) \geq u(\omega, L)$ .*
- (iii) *If  $\omega$  is in the 'No-Rent' configuration,  $u(s^L(C), L) = u(\omega, L)$  and  $u(s^H(C), H) = u(\omega, H)$ .*

<sup>13</sup>If  $C \in \Omega$ ,  $s(C) = (C, C)$ .

<sup>14</sup>That is, private values, common values Spence type and common values Rothschild-Stiglitz type.

**Assumption 3.** For  $\theta \neq \theta' \in \{L, H\}$  and any two contracts  $(\omega^\theta, \omega^{\theta'})$  and belief  $\mu$ , if

$$u(s^\theta(w^\theta, 1), \theta) < u(s^\theta(w^{\theta'}, 0), \theta)$$

and

$$u(s^\theta(w^\theta, 1), \theta) > u(s^\theta(w^{\theta'}, \mu), \theta)$$

then there exists  $\rho \in (0, 1)$  such that

$$u(s^\theta(w^\theta, 1), \theta) = u(s^\theta(w^{\theta'}, \rho\mu), \theta).$$

Assumption 1 ensures that  $s(C)$  exists. Due to external consistency, Assumption 1 is always satisfied whenever  $\Omega$  is a closed set. Assumption 2 states that depending on the current state the optimal renegotiation proof outcome leaves at least one of the types without additional rent.<sup>15</sup> Assumption 3 needs some more explanation. Suppose the principal offers two contracts  $\omega^\theta$  and  $\omega^{\theta'}$  such that an agent with type  $\theta'$  chooses  $\omega^{\theta'}$  with probability one. Assume that after observing the choice of the agent the principal updates his belief, offers the optimal renegotiation-proof contracts and the negotiation ends thereafter. Suppose furthermore that in this case it is not better for a forward looking agent with type  $\theta$  to choose either of the contracts with probability one. Assumption 3 then ensures that there must exist a mixing probability such that the agent is indifferent between the two contracts if the principal believes that he mixes with such a probability.<sup>16</sup>

We are now in the position to state our main result:

**Theorem 1.** *Suppose Assumption 1 to 3 holds true.*

(i) *For each  $\Omega$  that satisfies Definition 2 there exists a Perfect Bayesian Equilibrium with equilibrium strategies  $\sigma^p$  and  $\sigma^\theta$  such that  $\Omega = \bigcup_{t \geq 0} f(h^p(t), \sigma_+^p(t), \sigma_+^\theta(t))$ .*

(ii) *Let  $(C^L, C^H) = s(C_0)$  with  $C^\theta = (s^\theta(\omega_0, \mu_0), \mu^\theta)$ . The equilibrium path for a given  $\Omega$  is characterized by:*

(a) *The principal offers in the first round  $\sigma^p(C_0) = \{s^L(\omega_0, \mu_0), s^H(\omega_0, \mu_0)\}$ .*

(b) *The agent of type L chooses contract  $s^L(\omega_0, \mu_0)$  with probability*

$$p^L = \frac{(1 - \mu^L)(\mu^H - \mu)}{(1 - \mu)(\mu^H - \mu^L)},$$

*that is,  $\sigma^L(\{s^L(\omega_0, \mu_0), s^H(\omega_0, \mu_0), \omega_0\}, \alpha(\{\mu_0\}) = 1) = (p^L, 1 - p^L, 0)$ .*<sup>17</sup>

<sup>15</sup>Note, that whether Assumption 2 is satisfied depends on the shape of  $\Omega$ .

<sup>16</sup>Whenever  $s(C)$  is continuous in  $\mu$ , the intermediate value theorem ensures that Assumption 3 holds true.  $s(C)$  is continuous in  $\mu$  whenever  $\Omega$  is a closed set.

<sup>17</sup>Recal that  $\sigma^\theta$  is a mapping from  $(\mathcal{M}_t, \omega_{t-1}, \alpha_{t-1})$  to  $\Delta(\mathcal{M}_t \cup \{\omega_{t-1}\})$ .

(c) The agent of type  $H$  chooses contract  $s^H(\omega_0, \mu_0)$  with probability

$$p^H = \frac{\mu^H(\mu - \mu^L)}{\mu(\mu^H - \mu^L)},$$

that is,  $\sigma^H(\{s^L(\omega_0, \mu_0), s^H(\omega_0, \mu_0), \omega_0\}, \alpha(\{\mu_0\}) = 1) = (1 - p^H, p^H, 0)$ .

(d) The negotiation ends in the following round as the principal does not propose a new contract.

*Proof.* The proof is relegated to Appendix A. □

The proof of Proposition 1 is a direct consequence of Definition 2. First, the strategy of the principal prescribes that in any current state of the negotiation that is not in  $\Omega$  she offers the optimal, feasible contracts in  $\Omega$ . Whenever the negotiation reaches a state in  $\Omega$  the principal ends the negotiation. If the agent observes an offer that is in  $\Omega$ , he chooses the contract that is optimal given his type (possibly mixing when indifferent). If the agent observes an offer that is not in  $\Omega$ , he chooses the contract that will lead to the optimal offer in the next period. The key at this point is to establish that there is one type of the agent who is best off by choosing one of the contracts with probability one. For the other type it is then optimal to mix between this contract and at most one other contract from the proposal. Second, we show that to prove that the proposed strategies form a perfect Bayesian equilibrium it is sufficient to consider only one-stage deviations of the principal. Third, from external consistency of  $\Omega$  it follows that once the principal has deviated and offered a contract that leads to an outcome that is not in  $\Omega$ , there exists a feasible outcome in  $\Omega$  that would make her better off. From internal consistency of  $\Omega$  it follows that once an outcome in  $\Omega$  has been reached, there is no profitable deviation that would lead to another outcome in  $\Omega$ .

## 5. APPLICATIONS

One of the main advantages of our approach is that in order to apply the results of Proposition 1 to specific principal-agent problems we merely need to construct the set  $\Omega$  of renegotiation-proof contracts. That is, we need to construct a set  $\Omega$  with the properties defined in Definition 2. In what follows we state three helpful results that will facilitate the construction of  $\Omega$ .

**Definition 3.** A state  $C$  of the negotiation is called efficient if  $X(C) = \{C\}$ .

That is, in an efficient state  $C$  there does not exist a set of outcomes which both the principal and the agent would weakly prefer. Thus, such a state must be part of any set of renegotiation-proof outcomes:

**Lemma 1.** *For every set  $\Omega$  that satisfies the conditions of Definition and every efficient state  $C$  it follows that  $C$  is an element of  $\Omega$ .*

*Proof.* Follows directly from external consistency: If  $C$  is the only element of  $X(C)$ , then  $C$  must be in  $\Omega$ .  $\square$

Note that Lemma 1 implies that if  $\omega^H \in \xi_H$ ,  $(\omega^H, 1)$  is a renegotiation-proof outcome. By the same token, if  $\omega^L \in \xi_L$ ,  $(\omega^L, 0)$  is a renegotiation-proof outcome.

**Lemma 2.** *A state of the negotiation  $C = (\omega, \mu)$  with  $\omega$  in the 'No-Rent' configuration is not renegotiation proof. That is, for every  $\Omega$  that satisfies the conditions of Definition 2,  $C \notin \Omega$ .*

*Proof.* Suppose the state of the negotiation is  $C = (\omega, \mu)$  with  $\omega$  in the 'No-Rent' configuration. Thus, there exist  $\omega^L \in \xi_L$  and  $\omega^H \in \xi_H$  such that  $((\omega^L, 0), (\omega^H, 1))$  is feasible starting from  $C$ . Moreover,  $(1 - \mu)\pi((\omega^L, 0)) + \mu\pi((\omega^H, 1)) > \pi(C)$ . As by Lemma 1  $((\omega^L, 0), (\omega^H, 1))$  are renegotiation-proof outcomes,  $C$  cannot constitute a renegotiation-proof outcome. This would violate internal consistency of Definition 2.  $\square$

If the current state of the negotiation is in the 'No-Rent' configuration, the optimal outcomes that leave both types of the agent with the same rent as the current state of the negotiation are feasible. Moreover, those outcomes are efficient and thus renegotiation proof. It follows from internal consistency of sets of renegotiation-proof outcomes that the current state of the negotiation cannot be renegotiation proof.

**Lemma 3.** *Let  $\Omega$  satisfy the conditions of Definition 2. If a state of the negotiation  $C$  is in  $\Omega$ , there does not exist a single feasible state that yields larger profit to the principal, that is,*

$$\forall C \in \Gamma, \text{ if there exists } C' \in X(C) \text{ such that } \pi(C') > \pi(C), \text{ then } C \notin \Omega.$$

*Proof.* Suppose there exists a  $C \in \Omega$  and a  $C'$  with  $\pi(C') > \pi(C)$ . It follows from internal consistency (Definition 2 (i)) that  $C' \notin \Omega$ . Thus, by external consistency (Definition 2 (ii)) there exists  $(C^L, C^H) \in X(C')$  with  $C^\theta \in \Omega$ ,  $\theta \in \{L, H\}$ . As,  $(C^H, C^L)$  are feasible starting from  $C'$  and  $(C', C')$  is feasible starting from  $C$ ,  $(C^H, C^L)$  are also feasible starting from  $C$ . Feasibility together with  $p\pi(C^L) + (1 - p)\pi(C^H) \geq \pi(C') > \pi(C)$  implies that internal consistency must be violated and thus  $C \notin \Omega$ .  $\square$

Lemma 3 states that for every outcome of the negotiation there does not exist a pooling state that would make the principal and both types of the agent better off. This is due to the fact that



such a state cannot be in the set of renegotiation proof outcomes as this would violate internal consistency. In this case however, external consistency ensures that there exist renegotiation-proof outcomes that improve on the original outcome which would violate internal consistency.

**5.1. Private values.** In the private values case, the utility of the principal is independent of the type of the agent, that is,  $v(\omega, \theta) = v(\omega)$ .<sup>18</sup> The situation corresponds, for example, to a monopolist selling different quantities (or qualities) of a good to bidders with heterogeneous valuations for the good.

In the standard single-period model without the potential for renegotiation, the principal offers two contracts. The contract for the  $H$  type is efficient and he is indifferent between his contract and the contract of the  $L$ -type. The contract of the  $L$  type is inefficient and provides him with the same utility as in his outside option. The exact position of the described contracts depends on the prior  $\mu_0$  of the principal. However, such contracts are not renegotiation proof in the sense defined above. With private values, only efficient outcomes are renegotiation proof:

**Proposition 1.** *With private values, the unique set of renegotiation-proof outcomes of the negotiation is*

$$\Omega = \{(\omega^L, 0); \omega^L \in \xi_L\} \cup \{(\omega^H, 1); \omega^H \in \xi_H\}.$$

*Proof.* It follows from Lemma 1 that the proposed  $\Omega$  has to be a subset of any set of renegotiation-proof outcomes, as every efficient state of the negotiation is renegotiation proof. It remains to show that no other state can be renegotiation proof.

Suppose the state of the negotiation is  $C^1 = (\omega^1, \mu^1)$  with  $\omega^1$  in the 'H-Rent' configuration. As  $\omega^1$  is then to the left of both efficient contract curves and the principals utility is independent of the type of the agent, the indifference curve of the principal is steeper than the indifference curves of both of the agents. Thus, independent of  $\mu^1$ , there exists a single contract  $\bar{\omega}^1$  that makes everyone better off. It follows from Lemma 3 that  $C^1$  cannot constitute a renegotiation-proof outcome.

Suppose the state of the negotiation is  $C^2 = (\omega^2, \mu^2)$  with  $\omega^2$  in the 'L-Rent' configuration. As  $\omega^2$  is then to the right of both efficient contract curves and the principals utility is independent of the type of the agent, the indifference curve of the principal is flatter than the indifference curves of both of the agents. Thus, independent of  $\mu^2$ , there exists a single contract  $\bar{\omega}^2$  that makes everyone better off. It follows from Lemma 3 that  $C^2$  cannot constitute a renegotiation-proof outcome. The case that in the state of the negotiation the current signed contract is in the 'No-Rent' configuration is covered by Lemma 2. □

---

<sup>18</sup>This is the case considered in Strulovici (2014).

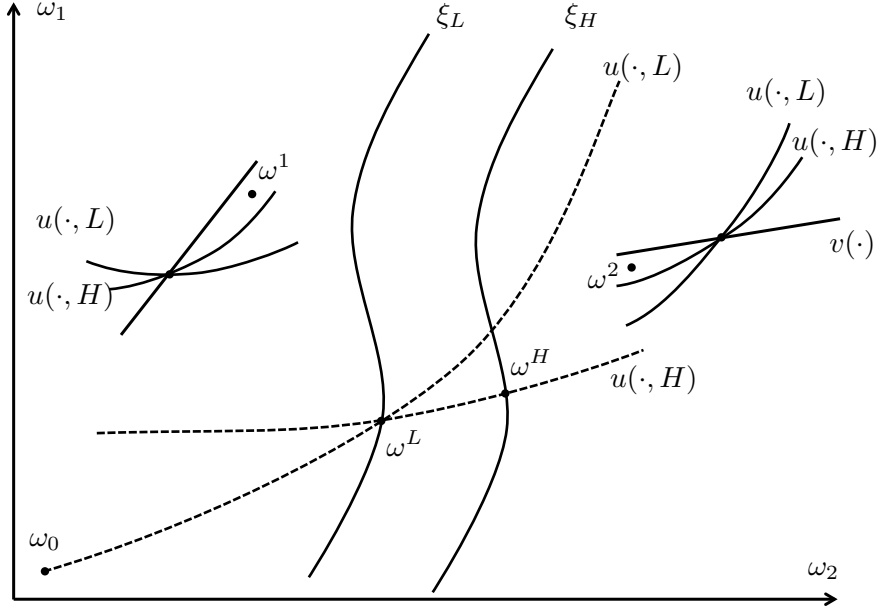


FIGURE 2. The equilibrium outcome of the negotiation game with private values.

Proposition 1 is the direct consequence of the fact that if the utility of the principal is independent of the type of the agent and the contract in the current state is in the 'H-Rent' ('L-Rent') configuration, there exists a pooling outcome that would make the principal and the agent strictly better off. Lemma 3 then implies that no such state can constitute a renegotiation-proof outcome. Given the set of renegotiation-proof outcomes it is easy to verify that Assumption 1 to 3 hold and Proposition 1 applies.

**Corollary 1.** *With private values, there exists a Perfect Bayesian Equilibrium of the negotiation game such that in the first round the principal offers the contracts  $\omega^L \in \xi_L$  and  $\omega^H \in \xi^H$  with  $u(\omega^L, L) = u(\omega_0, L)$  and  $u(\omega^H, H) = u(\omega^L, H)$ . The agent of type  $\theta$  chooses  $\omega^\theta$  with probability 1 and the principal ends the negotiation in the following round.*

The principal offers efficient contracts such that the  $L$  type receives the same utility as in his outside option and the  $H$  type is indifferent between his efficient contract and the contract of the  $L$  type. It is remarkable that in contrast to the single-period model the result is independent of the prior of the principal. This is illustrated in Figure 2.

**5.2. Common Values: The 'Spence' case.** We turn our attention to the case that the utility of the principal depends on the type of agent who signs the contract. We start by considering the 'Spence' case. The distinctive characteristic of the 'Spence' case is that the principal and the agent agree on the marginal trade-off's between types. Thus, the indifference curve of the principal for the

$H$  type is steeper than that for the  $L$  type. For example, in the education model in Spence (1973), the  $L$  type has larger costs of providing more effort. The same time, the additional productivity gained through additional education of the  $L$  type is smaller than of the  $H$  type.

As in the private values case, without the potential for renegotiation, the principal offers two contracts. The contract for the  $H$  type is efficient and he is indifferent between his contract and the contract of the  $L$ -type. The contract of the  $L$  type is inefficient and provides him with the same utility as in his outside option. Again, the exact position of the described contracts depends on the prior  $\mu_0$  of the principal. However, as in the private values case, only efficient contracts are renegotiation proof.

**Proposition 2.** *In the 'Spence' case, the unique set of renegotiation-proof outcomes of the negotiation is*

$$\Omega = \{(\omega^L, 0); \omega^L \in \xi_L\} \cup \{(\omega^H, 1); \omega^H \in \xi_H\}.$$

*Proof.* It follows from Lemma 1 that the proposed  $\Omega$  has to be a subset of any set of renegotiation-proof outcomes, as every efficient state of the negotiation is renegotiation proof. It remains to show that no other state can be renegotiation proof.

Suppose the state of the negotiation is  $C^1 = (\omega^1, \mu^1)$  with  $\omega^1$  in the ' $H$ -Rent' configuration. As  $\omega^1$  is then to the left of both efficient contract curves, the indifference curve of the principal for the  $L$  type is steeper than the indifference curve of the  $L$  type. Moreover, the indifference curve of the  $L$  type is steeper than the indifference curve of the  $H$  type. In the 'Spence' case the indifference curve of the principal for the  $H$  type is steeper than the indifference curve for the  $L$  type. It follows that in  $\omega^1$  both indifference curves of the principal are steeper than both indifference curves of the agent. Thus, independent of  $\mu^1$ , there exists a single contract  $\bar{\omega}^1$  that makes everyone better off. It follows from Lemma 3 that  $C^1$  cannot constitute a renegotiation-proof outcome.

Suppose the state of the negotiation is  $C^2 = (\omega^2, \mu^2)$  with  $\omega^2$  in the ' $L$ -Rent' configuration. As  $\omega^2$  is then to the right of both efficient contract curves, the indifference curve of the  $H$  type is steeper than the indifference curve of the principal for the  $H$  type. Moreover, the indifference curve of the  $L$  type is steeper than the indifference curve of the  $H$  type. As above, the indifference curve of the principal for the  $H$  type is steeper than the indifference curve for the  $L$  type. It follows that both indifference curves of the agents are steeper than both indifference curves of the principal. Thus, independent of  $\mu^2$ , there exists a single contract  $\bar{\omega}^2$  that makes everyone better off. It follows from Lemma 3 that  $C^2$  cannot constitute a renegotiation-proof outcome. The case that the state of the negotiation is in the 'No-Rent' configuration is covered by Lemma 2.  $\square$

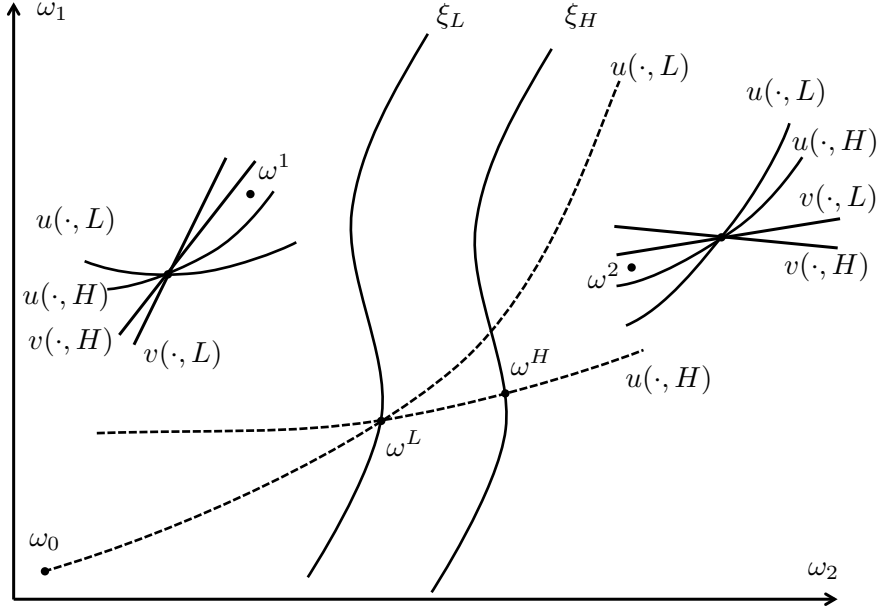


FIGURE 3. The equilibrium outcome of the negotiation game with common values 'Spence' type.

Interestingly, changing from private values to common values does not change the fact that only efficient contracts are renegotiation proof. As with private values, whenever the contract of the current state is in the 'H-Rent' ('L-Rent') configuration, there exists a pooling contract that would make the principal and the agent strictly better off. This is a direct consequence of the fact, that both indifference curves of the principal are steeper (less steep) than both indifference curves of the agent. Thus, Lemma 3 implies that no such state can constitute a renegotiation-proof outcome. This is illustrated in Figure 3. Given the set of renegotiation-proof outcomes it is easy to verify that Assumption 1 to 3 hold and Proposition 1 applies.

**Corollary 2.** *With common values 'Spence' type, there exists a Perfect Bayesian Equilibrium of the negotiation game such that in the first round the principal offers the contracts  $\omega^L \in \xi_L$  and  $\omega^H \in \xi_H$  with  $u(\omega^L, L) = u(\omega_0, L)$  and  $u(\omega^H, H) = u(\omega^L, H)$ . The agent of type  $\theta$  chooses  $\omega^\theta$  with probability 1 and the principal ends the negotiation in the following round.*

It could be argued that this result does not come as a surprise as efficient contracting could have been expected given that frictionless renegotiation should lead to an exploitation of all gains from trade. However, as we will show in the following section, this is not true if the agent and the principal rank the marginal trade-off's between types differently.

**5.3. Common Values: The 'Rothschild-Stiglitz' case.** The distinctive characteristic of the 'Rothschild-Stiglitz' case is that the agent and the principal do not agree on the marginal trade-off's between types. For example, in the insurance model in Rothschild and Stiglitz (1976), the  $L$  type has a lower risk probability. On one hand, for a marginal increase in premium he has to be compensated by a larger increase in indemnity than the  $H$  type. Thus, the  $L$  type indifference curve is steeper than the indifference curve of the  $H$  type. On the other hand, a marginal increase in indemnity is more costly for the principal if the agent is of the  $H$  type than if the agent is of the  $L$  type.

Under the assumption that the efficient contract curve of the  $L$  type lies to the left of the efficient contract curve of the high type, the solution of the one-shot negotiation is similar to the solution of the one-shot negotiation in the private values and the 'Spence' case: the principal offers two contracts. The contract for the  $H$  type is efficient and he is indifferent between his contract and the contract of the  $L$ -type. The contract of the  $L$  type is inefficient and provides him with the same utility as in his outside option. The exact position of the described contracts depends on the prior  $\mu_0$  of the principal.

If renegotiation is taken into account, the renegotiation-proof outcomes in the 'Rothschild-Stiglitz' case can be remarkable different from the 'Spence' and private values case as a set containing only efficient outcomes may not be renegotiation-proof.

**Proposition 3.** *Let  $\omega^L \in \xi_L$  and  $\omega^H \in \xi^H$  such that  $u(\omega^L, L) = u(\omega, L)$  and  $u(\omega^H, H) = u(\omega^L, H)$  for some  $\omega \in \mathbb{R}_+^2$ . In the 'Rothschild-Stiglitz' case, if there exist  $\omega \in \mathbb{R}_+^2$  such that  $v(\omega, H) > v(\omega^H, H)$ , then  $\{(\omega^L, 0); \omega^L \in \xi_L\} \cup \{(\omega^H, 1); \omega^H \in \xi_H\}$  is not a set of renegotiation-proof outcomes.*

*Proof.* As  $v(\omega, H) > v(\omega^H, H)$  there exist an  $\mu < 1$  such that  $(1 - \mu)v(\omega, L) + \mu v(\omega, H) > (1 - \mu)v(\omega^L, L) + \mu v(\omega^H, H)$ . Thus, starting from outcome  $(\omega, \mu)$ ,  $(\omega^L, 0)$  and  $(\omega^H, 1)$  are not feasible. Moreover, as  $(\omega^L, 0)$  and  $(\omega^H, 1)$  are the cheapest efficient outcomes there does not exist other efficient outcomes that are feasible. Thus,  $\{(\omega^L, 0); \omega^L \in \xi_L\} \cup \{(\omega^H, 1); \omega^H \in \xi_H\}$  violates external consistency of Definition 2.  $\square$

The main difference between the 'Rothschild-Stiglitz' case and other cases is that in the other cases it is always feasible to reach efficient allocations from any state of the negotiation game. This is not true for the 'Rothschild-Stiglitz' case. To see this, consider some state with a contract in the 'H-Rent' configuration. As, for example, in the Spence case the  $H$ -type indifference curve of the principal lies above the  $L$ -type indifference curve, the optimal efficient contracts for the principal lie both on higher indifference curves than the contract of the original state. This not need be the

case in the 'Rothschild-Stiglitz' case as the  $H$ -type indifference curve of the principal is below the  $L$ -type indifference curve. In this case it could be that the optimal efficient contracts are such that the principal receives a strictly lower utility from contracting with the  $H$  type than in the contract of the original state. Thus, if the probability of facing the  $H$  type in the original state is high, the principal is strictly worse off with the efficient contracts. It follows that a set containing only efficient outcomes may not be renegotiation-proof.

Proposition 3 can be seen as an impossibility result. However, to show that inefficient equilibria of the negotiation game exist we need to show that indeed a set of renegotiation-proof outcomes  $\Omega$  exists. To do so we use internal and external consistency to construct such a set. For the construction we impose some more structure on our very general set-up by making the following assumption:

**Assumption 4.** *The utility functions of the principal and the agent are additively separable. That is,*

- (i)  $u(\omega, \theta) = f(\omega_2, \theta) + \lambda_\theta \omega_1$  and  $v(\omega, \theta) = g(\omega_2, \theta) - \omega_1$  with  $\lambda_\theta \in \mathbb{R}_+$  and
- (ii) if  $-f_1(\omega_2, L)/\lambda_L = g_1(\omega_2, L)$  then  $-f_1(\omega_2, H)/\lambda_H < g_1(\omega_2, H)$ .

If Assumption 4 holds true, we can use internal and external consistency to construct a set of renegotiation-proof contracts:

**Proposition 4.** *Suppose Assumption 4 holds true. A set  $\Omega$  of renegotiation-proof outcomes exists in the 'Rothschild-Stiglitz' case.*

*Proof.* The proof is relegated to Appendix A. □

The construction proceeds inductively. We start by working our way along all outcomes with contracts on a particular indifference curve of the agent. Inductively, we include those outcomes in  $\Omega$  that cannot be strictly improved upon by feasible outcomes that are already included in  $\Omega$ . We then use the additive separability to extend the construction to all indifference curves.

As suggested by Proposition 3 the constructed  $\Omega$  will contain inefficient outcomes. However, whether the final outcome will be inefficient depends on the model parameters:<sup>19</sup>

**Corollary 3.** *Let  $\omega^L \in \xi_L$  and  $\omega^H \in \xi^H$  such that  $u(\omega^L, L) = u(\omega_0, L)$  and  $u(\omega^H, H) = u(\omega^L, H)$ . With common values 'Rothschild-Stiglitz' type:*

---

<sup>19</sup>Given the set of renegotiation-proof outcomes constructed in the proof of Proposition 3 it is straightforward to verify that Assumption 1 to 3 hold and Proposition 1 applies.

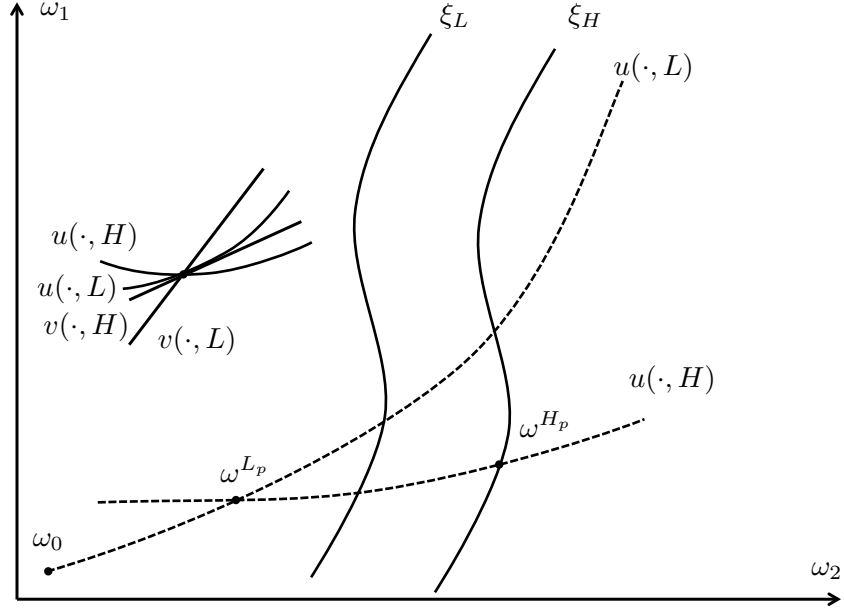


FIGURE 4. The equilibrium outcome of the negotiation game with common values 'Rothschild-Stiglitz' type.

- (i) If  $v(\omega_0, H) < v(\omega^H, H)$ , there exists a Perfect Bayesian Equilibrium of the negotiation game such that in the first round the principal offers the contracts  $\omega^L$  and  $\omega^H$ . The agent of type  $\theta$  chooses  $\omega^\theta$  with probability 1 and the principal ends the negotiation in the following round.
- (ii) Otherwise, there exist a  $\mu' < 1$  such that for all  $\mu_0 \geq \mu'$  there exists a Perfect Bayesian Equilibrium of the negotiation game in which the principal offers contracts  $\omega^{L_p} \notin \xi_L$  and  $\omega^{H_p} \in \xi_H$  with  $u(\omega^{L_p}, L) = u(\omega_0, L)$  and  $u(\omega^{H_p}, H) = u(\omega^{L_p}, H)$ . The L-type agent chooses  $\omega^{L_p}$  with probability 1 and the H-type agent chooses  $\omega^{H_p}$  with a probability strictly below 1. The principal ends the game in the following round.

*Proof.* Follows immediately from the construction of  $\Omega$  in the proof of Proposition 3. □

The condition  $v(\omega_0, H) < v(\omega^H, H)$  is restrictive and holds only for extreme realizations of the utility of the agent and the principal. Thus, the negotiation game will in most cases yield an inefficient outcome. Moreover,  $v(\omega_0, H) > v(\omega^H, H)$  does not imply that efficient outcomes are not feasible ex-ante. Hence, the outcome of the negotiation game can be inefficient, even if efficient outcomes were feasible ex-ante. The results are illustrated in Figure 4.

Note that in contrast to the one-shot negotiation that also yields inefficient outcomes a full separation of types does not occur if renegotiation is taken into account. On the equilibrium path, the H type always chooses the L-type's contract with a positive probability.

## 6. CONCLUDING REMARKS

In our opinion, one of the main contributions of this article is the characterization of the set of renegotiation-proof outcomes. Using internal and external consistency powerful yet simple results are achieved. The main advantage of this approach is that in contrast to other definitions of renegotiation-proofness, we do not assume that the outcome has to be efficient. Thus, allowing for inefficient outcomes as in the 'Rothschild-Stiglitz' case.

In the case at hand we were able to prove that the renegotiation-proof outcomes can indeed be supported by Perfect Bayesian Equilibrium of a general negotiation game. As long the type space remains finite the presented analysis can be extended even though one has to be careful with the definition of the (mixed) strategies of the agent. However, if type spaces become more complicated, the explicit derivation of strategies in the negotiation game becomes intractable. Nevertheless, we believe that the presented analysis based on the properties of renegotiation-proof sets still remains valid as it really captures the sequential rationality of the principal in a dynamic game. Thus, extending the 'axiomatic' analysis of renegotiation-proof outcomes to more complicated settings is insightful and should be part of the future research agenda. Examples of this approach include Asheim and Nilssen (1997), Gretschko and Wambach (2015), Vartiainen (2013), and Vartiainen (2014).



**Proof of Proposition 1.**

*Proof.* The proof proceeds in four steps:

Step 1: We specify the equilibrium strategies of the principal.

Step 2: We prove a Lemma that facilitates the exposition of the strategies of the agent.

Step 3: We specify the equilibrium strategies of the agent.

Step 4: We specify the belief updating of the principal and the agent

Step 5: We show that to prove that the proposed strategies form a PBE it is sufficient to consider only one-shot deviations.

Step 6: We prove that the proposed strategies form a PBE.

**Step 1: Definition of the equilibrium strategy of the principal**

The proposed equilibrium strategy of the principal in stage  $t$  only depends on the current state  $C_{t-1} = (\omega_{t-1}, \mu_{t-1})$  of the negotiation. If  $C_{t-1}$  is in  $\Omega$ , the principal offers  $\mathcal{M}_t = \emptyset$ . If  $C_{t-1}$  is not in  $\Omega$ , the principal offers  $\mathcal{M}_t = \{s^L(C_{t-1}), s^H(C_{t-1})\}$ . That is, the principal ends the game if the current state  $C_{t-1}$  is in  $\Omega$ . Otherwise, she proposes the optimal contracts in  $\Omega$  starting from state  $C_{t-1}$ :

$$\sigma_t^p(h^p(t)) = \sigma^p(C_{t-1}) = \begin{cases} \{s^L(C_{t-1}), s^H(C_{t-1})\} & \text{if } C_{t-1} \notin \Omega \\ \emptyset & \text{if } C_{t-1} \in \Omega \end{cases}.$$

**Step 2: A useful result**

**Lemma 4.** *For any proposal  $\mathcal{M}_t$  of the principal at least one of the following holds true:*

- (i) *there exists a contract  $\omega^L \in \mathcal{M}_t \cup \{\omega_{t-1}\}$  such that  $u(s^L(\omega^L, 0), L) \geq u(s^L(\omega, \mu), L)$ , for all  $\omega \in \mathcal{M}_t \cup \{\omega_{t-1}\}$  and all  $\mu \in [0, 1]$ .*
- (ii) *there exists a contract  $\omega^H \in \mathcal{M}_t \cup \{\omega_{t-1}\}$  such that  $u(s^H(\omega^H, 1), H) \geq u(s^H(\omega, \mu), H)$ , for all  $\omega \in \mathcal{M}_t \cup \{\omega_{t-1}\}$  and all  $\mu \in [0, 1]$ .*

*Proof.* Suppose first that all  $\omega \in \mathcal{M}_t \cup \{\omega_{t-1}\}$  are in the 'H-Rent' configuration. In this case, Assumption 2 (i), ensures that the contract  $\omega^L$  that maximizes  $u(\cdot, L)$  over  $\mathcal{M}_t \cup \{\omega_{t-1}\}$  has property (i). If all  $\omega \in \mathcal{M}_t \cup \{\omega_{t-1}\}$  are in the 'L-Rent' configuration, Assumption 2 (ii) ensures that the contract  $\omega^H$  that maximizes  $u(\cdot, H)$  over  $\mathcal{M}_t \cup \{\omega_{t-1}\}$  has property (ii). If all  $\omega \in \mathcal{M}_t \cup \{\omega_{t-1}\}$  are in the 'No-Rent' configuration, Assumption 2 (iii) ensures that the contract  $\omega^L$  that maximizes  $u(\cdot, L)$  over  $\mathcal{M}_t \cup \{\omega_{t-1}\}$  has property (i) and that the contract  $\omega^H$  that maximizes  $u(\cdot, H)$  over  $\mathcal{M}_t \cup \{\omega_{t-1}\}$  has property (ii).

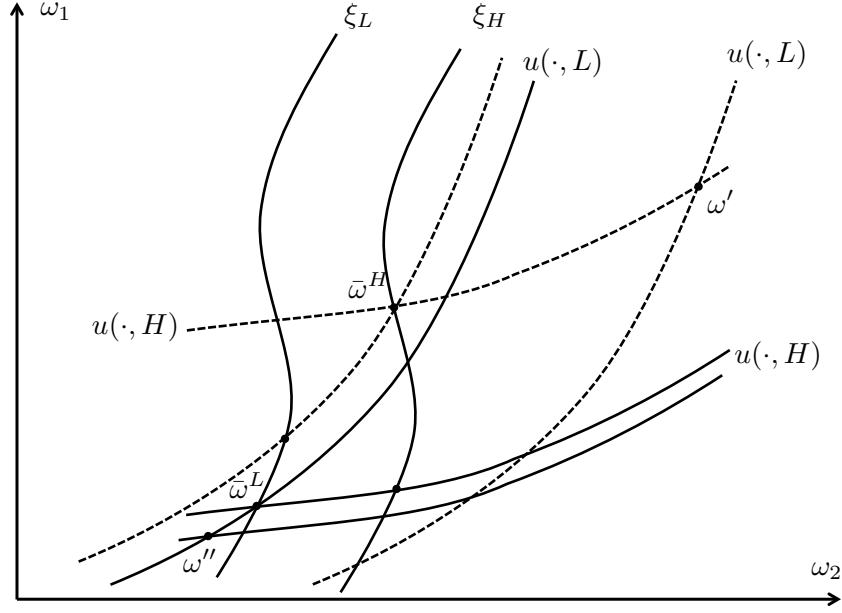


FIGURE 5. Sketch of proof of Lemma 4.

Finally, let  $\mathcal{M}_t \cup \{\omega_{t-1}\}$  be arbitrary. For a contradiction, suppose that (i) and (ii) do not hold true. As (i) does not hold there must exist a contract  $\omega' \in \mathcal{M}_t \cup \{\omega_{t-1}\}$  and  $\mu' > 0$  such that  $u(s^L(\omega', \mu'), L) > u(s^L(\omega, 0), L)$  for all  $\omega \in \mathcal{M}_t \cup \{\omega_{t-1}\}$ . Due to Assumption 2,  $\omega'$  must lie in the 'L-Rent' configuration. With the same argument there exists  $\omega'' \in \mathcal{M}_t \cup \{\omega_{t-1}\}$  in the 'H-Rent' configuration and  $\mu'' > 0$  such that  $u(s^H(\omega'', \mu''), H) > u(s^H(\omega, 0), H)$  for all  $\omega \in \mathcal{M}_t \cup \{\omega_{t-1}\}$ . Observe that  $u(s^L(\omega', \mu'), L)$  is bounded above by  $u(\bar{\omega}^H, L)$  where  $\bar{\omega}^H$  is the  $H$ -efficient contract that gives the  $H$ -type the same utility as  $\omega'$ , i.e.  $u(\omega', H) = u(\bar{\omega}^H, H)$  and  $\bar{\omega}^H \in \xi_H$ . This is due to Assumption 2 and the fact that efficient allocations are in  $\Omega$  (Lemma 1). As (i) does not hold true it must follow that  $u(\omega'', L) < u(\bar{\omega}^H, L)$ . By the same token,  $u(s^H(\omega'', \mu''), H)$  is bounded above by  $u(\bar{\omega}^L, H)$  where  $\bar{\omega}^L$  is the  $L$ -efficient contract that gives the  $L$ -type the same utility as  $\omega''$ , i.e.  $u(\omega'', L) = u(\bar{\omega}^L, L)$  and  $\bar{\omega}^L \in \xi_L$ . However, single crossing together with the fact that  $\xi_H$  lies to the right of  $\xi_L$  implies that  $u(\bar{\omega}^L, H) < u(\omega', H)$ . As  $u(\bar{\omega}^L, H) \geq u(s^H(\omega'', \mu''), H)$  this forms a contradiction to the assumption that (ii) does not hold true. The proof is illustrated in Figure 5.  $\square$

### Step 3: Definition of the equilibrium strategy of the agent

The strategy of the agent at stage  $t$  is more elaborate and depends on the current contract  $\omega_{t-1}$ , the set of proposed contracts of the principal  $\mathcal{M}_t$  and the agents second-order belief  $\alpha_{t-1}$ . We start by defining the strategy of the agent if  $\alpha_{t-1}(\{\mu_{t-1}\}) = 1$ , the current signed contract is  $\omega_{t-1}$  and the principal proposes  $\mathcal{M}_t = \{s^L((\omega_{t-1}, \mu_{t-1})), s^H((\omega_{t-1}, \mu_{t-1}))\}$ . In this case, let  $(C^L, C^H) = s(C_0)$

with  $C_\theta = (s^\theta(\omega_o, \mu_0), \mu^\theta)$ . The agent of type  $L$  chooses contract  $s^L(\omega_{t-1}, \mu_{t-1})$  with probability

$$p^L = \frac{(1 - \mu^L)(\mu^H - \mu)}{(1 - \mu)(\mu^H - \mu^L)},$$

that is,  $\sigma_t^L(h^a(t)) = \sigma^L(\{s^L(\omega_{t-1}, \mu_{t-1}), s^H(\omega_{t-1}, \mu_{t-1}), \omega_{t-1}\}, \alpha(\{\mu_{t-1}\}) = 1) = (p^L, 1 - p^L, 0)$ .<sup>20</sup>

The agent of type  $H$  chooses contract  $s^H(\omega_{t-1}, \mu_{t-1})$  with probability

$$p^H = \frac{\mu^H(\mu - \mu^L)}{\mu(\mu^H - \mu^L)},$$

that is,  $\sigma_t^H(h^a(t)) = \sigma^H(\{s^L(\omega_{t-1}, \mu_{t-1}), s^H(\omega_{t-1}, \mu_{t-1}), \omega_{t-1}\}, \alpha(\{\mu_{t-1}\}) = 1) = (1 - p^H, p^H, 0)$ .

Now suppose that again the current signed contract is  $\omega_{t-1}$ ,  $\alpha_{t-1}(\{\mu_{t-1}\}) = 1$  but  $\mathcal{M}_t \neq \{s^L(C_{t-1}), s^H(C_{t-1})\}$ . In this case, Lemma 4 ensures that there exists a type  $\theta$  and a contract  $\omega^\theta$  such that choosing  $\omega^\theta$  with probability 1 maximizes continuation pay-off given the equilibrium strategy of the principal. To save on notation we will only describe the case  $\theta = L$ .<sup>21</sup> In this case, the equilibrium strategy prescribes that the  $L$  type chooses  $\omega^L$  with probability 1. To define the equilibrium strategy of the  $H$  type we distinguish 3 cases:

- (i) If there exists  $\omega^H$  such that  $u(s^H(\omega^H, 1), H) > u(s^H(\omega^L, 0), H)$ , choose  $\omega^H \in \mathcal{M}_t \cup \{\omega_{t-1}\}$  with probability 1 that maximizes  $u(s^H(\cdot, 1), H)$
- (ii) If  $(s^H(\omega, 1), H) \leq u(s^H(\omega^L, \mu_{t-1}), H)$  for all  $\omega \in \mathcal{M}_t \cup \{\omega_{t-1}\}$ , choose  $\omega^L$  with probability 1.
- (iii) If neither of the two hold, choose with probability  $(1 - \rho)/(1 - \rho\mu_{t-1})$   $\omega^H \in \mathcal{M}_t \cup \{\omega_{t-1}\}$  that maximizes  $u(s^H(\cdot, 1), H)$  where  $\rho$  is such that  $u(s^H(\omega^H, 1), H) = u(s^H(\omega^L, \rho\mu_{t-1}), H)$ , choose  $\omega^L$  with the complementary probability.<sup>22</sup>

#### Step 4: Belief updating

Suppose, the principal offered  $\mathcal{M}_t = \{s^L(C_{t-1}), s^H(C_{t-1})\}$ . If the agent chooses,  $s^\theta(C_{t-1})$ , the principal updates her belief to  $\mu_t = \mu^\theta$ . If the principal o

As above, we only define belief updating for the case (i) of Lemma 4, that is, the case in which the  $L$ -type agent chooses  $\omega^L$  with probability one. The other case is defined in an analogous manner. If the principal observes that the agent chooses  $\omega^H$ , she updates her belief to  $\mu_t = 1$ , i.e., she believes that the agent is of the  $H$  type for sure. If the principal observes that contract  $\omega^L$  is chosen, she updates her belief to  $\mu_t = \tilde{\rho}\mu_{t-1}$  with  $\tilde{\rho} = 0$  in case (i),  $\tilde{\rho} = 1$  in case (ii), and  $\tilde{\rho} = \rho$  in case (iii). If the principal observes a choice  $\omega \notin \{\omega^L, \omega^H\}$ , she updates her belief to  $\mu_t = 1$ , i.e., she believes

<sup>20</sup>Recal that  $\sigma^\theta$  is a mapping from  $(\mathcal{M}_t, \omega_{t-1}, \alpha_{t-1})$  to  $\Delta(\mathcal{M}_t \cup \{\omega_{t-1}\})$ .

<sup>21</sup>The case  $\theta = H$  proceeds in exactly the same manner.

<sup>22</sup>Assumption 3 guarantees the existence of such a  $\rho$ .

that the agent is of the  $H$ -type for sure. The second-order beliefs of the agent are consistent with the beliefs of the principal. That is,  $\alpha_t(\{\mu_t\}) = 1$ .

This construction defines the strategies and beliefs for every possible history of the game.

### Step 5: One-shot deviations

As the negotiation game is played without discounting, it is not continuous at infinity and the standard one-shot deviation principle does not apply directly. However, we will show that irrespective of the strategy of the agent for every profitable deviation of the principal that may result in infinite history there is another profitable deviation strategy that only results in finite histories. In this case, the standard finite-deviations argument applies and to establish equilibrium it is sufficient to consider only one-shot deviations of the principal. For the agent, we will show that given the equilibrium strategy of the principal there is no deviation that results in an infinite history and makes the agent better off.

**Principal:** Let  $\sigma^p$  denote the strategy of the principal. Suppose, there exists another strategy  $\bar{\sigma}^p$  that may result in an infinite history and improves on the expected pay-off of  $\sigma^p$  by at least some  $\epsilon > 0$ . Two cases are relevant.

Case 1: There exists a history  $h'(t')$  of the game such that given  $\bar{\sigma}^p$  an infinite history  $h(\infty)$  with  $h_{t'}(\infty) = h'(t')$  will be reached with probability higher than some fixed  $\delta > 0$ . In this case, with probability higher than  $\delta$  the final payoff of the principal will be  $v(h(\infty)) = -\infty$ . Hence, the expected continuation pay-off starting from  $h'(t')$  is  $-\infty$ . It follows that replacing  $\bar{\sigma}^p(h'(t'))$  by  $\mathcal{M}_{t+1} = \emptyset$  yields a strictly higher payoff. Thus, there exists a strategy  $\tilde{\sigma}^p$  that results only in finite histories and improves the payoff of the principal as compared to  $\sigma^p$  by at least  $\epsilon$ .

Case 2: There exists no history  $h'(t')$  of the game such that given  $\sigma$  an infinite history  $h(\infty)$  with  $h_{t'}(\infty) = h'(t')$  will be reached with probability higher than some fixed  $\delta > 0$ . In this case, for a given history  $h'(t')$  denote by  $\mathcal{H}_{h'(t')} = \{h(t) \in \mathcal{H} \mid h_{t'}(t) = h'(t'), t > t'\}$  the set of all terminal histories that contain  $h'(t')$  and denote by  $P_{\bar{\sigma}^p}$  the probability distribution on  $\mathcal{H}_{h'(t')}$  induced by  $\bar{\sigma}^p$ . It follows that for every  $\delta > 0$  there exist a  $t_\delta > t'$  such that for every history  $\bar{h}(t_\delta)$  that coincides with  $h'(t')$  up to stage  $t'$ ,  $P_{\bar{\sigma}^p}(\{h(t) \in \mathcal{H}_{h'(t')} \text{ and } h_{t_\delta}(t) = \bar{h}(t_\delta)\}) < \delta$ . That is, the ex-ante probability at  $h'(t')$  that the game has not ended by stage  $t_\delta$  is smaller than  $\delta$ . Thus, it is possible to choose  $\delta$  sufficiently small such that the change in expected payoff starting from  $h'(t')$  of the principal from replacing  $\bar{\sigma}^p(\bar{h}(t_\delta))$  by  $\mathcal{M}_{t_\delta+1} = \emptyset$  is at most  $\epsilon/2$ . Hence, there exists a strategy  $\tilde{\sigma}^p$  that results only in finite histories and improves the payoff of the principal as compared to  $\sigma^p$  by at least  $\epsilon/2$ . Using the one-stage deviation principle for finite games yields the result.

**Agent:** Given the equilibrium strategy of the principal, the only deviation of the agent that would result in an infinite history is to choose  $\omega_t = \omega_0$  in every stage  $t$ . This is clearly not a profitable deviation for the agent.

### Step 6: Verification of equilibrium

As above, we will verify the equilibrium for case (i) of Lemma 4.

**Agent:** Suppose the negotiation reached round  $t$ , the current signed contract is  $\omega_{t-1}$ , the belief of the agent is  $\alpha_{t-1}(\{\mu_{t-1}\}) = 1$ , and the offer of the principal is  $\mathcal{M}_t = \{s^L(C_{t-1}), s^H(C_{t-1})\}$  with  $C_{t-1} = (\omega_{t-1}, \mu_{t-1})$ . Given the strategy of the principal, if the agent chooses either  $s^L(C_{t-1})$  or  $s^H(C_{t-1})$ , the negotiation will be over in round  $t + 1$ . Thus, it is optimal for the agent to choose the best contract available. Whenever the agent is indifferent between  $s^L(C_{t-1})$  and  $s^H(C_{t-1})$  any mixing between the contracts is optimal. Thus, choosing contract  $s^\theta(C_{t-1})$  with probability  $p^\theta$  is optimal as by Definition 1,  $p^\theta < 1$  if and only if  $u(s^\theta(C_{t-1}), \theta) = u(s^{\theta'}(C_{t-1}), \theta)$ . If the agent chooses  $\omega_{t-1}$ , the principal updates his belief to  $\mu_t = 1$  and proposes  $\mathcal{M}_{t+1} = \{s^L(\omega_{t-1}, 1), s^H(\omega_{t-1}, 1)\}$ . Neither type of the agent is strictly better off with choosing one of those contracts as compared to  $s^L(C_{t-1})$  or  $s^H(C_{t-1})$ . Thus, choosing  $\omega_{t-1}$  in round  $t$  cannot constitute a profitable deviation.

Now suppose the offer of the principal is  $\mathcal{M}_t \neq \{s^L(C_{t-1}), s^H(C_{t-1})\}$ . In this case, for whatever contract the agent chooses the principal will update her belief, the new state will be  $C_t$  and the principal will propose  $\mathcal{M}_{t+1} = \{s^L(C_t), s^H(C_t)\}$  and end the negotiation afterward. Thus, it is optimal of the agent of type  $\theta$  to choose a contract in  $\mathcal{M}_t \cup \{\omega_{t-1}\}$  such that  $s^\theta(C_t)$  is maximized. Precisely this logic is reflected in the definition of the strategy of the agent. Thus, the agent behaves optimal given the strategy of the principal.

**Principal:** As established above, it is sufficient to consider only one-shot deviations of the principal. We will consider two cases: the current state of the negotiation is not in  $\Omega$  and the current state is in  $\Omega$ .

Case 1: The negotiation reached round  $t$  and the state is  $C_{t-1} \notin \Omega$ . Ending the negotiation by proposing  $\mathcal{M}_t = \emptyset$  is dominated by the equilibrium strategy of proposing  $\mathcal{M}_t = \{s^L(C_{t-1}), s^H(C_{t-1})\}$ . This follows directly from the definition of  $s(C_{t-1})$ . Thus, suppose the principal offers a nonempty  $\mathcal{M}_t \neq \{s^L(C_{t-1}), s^H(C_{t-1})\}$  in round  $t$  and follows her equilibrium strategy afterward. In this case, the agent will choose either contract  $\omega^L \in \mathcal{M}_t \cup \{\omega_{t-1}\}$  or contract  $\omega^H \in \mathcal{M}_t \cup \{\omega_{t-1}\}$  as defined in the strategy of the agent above. In round  $t + 1$  the principal will update his belief and the state of the negotiation is either  $(\omega^H, 1)$  or  $(\omega^L, \tilde{\rho}\mu_{t-1})$ . Observe in that case replacing  $\omega^H$  with  $\bar{\omega}^H \in \xi^H$  such that  $u(\bar{\omega}^H, H) = u(\omega^H, H)$  makes the principal strictly better off. Thus, it is without loss to assume that  $(\omega^H, 1) \in \Omega$ . Two cases are relevant. First,  $(\omega^L, \tilde{\rho}\mu_{t-1}) \in \Omega$ , in this

case following her equilibrium strategy, the principal ends the negotiation. From the definition of  $s(C)$  it follows that this cannot make the principal strictly better off than following the proposed equilibrium strategy in round  $t$ . Second,  $(\omega^L, \tilde{\rho}\mu_{t-1}) \notin \Omega$ . In this case, after  $\omega^L$  was chosen by the agent, the principal, following her equilibrium strategy, proposes  $\mathcal{M}_{t+1} = \{s^L(C_t), s^H(C_t)\}$ , the agent chooses one of the contracts according to his equilibrium strategy and the principal updates her belief such that  $C_{t+1} \in \Omega$  and the negotiation ends. If  $s((\omega^L, \tilde{\rho}\mu_{t-1})) = (C^L, C^H)$  is feasible starting from  $C_{t-1}$ , by definition of  $s(\cdot)$  the principal cannot be better off than by having followed the proposed equilibrium strategy and proposed  $\mathcal{M}_t = \{s^L(C_{t-1}), s^H(C_{t-1})\}$ . Thus, it remains to establish that  $s((\omega^L, \tilde{\rho}\mu_{t-1})) = (C^L, C^H)$  is feasible starting from  $C_{t-1}$ . To do so suppose that  $\omega_t$  is in the 'H-Rent' configuration.<sup>23</sup> Let  $C^\theta = (\tilde{\omega}^\theta, \tilde{\mu}^\theta)$ . Due to single crossing and Assumption 2 it follows that  $u(s^H((\omega^L, \tilde{\rho}\mu_{t-1})_t), H) > u(s^L((\omega^L, \tilde{\rho}\mu_{t-1})), H)$ . Thus, by the definition of feasibility it follows that  $\tilde{\mu}^H = 1$ . To establish feasibility of  $(C^L, C^H)$  starting from  $C_{t-1}$  we check condition (iii) of Definition 1.<sup>24</sup> We have to show that there exist a  $p$  such that  $p\mu^L + (1-p)\mu^H = \mu_{t-1}$ . As  $\mu^H = 1$  it suffices to show that  $\mu^L \leq \mu_{t-1}$ . As  $\mu_t = \tilde{\rho}\mu_{t-1}$ , it follows that  $\mu_{t-1} \geq \mu_t$ . However, as  $(C^L, C^H)$  is feasible starting from  $C_t$  it follows that  $\mu_{t-1} \geq \mu_t \geq \mu^L$ .

Case 2: The negotiation reached round  $t$  and the state is  $C_{t-1} \in \Omega$ . Suppose the principal deviates from the proposed equilibrium strategy and instead of ending the negotiation proposes a nonempty set of contracts  $\mathcal{M}_t$  and follows his equilibrium strategy afterward. The agent will choose either contract  $\omega^L \in \mathcal{M}_t \cup \{\omega_{t-1}\}$  or contract  $\omega^H \in \mathcal{M}_t \cup \{\omega_{t-1}\}$  as defined in the strategy of the agent above. In round  $t+1$  the principal will update his belief and the state of the negotiation is either  $(\omega^H, 1)$  or  $(\omega^L, \tilde{\rho}\mu_{t-1})$ . Again it is without loss to assume that  $(\omega^H, 1) \in \Omega$ . Two cases are relevant. First,  $(\omega^L, \tilde{\rho}\mu_{t-1}) \in \Omega$ , in this case, following her equilibrium strategy, the principal ends the negotiation. As  $(\omega^L, \tilde{\rho}\mu_{t-1}) \in \Omega$  and  $(\omega^L, \tilde{\rho}\mu_{t-1})$  is feasible starting from  $C_{t-1}$ , it follows from internal consistency (Definition 2 (i)) that executing  $(\omega^L, \tilde{\rho}\mu_{t-1})$  cannot make the principal strictly better off compared to following his equilibrium strategy and executing  $C_{t-1}$ . Second,  $(\omega^L, \tilde{\rho}\mu_{t-1}) \notin \Omega$ . In this case, following her equilibrium strategy, the principal proposes  $\mathcal{M}_{t+1} = \{s^L(C_t), s^H(C_t)\}$ , the agent chooses one of the contracts according to his equilibrium strategy, the principal updates her belief such that  $C_{t+1} \in \Omega$  and the negotiation ends. As in Case 1 above,  $s((\omega^L, \tilde{\rho}\mu_{t-1})) = (C^L, C^H)$  is feasible starting from  $C_{t-1}$ . Thus, from internal consistency (Definition 2 (i)) it follows that proposing  $(C^L, C^H)$  cannot make the principal better off compared to following his equilibrium strategy and executing  $C_{t-1}$ .  $\square$

<sup>23</sup>This is consistent with case (i) of Lemma 4. The case that  $\omega_t$  is in the 'No-Rent' or 'L-Rent' configurations would proceed in an analogous manner.

<sup>24</sup>Conditions (i), (ii), and (iv) in Definition 1 are satisfied trivially.

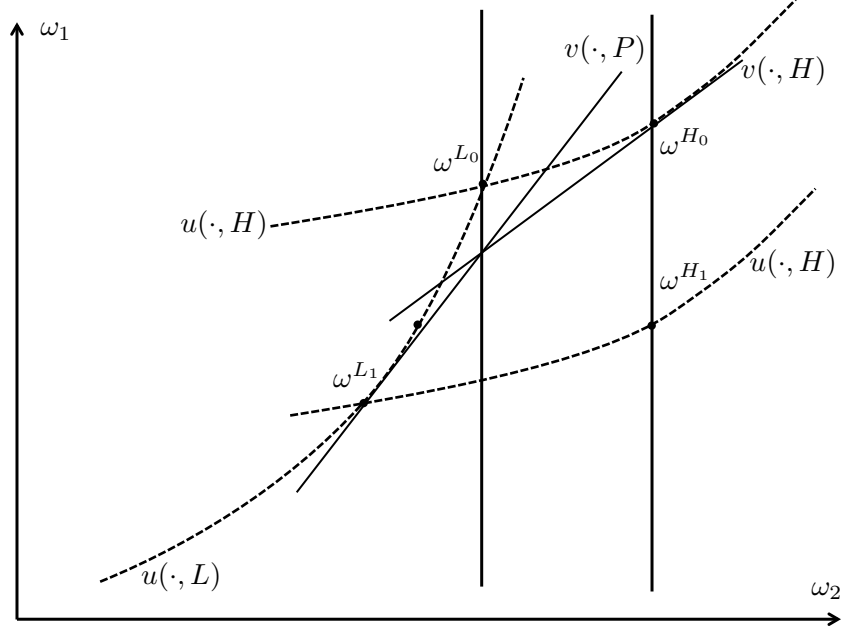


FIGURE 6. Construction of  $\Omega$  in the case of common values 'Rothschild-Stiglitz' type.

#### Proof of Proposition 4.

*Proof.* The construction of the the set of renegotiation-proof outcomes draws on ideas introduced in Asheim and Nilssen (1997). It differs in the following two respects. First, Asheim and Nilssen (1997) consider an insurance market. Thus, the efficient contract curves for the  $L$  type and the  $H$  type are the same. We extend the analysis to problems with distinct efficient contract curves. Second, they assume that no overinsurance takes place. Thus, their construction does not extend to the ' $L$ -Rent' configuration which is the case for our construction.

To show that a set of renegotiation-proof outcomes exists, we will construct such a set. The main idea of the construction is to make use of internal consistency of renegotiation-proof outcomes in Definition 2. We start by working our way along all outcomes with contracts on a particular indifference curve. Inductively, we will include those outcomes in  $\Omega$  that cannot be strictly improved upon by feasible outcomes that are already included in  $\Omega$ .

We start the construction for outcomes with contracts that are in the ' $H$ -Rent' configuration. Take some  $\omega^{L_0} \in \xi_L$ . Denote by  $\omega^{H_0} \in \xi_H$  the  $H$ -efficient contract that gives the  $H$  type the same utility as  $\omega^{L_0}$ . That is,  $u(\omega^{H_0}, H) = u(\omega^{L_0}, H)$ . From Lemma 1 it follows that  $C^{L_0} = (\omega^{L_0}, 0)$  and  $C^{H_0} = (\omega^{H_0}, 1)$  are in  $\Omega$ . Observe that  $(\omega, \mu)$  such that  $u(\omega, L) = u(\omega^{L_0}, L)$ ,  $u(\omega, H) \leq u(\omega^{H_0}, H)$  and  $v(\omega, H) < v(\omega^{H_0}, H)$  cannot constitute a renegotiation-proof outcome as  $(C^{L_0}, C^{H_0})$  is feasible and makes everyone strictly better off. This would violate internal consistency. Thus, for the

construction of  $\Omega$ , we will only consider outcomes such that given her belief the principal is indifferent between  $\omega$  and a potential negotiation towards  $(C^{L_0}, C^{H_0})$ . That is, we consider

$$\Gamma^-(C^{L_0}) := \{C = (\omega, \mu); u(\omega, L) = u(\omega^{L_0}, L), \mu < 1, (1 - \mu)\pi(C^{L_0}) + \mu\pi(C^{H_0}) = \pi(C)\}.$$

For all  $C = (\omega, \mu)$  and  $\bar{C} = (\bar{\omega}, \bar{\mu})$  in  $\Gamma^-(C^{L_0})$  it holds that whenever  $u(\omega, H) < u(\bar{\omega}, H)$  it follows  $\mu > \bar{\mu}$ . That is, moving along the indifference curve of the  $L$  type,  $\mu$  decreases for all  $C$  that are in  $\Gamma^-(C^{L_0})$ . From Lemma 3 it follows that only outcomes can be renegotiation proof for which no pooling outcome exists that would make everyone better off. Hence, whenever there exist an outcome  $(\omega^{L_1}, \mu^{L_1}) \in \Gamma^-(C^{L_0})$  such that the pooling indifference curve of the principal for  $\mu^{L_1}$  is tangent to the  $L$  type indifference curve in  $\omega^{L_1}$ , all outcomes with  $C = (\omega, \mu) \in \Gamma^-(C^{L_0})$  and  $u(\omega, H) \leq u(\omega^{L_1}, H)$  cannot constitute a renegotiation-proof outcome. Thus, include all outcomes  $C = (\omega, \mu) \in \Gamma^-(C^{L_0})$  in  $\Omega$  with  $u(\omega, H) > u(\omega^{L_1}, H)$ . Define  $C^{L_1} = (\omega^{L_1}, \mu^{L_1})$  and  $C^{H_1} = (\omega^{H_1}, 1)$  with  $\omega^{H_1} \in \xi_H$  and  $u(\omega^{H_1}, H) = u(\omega^{L_1}, H)$ . Continue the construction by considering

$$\Gamma^-(C^{L_1}) := \{C = (\omega, \mu); u(\omega, L) = u(\omega^{L_1}, L), \\ \mu < 1, (1 - \eta)\pi(C^{L_1}) + \eta\pi(C^{H_1}) = \pi(C), (1 - \eta)\mu^{L_1} + \eta = \mu\}.$$

As above, there may exist an outcome  $C^{L_2} = (\omega^{L_2}, \mu^{L_2}) \in \Gamma^-(C^{L_1})$  such that the pooling indifference curve of the principal for  $\mu^{L_2}$  is tangent to the  $L$  type indifference curve. In this case, include all outcomes  $C = (\omega, \mu) \in \Gamma^-(C^{L_0})$  in  $\Omega$  with  $u(\omega, H) > u(\omega^{L_1}, H)$ . Proceed by constructing  $\Gamma^-(C^{L_2})$ . The construction stops at  $\Gamma^-(C^{L_n})$  if in  $\Gamma^-(C^{L_n})$  there is no outcome such that the pooling indifference curve of the principal is tangent to indifference curve. In this case, include all outcomes in  $\Gamma^-(C^{L_n})$  in  $\Omega$ . The construction is sketched in Figure 6.

So far we have used internal consistency to construct a set renegotiation-proof outcomes along one particular indifference curve of the  $L$ -type agent. In the next step, we extend the construction to all potential outcomes in the ' $H$ -Rent' configuration. Again, the main idea is on one hand to include those outcomes in  $\Omega$  that cannot be strictly improved upon by feasible outcomes that are already in  $\Omega$  and on the other hand to only include outcomes that would not strictly improve on outcomes that are already in  $\Omega$ . Recall that in the RS case we assumed that the utility functions of the agent and the principal are additively separable. Thus, for all  $((\omega_1, \omega_2), \mu)$  that are in  $\Omega$  so far we also include  $((\lambda\omega_1, \omega_2), \mu)$  in  $\Omega$  for all  $\lambda > 0$ . From additive separability of the utility functions and the construction above, internal consistency of Definition 2 is not violated for all outcomes  $C = (\omega, \mu)$



in  $\Omega$  along the indifference curve of the  $L$ -type agent that contains  $\omega$ . Moreover, in the ' $H$ -Rent' configuration, for each outcome  $C = (\omega, \mu)$  whenever there is no feasible outcome  $C' = (\omega', \mu')$  with  $\omega'$  on the same  $L$  type indifference curve as  $\omega$  that strictly improves on  $C$ , there is also no feasible outcome which gives the  $L$  type a larger utility. Thus, the constructed  $\Omega$  does not violate internal consistency of Definition 2.

Before we proceed with the construction in the ' $L$ -Rent' configuration, we verify that the constructed  $\Omega$  does not violate external consistency of Definition 2 for outcomes with contracts in the ' $H$ -Rent' configuration. That  $\Omega$  does not violate internal consistency is a direct consequence of the construction. Take some outcome  $C = (\omega, \mu)$ . Two cases are relevant. First, there exist a  $\mu'$  such that  $(\omega, \mu') \in \Gamma^-(C^{L_k}) \cap \Omega$  for some  $k$ . In this case, whenever  $\mu < \mu'$ ,  $C^{L_k}$  and  $C^{H_k}$  are feasible. Whenever,  $\mu > \mu'$ ,  $(\omega, \mu')$  and  $(\bar{\omega}, 1)$  with  $u(\bar{\omega}, H) = u(\omega, H)$  and  $\bar{\omega} \in \xi_H$  are feasible. Second, no such  $\mu'$  exist. In this case  $C^{L_0}$  and  $C^{H_0}$  are feasible.

The construction of  $\Omega$  for outcomes with contracts in the ' $L$ -Rent' configuration mirrors the construction above for the ' $H$ -Rent' configuration. Take some  $\omega^{H_0} \in \xi_H$ . Denote by  $\omega^{L_0} \in \xi_L$  the  $L$ -efficient contract that gives the  $L$  type the same utility as  $\omega^{H_0}$ . That is,  $u(\omega^{L_0}, H) = u(\omega^{H_0}, H)$ . As above inductively define,  $C^{H_k}$  and  $C^{L_k}$  as well as

$$\Gamma^-(C^{H_k}) := \{C = (\omega, \mu); u(\omega, H) = u(\omega^{H_k}, H), \\ \mu > 0, (1 - \eta)\pi(C^{L_k}) + \eta\pi(C^{H_k}) = \pi(C), (1 - \eta)\mu^{H_k} = \mu\}.$$

The rest of the construction proceeds in exactly the same manner as the construction of renegotiation-proof outcomes with contracts in the ' $H$ -Rent' configuration and is therefore omitted. From Lemma 2 it follows that outcomes with contracts in the ' $No$ -Rent' configurations cannot be part of a set of renegotiation-proof outcomes. This completes the construction.  $\square$

## REFERENCES

- ASHEIM, G. AND T. NILSSEN (1997): "Insurance monopoly and renegotiation," *Economic Theory*, 354, 341–354.
- BEAUDRY, P. AND M. POITEVIN (1993): "Signalling and Renegotiation in Contractual Relationships," *Econometrica*, 61, 745–782.
- BESTER, H. AND R. STRAUZ (2004): "Contracting with Imperfect Commitment and the Revelation Principle: The Single Agent Case," *Econometrica*, 69, 1077–1098.

- FUDENBERG, D. AND J. TIROLE (1990): “Moral Hazard and Renegotiation in Agency Contracts,” *Econometrica*, 58, 1279–1319.
- GRETSCHKO, V. AND A. WAMBACH (2015): “Inefficiencies in Contract (Re-)Negotiation with Infinite Type Spaces,” *mimeo*.
- HART, O. AND J. TIROLE (1988): “Contract Renegotiation and Coasian Dynamics,” *The Review of Economic Studies*, 55, 509–540.
- HÖRNER, J. AND L. SAMUELSON (2011): “Managing Strategic Buyers,” *Journal of Political Economy*, 119, 379–425.
- KRASA, S. (1999): “Unimprovable Allocations in Economies with Incomplete Information,” *Journal of Economic Theory*, 87, 144–168.
- MAESTRI, L. (2012): “Dynamic Contracting under Adverse Selection and Renegotiation,” *mimeo*.
- ROTHSCHILD, M. AND J. STIGLITZ (1976): “Equilibrium in Competitive Insurance Markets : An Essay on the Economics of Imperfect Information,” *The Quarterly Journal of Economics*, 90, 629–649.
- SKRETA, V. (2006): “Sequentially Optimal Mechanisms,” *The Review of Economic Studies*, 73, 1085–1111.
- (2013): “Optimal Auction Design under Non-Commitment,” *mimeo*, 1–37.
- SPENCE, M. (1973): “Job Market Signaling,” *The Quarterly Journal of Economics*, 87, 355–374.
- STRULOVICI, B. (2014): “Contract Negotiation and the Coase Conjecture,” *mimeo*.
- VARTIAINEN, H. (2013): “Auction Design Without Commitment,” *Journal of the European Economic Association*, 11, 316–342.
- (2014): “Mechanism design without commitment with an application to bilateral bargaining,” *mimeo*.