

## Nash meets Rubinstein in final-offer arbitration

Muhamet Yildiz\*

MIT Economics Department

myildiz@mit.edu

<http://econ-www.mit.edu/faculty/myildiz/index.htm>

### **Abstract**

I consider a final-offer arbitration model in which the offers are submitted sequentially, the parties are allowed to accept offers, and the arbitrator maximizes Nash's social welfare function. I show that backwards induction in this three-period model leads to the subgame-perfect equilibrium outcome of Rubinstein's infinite-horizon alternating-offer bargaining game.

**Keywords:** Nash Bargaining, Rubinstein, bargaining, final-offer arbitration

**JEL Classification:** C78.

---

\*I thank Robert Gibbons, John Kennan, Alp Simsek, and Robert Wilson for helpful comments.

# 1 Introduction

Bargaining theory has two pillars: Nash's axiomatic bargaining solution, and Rubinstein's solution to the infinite-horizon bargaining with alternating offers. Nash (1950) has shown that if one wants to satisfy certain assumptions, then he must maximize the multiplication of the payoffs, which will be called the *Nash's social welfare function* hereafter. Rubinstein (1982) has shown that an infinite-horizon bargaining game with alternating offers has a unique subgame-perfect equilibrium. The two solutions are close when the discount rate is close to 1 (Binmore, Rubinstein, and Wolinsky (1987)), but they are distinct in general.

On the other hand, in formal negotiations, some alternative dispute resolution mechanisms are used if the parties do not reach an agreement before a certain deadline. An important mechanism is *final-offer arbitration*, introduced by Stevens (1966), in which the arbitrator has to choose between the offers submitted by the parties. In the United States, final-offer arbitration is frequently used to resolve interest disputes in public-sector employment and to determine the salaries of professional baseball players.

It turns out that these three important pieces in dispute resolution share an interesting story. If the arbitrator maximizes Nash's social-welfare function in the final-offer arbitration, then the unique subgame-perfect equilibrium outcome of arbitration is precisely the unique subgame-perfect equilibrium outcome in Rubinstein's infinite-horizon bargaining game with alternating offers. The equivalence remains intact when there are negotiations before the arbitration.

In formal negotiations, the parties often impose a deadline. When the deadline is short, however, the mechanism used at the deadline usually has a large impact on the outcome of the negotiation in equilibrium, introducing a bias in favor of one party. The result above points out a way to impose a deadline without affecting the bargaining outcome, regardless of how short the deadline is. It is using final-offer arbitration with an arbitrator who maximizes Nash's social-welfare function.

A more theoretical contribution of this paper to the bargaining theory is that the equilibrium dynamics of the model developed here does not resemble to the dynamics in usual bargaining models. As the paper demonstrates, unlike in the usual models, the actions of players have great impact on the equilibrium behavior of future players, and that is why the outcome of two-period negotiation here is the same as the outcome of the infinite-period negotiation in usual setup.

## 2 Final-Offer Arbitration with Sequential Offers

There are two negotiators, namely 1 and 2, and an arbitrator. Negotiators are to jointly choose a pair  $(x, y)$  from a convex, compact set  $X$ , where  $X \subset \mathbb{R}_+^2$  is understood to be the set of all feasible expected utility pairs for the negotiators after normalizing the disagreement payoffs to  $(0, 0) \in X$ . There are three dates  $t \in \{0, 1, 2\}$ . At date  $t = 0$ , Negotiator 1 submits an offer  $(x_0, y_0) \in X$  to the arbitrator. The offer is observable to Negotiator 2, who decides whether to accept the offer, ending the game with  $(x_0, y_0)$ , or wait and submit her own offer  $(x_1, y_1) \in X$  to the arbitrator at  $t = 1$ . Again, at  $t = 1$ , the offer  $(x_1, y_1)$  is observable, and Negotiator 1 decides whether to accept. If he accepts the offer,  $(x_1, y_1)$  is chosen and the game ends. Otherwise, the arbitrator makes the decision at  $t = 2$ , by choosing

$$(x_2, y_2) \in \{(x_0, y_0), (x_1, y_1)\}.$$

(In final-offer arbitration, the arbitrator has to choose one of the submitted offers.) If  $(x, y)$  is chosen at  $t$ , then the payoffs of negotiators 1 and 2 are  $\delta^t x$  and  $\delta^t y$ , respectively. The arbitrator's utility function at  $t = 2$  is Nash's (1950) social welfare function:

$$u_A(x, y) = xy.$$

The arbitrator's time preferences need not be specified because he moves only once. This perfect-information game is referred to as *sequential final-offer arbitration model*.

I make the following standard assumption: the function

$$f : x \mapsto \max \{y \mid (x, y) \in X\}$$

is concave, continuous, strictly decreasing;  $f(0) > 0$ , and  $f(\bar{x}) = 0$  for some  $\bar{x}$ .

There are two crucial modeling assumptions. First, the parties are allowed to accept the offers before the arbitrator makes a decision. This assumption commonly holds in real life. In fact, in the Major Baseball League, in 80% of the cases that are submitted to an arbitrator, the parties settle before the arbitrator makes a decision, which takes a month (see for example Wilson (1994)). Second, the offers are made sequentially. This assumption is sometimes natural because the party who files a case with an arbitrator has an incentive to submit his offer with the application and let the other party know what the offer is.

## 3 Transferable-Utility Case

In order to illustrate the main idea, consider the transferable utility case, in which  $X = \{(x, y) \mid x + y = 1\}$ . Here, I exclude the Pareto-inferior payoffs for clarity. At the last period, maximizing Nash's social welfare function, the arbitrator chooses the offer that is closer to

$(1/2, 1/2)$ . Now suppose that Negotiator 2 offers  $(x_1, y_1)$  such that  $x_1$  is closer to  $1/2$  than  $x_0$  is. Negotiator 1 would not reject such an offer because the arbitrator would select  $(x_1, y_1)$  in the next period anyway. If  $x_0$  is closer to  $1/2$ , then he accepts  $(x_1, y_1)$  if  $x_1 \geq \delta x_0$ , as  $(x_0, y_0)$  would be selected next day if he rejects the offer. Therefore, in equilibrium, Negotiator 2 counteroffers  $(x_1^*(x_0), 1 - x_1^*(x_0))$  with

$$x_1^*(x_0) = \min \{ \delta x_0, 1 - x_0 \}. \quad (1)$$

Notice that  $x_1^*(x_0)$  is maximized at  $x_0^* = 1/(1 + \delta)$ , which is the equilibrium offer in Rubinstein's model. In the first period, if Negotiator 1 offers  $x_0^*$ , then the other negotiator would accept that offer because she is indifferent between paying  $x_0^*$  in the first period and paying  $x_1^*(x_0^*) = \delta/(1 + \delta)$  in the next period. If he asks a higher payoff  $x_0 > x_0^*$ , she rejects it because such a *high* demand  $x_0$  allows her to counteroffer a *low* payoff  $x_1^*(x_0)$  in the next period, enticing her to wait. Indeed,

$$1 - x_0 < \delta/(1 + \delta) < \delta x_0 = \delta(1 - x_1^*(x_0)),$$

showing that accepting  $x_0$  at  $t = 0$ , which yields  $1 - x_0$ , is worse than counteroffering  $x_1^*(x_0)$  at  $t = 1$ , which yields  $\delta(1 - x_1^*(x_0))$ . But since the counteroffer depends on the initial offer, Negotiator 1 may still prefer to offer such  $x_0$  if it leads to a better counteroffer  $x_1^*(x_0)$ . It turns out that this is not the case:  $\delta x_1^*(x_0) < \delta/(1 + \delta) < x_0^*$ . Could he get a higher payoff than  $x_0^*$  by offering  $x_0 < x_0^*$  and receiving a better counteroffer  $x_1^*(x_0)$ ? The answer, again, turns out to be negative because  $x_1^*(x_0)$  is increasing in that region. Therefore, he offers  $x_0^* = 1/(1 + \delta)$  in the first period, and the offer is accepted—as in Rubinstein's infinite-horizon model.

The dynamics of the counteroffers here does not resemble to the dynamics in usual bargaining models. In usual bargaining, the past actions do not have any impact on the future behavior. When Negotiator 1 makes his offer, he takes the future counteroffers and the other negotiator's acceptance threshold given. On the other hand, in the sequential final-offer arbitration model, the past actions do affect the future equilibrium actions. The arbitrator's decision depends on the two offers, and the counteroffer depends on the initial offer. Now Negotiator 1 tries to affect the counteroffer and possibly the arbitrator's choice by his offer. In doing so, he ends up offering what he would have offered in infinite-horizon bargaining model of Rubinstein. Two main forces lead him to make such an offer. First, a higher optimal counteroffer  $x_1^*(x_0)$  makes Negotiator 2 more willing to accept  $x_0$ . Second, the optimal counteroffer,  $x_1^*(x_0) = \min \{ \delta x_0, 1 - x_0 \}$ , is maximized at the equilibrium offer in Rubinstein's model.

## 4 Preliminaries

This section describes the subgame-perfect equilibrium (henceforth SPE) in Rubinstein's model and the indifference curves of Nash's social welfare function.

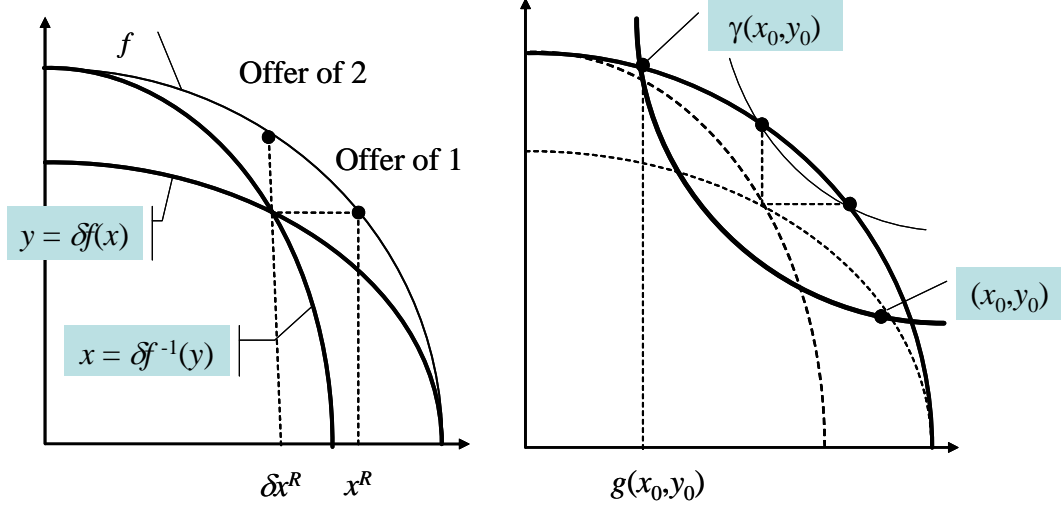


Figure 1: Rubinstein's solution (on left) and the indifference curves of Nash's social welfare function (on right)

**Rubinstein's solution** There is a unique SPE in Rubinstein's (1982) infinite-horizon, alternating-offer bargaining model. According to the SPE, Negotiator 1 offers a pair  $(x^R, y^R)$  with  $y^R = f(x^R)$ , and Negotiator 2 accepts it. Rubinstein's solution is illustrated on the left panel of Figure 1. Fixing the payoffs of Negotiator 1, scale down the payoffs of Negotiator 2 by  $\delta$ , obtaining a curve that is defined by  $y = \delta f(x)$ . Now fixing the payoffs of Negotiator 2, scale down the payoffs of Negotiator 1, obtaining another curve, defined by  $x = \delta f^{-1}(y)$ . The unique intersection of these curves is  $(\delta x^R, \delta f(\delta x^R))$ , which is of course defined by the equation

$$f(x^R) = \delta f(\delta x^R). \quad (2)$$

**Indifference Curves of Nash's Social Welfare Function** For any  $(x_0, y_0)$ , the indifference curve  $\{(x, y) | xy = x_0 y_0\}$  that contains  $(x_0, y_0)$  is plotted on the right-hand panel of Figure 1. If  $(x_0, y_0)$  is the Nash bargaining solution, then the indifference curve is tangent to the Pareto frontier, with a unique intersection at  $(x_0, y_0)$ . Otherwise, the indifference curve intersects the Pareto frontier twice. Let the best intersection for Negotiator 2 be  $\gamma(x_0, y_0) \equiv (g(x_0, y_0), f(g(x_0, y_0)))$ . Note that  $g(x_0, y_0)$  is the smaller of the two solutions to the equation

$$g(x_0, y_0) f(g(x_0, y_0)) = x_0 y_0. \quad (3)$$

The indifference curves of Nash's social welfare function and the Rubinstein's solution  $x^R$  have a special relation. As seen in Figure 1, the offer  $(x^R, f(x^R))$  of Negotiator 1 and the

offer  $(\delta x^R, f(\delta x^R))$  of Negotiator 2 are on the same indifference curve. Indeed, multiplying both sides of (2) by  $x^R$ , one obtains

$$x^R f(x^R) = \delta x^R f(\delta x^R).$$

Since  $\delta x^R < x^R$ , this yields

$$g(x^R, f(x^R)) = \delta x^R. \quad (4)$$

This equality is another way to define  $x^R$ , as it has a unique solution, and it will be the crucial step in the proof of the main result.

## 5 Main Result

**Proposition 1** *The outcome of the unique subgame-perfect equilibrium of the sequential final-offer arbitration model is the same as the outcome of the unique subgame-perfect equilibrium of Rubinstein's alternating-offer bargaining model: Negotiator 1 offers*

$$(x_0, y_0) = (x^R, f(x^R)),$$

and Negotiator 2 accepts the offer.

**Proof.** I will now apply backward induction, by breaking the ties towards an equilibrium.

At  $t = 2$ , given  $(x_0, y_0)$  and  $(x_1, y_1)$ , the arbitrator will choose

$$(x_2, y_2) = \begin{cases} (x_1, y_1) & \text{if } x_1 y_1 \geq x_0 y_0 \\ (x_0, y_0) & \text{otherwise.} \end{cases}$$

(Since Negotiator 2 moves second, in equilibrium, arbitrator chooses  $(x_1, y_1)$  in case of indifference.)

Now consider a  $t = 1$  history in which Negotiator 2 offers  $(x_1, y_1)$  after rejecting  $(x_0, y_0)$ . Suppose that  $x_1 y_1 \geq x_0 y_0$ . Then, Negotiator 1 foresees that if he rejects the offer  $(x_1, y_1)$ , in the next period, the arbitrator will choose the same decision  $(x_1, y_1)$ . That is clearly worse than accepting the offer at  $t = 1$ . Therefore, Negotiator 1 accepts the offer. Now suppose that  $x_1 y_1 < x_0 y_0$ . Then, rejection leads to implementing  $(x_0, y_0)$  at  $t = 2$  with payoff  $\delta^2 x_0$  to Negotiator 1, while acceptance leads to implementing  $(x_1, y_1)$  at  $t = 1$  with payoff  $\delta x_1$  to Negotiator 1. Hence, Negotiator 1 will accept the offer if and only if

$$x_1 \geq \delta x_0.$$

Now consider the node at which Negotiator 2 is to make an offer after rejecting  $(x_0, y_0)$ . It is not a best reply for Negotiator 2 to make an offer that will be rejected: as we have just seen,

such an offer will lead to the choice of  $(x_0, y_0)$  at  $t = 2$ , while he can get  $\gamma(x_0, y_0)$  accepted at  $t = 1$ , leading to a higher payoff. Therefore, Negotiator 2 offers  $(x_1^*(x_0, y_0), y_1^*(x_0, y_0))$  where

$$x_1^*(x_0, y_0) = \min \{g(x_0, y_0), \delta x_0\} \quad (5)$$

and

$$y_1^*(x_0, y_0) = f(x_1^*(x_0, y_0)). \quad (6)$$

Now consider  $t = 0$ . I will show that Negotiator 1 will offer  $(x^R, f(x^R))$ , and the offer will be accepted, as desired. The proof consists of three steps.

*Step 1:* If Negotiator 1 offers  $(x^R, f(x^R))$ , it is accepted, yielding the payoff of  $x^R$  for Negotiator 1.

*Proof of Step 1:* By (4),  $x_1^*(x^R, f(x^R)) = \delta x^R$  and hence  $y_1^*(x_0, y_0) = f(\delta x^R)$ . That is, if the offer  $(x^R, f(x^R))$  is rejected, then  $(\delta x^R, f(\delta x^R))$  will be implemented at  $t = 1$ —precisely as in the SPE of Rubinstein’s model. As in that model, Negotiator 2 is indifferent between accepting and rejecting the offer, and he accepts in equilibrium.

*Step 2:* Any offer  $(x_0, y_0)$  with  $x_0 > x^R$  is rejected and leads to a payoff strictly less than  $x^R$  for Negotiator 1.

*Proof of Step 2:* Note that  $(x_0, y_0)$  is in a lower indifference curve. Hence,

$$g(x_0, y_0) < g(x^R, f(x^R)) = \delta x^R < \delta x_0, \quad (7)$$

where the equality is by (4). Hence, by (5) and (6),

$$x_1^*(x_0, y_0) = g(x_0, y_0) \quad (8)$$

and

$$y_1^*(x_0, y_0) = f(g(x_0, y_0)) > f(\delta x^R), \quad (9)$$

where the inequality follows from (7) and the fact that  $f$  is strictly decreasing. Therefore,

$$\delta y_1^*(x_0, y_0) > \delta f(\delta x^R) = f(x^R) > f(x_0) \geq y_0,$$

where the equality is by (2), the next inequality is by the fact that  $f$  is strictly decreasing, and the last inequality is by definition of  $f$ . Since Negotiator 2 gets  $y_0$  from acceptance and  $\delta y_1^*(x_0, y_0)$  from rejection, she rejects the offer  $(x_0, y_0)$ , as claimed. To see the second part of the claim, note that, by (5) and (7), the continuation payoff for Negotiator 1 after such an offer is

$$\delta x_1^*(x_0, y_0) = \delta g(x_0, y_0) < \delta^2 x^R.$$

*Step 3:* Any offer  $(x_0, y_0)$  with  $x_0 < x^R$  leads to a payoff strictly less than  $x^R$  for Negotiator 1.

*Proof of Step 3:* If  $(x_0, y_0)$  is accepted, the payoff is  $x_0 < x^R$ . If  $(x_0, y_0)$  is rejected, then the continuation payoff, by (5), is

$$\delta x_1^*(x_0, y_0) \leq \delta^2 x_0 < x^R,$$

proving the claim.

Step 1 shows that the continuation value of offering  $(x^R, f(x^R))$  is  $x^R$ , which is strictly higher than the continuation value from any other offer, as established by Steps 2 and 3. Therefore, at  $t = 0$ , Negotiator 1 offers  $(x^R, f(x^R))$ , and it is accepted by Step 1. ■

Proposition 1 establishes that in a model of final-offer arbitration in which the parties are allowed to accept the other parties' offers the subgame-perfect equilibrium outcome is the same as the one in Rubinstein's infinite-horizon bargaining model. As we will see later, this equivalence remains intact when the parties are allowed to negotiate before the arbitration. It is crucial for Proposition 1 that the offers are made sequentially. If the parties submit the offers simultaneously, as in the existing literature, the result is different. In that case, both parties offer the arbitrator's ideal payoff, which is the Nash's bargaining solution in our case (Crawford, 1979).

A basic intuition can be gleaned from the optimal counteroffer

$$x_1^*(x_0, f(x_0)) = \min \{g(x_0, f(x_0)), \delta x_0\}$$

to a Pareto optimal offer  $(x_0, f(x_0))$ , where the share of Negotiator 2 is  $f(x_1^*)$ . In choosing  $x_0$ , Negotiator 1 has an incentive to maximize the counteroffer  $x_1^*$  because Negotiator 2 is more willing to accept offers when  $x_1^*$  is higher (and  $f(x_1^*)$  is lower). Clearly, as shown in Figure 2,  $\delta x_0$  is increasing in  $x_0$  and  $g(x_0, f(x_0))$  is decreasing in  $x_0$ . Therefore,  $x_1^*(x_0, f(x_0))$  is maximized at  $x_0$  with

$$g(x_0, f(x_0)) = \delta x_0.$$

This is precisely the equation (4) that defines the Rubinstein's offer  $x^R$ . The equality is quite intuitive. If Negotiator 1 makes a greedy offer at  $t = 0$ , then Negotiator 2 would counter it with a slightly more equitable offer, which leads to a slightly higher social welfare according to Nash. (The counteroffer is accepted because it would have been selected by the arbitrator anyway.) On the other hand, if Negotiator 1 makes a less greedy offer and Negotiator 2 happens to reject it, then Negotiator 2's counteroffer would simply extract the gain from implementing the original offer at  $t = 2$  rather than a period later. The equation equalizes these two incentives, which is also the property of Rubinstein's solution under equal discount rates.

## 6 Negotiation before Arbitration

I now introduce more rounds of alternating-offer negotiation prior to arbitration. If the parties are to go to arbitration exogenously at some given date  $T$  and if the arbitrator chooses



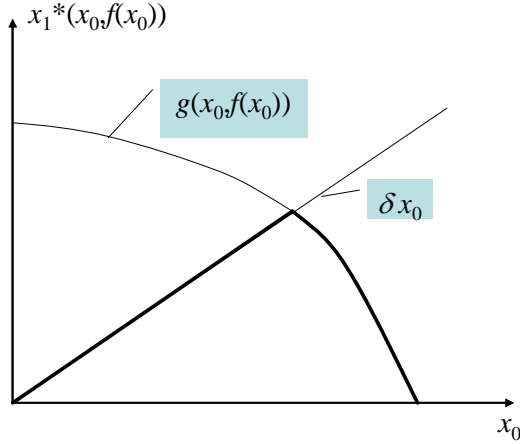


Figure 2: The payoff  $x_1^*$  of player 1 counteroffered by player 2, as a function of the initial offer  $x_0$  of player 1.

between the last two offers in the negotiation, then the SPE outcome will be precisely as in the Rubinstein's infinite-horizon model. This is simply because at round  $T - 2$  the outcome will be as in the Rubinstein's infinite-horizon model, and this leads to the Rubinstein's SPE actions in the earlier rounds by the fixed-point property of the alternating-offer bargaining. Here, I will consider a more interesting model of endogenous arbitration instead.<sup>1</sup>

**Endogenous Final-Offer Arbitration Model** Suppose that the possible dates are all natural numbers  $\{0, 1, 2, \dots\}$ . Negotiators 1 and 2 negotiate. At each  $t$ , one of the negotiators—Negotiator 1 at even dates and Negotiator 2 at odd dates—makes an offer, and the other negotiator decides whether to

- accept the offer, ending the game, or
- reject it and file for an arbitration in the next period, or
- reject it and remain in the negotiation.

In the arbitration, the party who files the case submits an offer to the arbitrator. The other party either accepts it or submits a counteroffer to the arbitrator (rejecting the former offer). The party who files the case, may accept the counteroffer or rejects it, in which case the arbitrator selects between the two offers and the game ends. The payoffs and the set  $X$  are as before.

Note that the only difference between filing a case and remaining in the negotiation is that in the latter case the party also triggers an arbitration procedure with his offer, ending

---

<sup>1</sup>This extension came out of a discussion with Alp Simsek.

the game after two more periods. Moreover, as in the usual bargaining models with outside options, only the responder has an option to file a case, presumably the proposer could not file a case when he has an offer that is not rejected yet.

The main result extends to this case intact:

**Proposition 2** *The unique subgame-perfect equilibrium outcome of the endogenous final-offer arbitration model is the same as the unique subgame-perfect equilibrium outcome of Rubinstein's infinite-horizon alternating-offer bargaining model: Negotiator 1 offers*

$$(x_0, y_0) = (x^R, f(x^R)),$$

*and Negotiator 2 accepts the offer.*

**Proof.** Since we can take the arbitration procedure as the outside option of the responder, the result follows from the result of Binmore, Shaked and Sutton (1989). Now, by Proposition 1, the value of going into final-offer arbitration, which is the responder's outside option, is precisely the same as remaining in the negotiation. Therefore, the unique subgame-perfect equilibrium outcome is the same as the one in Rubinstein's model with no outside option. ■

## References

- [1] Binmore, Ken, Ariel Rubinstein and Asher Wolinsky (1986): "The Nash bargaining solution in economic modelling," *Rand Journal of Economics*, 17, 176-188.
- [2] Binmore, Ken, Avner Shaked, and John Sutton (1989): "An Outside-option Experiment," *Quarterly Journal of Economics* 104, 753-770.
- [3] Crawford, Vincent (1979): "On Compulsory-Arbtration Schemes", *Journal of Political Economy*, 87, 131-159.
- [4] Nash, John (1950): "The Bargaining Problem", *Econometrica*, 18-2, 155-162.
- [5] Rubinstein, Ariel (1982): "Perfect Equilibrium in a Bargaining Model," *Econometrica*, 50-1, 97-110.
- [6] Stevens, Carl (1966): "Is Compulsory Arbitration Compatible with Bargaining?", *Industrial Relations*, 5, 38-52.
- [7] Wilson, Robert (1994): "Negotiation with Private Information: Litigation and Strikes", *Nancy L. Schwartz Memorial Lecture*.