

How Best to Disagree in Order to Agree?

Emily Tanimura, Sylvie Thoron

▶ **To cite this version:**

Emily Tanimura, Sylvie Thoron. How Best to Disagree in Order to Agree?. International Game Theory Review, 2016. hal-01704885

HAL Id: hal-01704885
https://hal.science/hal-01704885

Submitted on 8 Feb 2018

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

How Best to Disagree in Order to Agree?

Emily TANIMURA

Centre d'Economie de la Sorbonne, University of Paris 1, 106-112 bld de l'hopital, 75013, Paris FRANCE

e-mail: emily.tanimura@univ-paris1.fr

Sylvie THORON *Corresponding author

LIPHA Paris Est, University of Paris XII, 61 avenue du General de Gaulle, 94010 Creteil Cedex France

e-mail: sylvie.thoron@u-pec.fr

Abstract: We propose a new non-cooperative approach to implement a cooperative bargaining solution. This approach relies on a mechanism which specifies what happens when, at the end of the bargaining phase, players' propositions are still not compatible. The mechanism uses a random lottery and the history of proposals. At equilibrium, under the threat of this mechanism, provided that the use of it carries a small cost, players reach consensus after a finite number of proposals. No discounting is needed. The equilibrium strategies implement the full Raiffa bargaining process and players agree on a solution arbitrarily close to the Raiffa solution.

Keywords and phrases: non-cooperative bargaining theory, Raiffa bargaining solution, risk aversion, outside option, disagreement.

*We would like to thank for their detailed comments Francis Bloch, Steven Brams, Alan Kirman, Jean-Francois Laslier, Ariel Rubinstein, and also Ed Hopkins, Jozsef Sakovics and all the participants of the seminar at the University of Edinburgh, Francoise Forges and all the participants of the LEDa seminar at the University of Paris-Dauphine.

†

1. Introduction

In bargaining theory a number of cooperative solutions have been proposed and axioms provided. The basic problem has been to implement these solutions with a non-cooperative procedure. We propose a new approach which relies on a mechanism that specifies what happens when, at the end of the bargaining phase, players' propositions are not compatible. The mechanism imposes a specific procedure which draws one or the other of the last two non-compatible proposals, as in a final offer arbitration (Stevens 1966). However, here, the arbitrator is just a random lottery and the parties have the possibility to reject its outcome, in which case the mechanism goes back through the history of pairs of proposals of the bargaining phase. Although the mechanism may seem complex, its very existence ensures that it will never be used. At equilibrium, under the threat of this mechanism, provided that its use carries a small cost, players reach consensus after a finite number of proposals. No discounting is needed. The equilibrium strategies implement the full Raiffa bargaining process and players agree on a solution arbitrarily close to the Raiffa solution.

While the Raiffa bargaining solution does not hold as prominent a place in the bargaining literature as the Nash (1950), or the Kalai Smorodinsky (1975) solutions, it has clearly been the object of a recent revival of interest, mainly in the form of new axiomatic characterizations. An axiomatic characterization of the Raiffa solution was not provided in Raiffa's initial work but given later by Salonen (1988), Livne (1989), or in recent alternative axiomatizations by Anbarci and Sun (2013) or Trockel (2009). In addition, Diskin et al (2011) give an axiomatization which characterizes not only the final outcome but the Raiffa bargaining process of which the Raiffa solution is the limit. Myerson (1991) provided the first implementation of the Raiffa solution in a bargaining game where players are drawn alternatively as proposers for T periods or until the proposer's offer is accepted. At the equilibrium,

players find an agreement immediately and when the number of periods is large, the outcome approaches the Raiffa solution. More recently, Trockel (2011) showed that by modifying the Myerson game, an exact and not only asymptotic implementation can be obtained.

The procedure we propose implements Raiffa asymptotically as in the Myerson game but differs from the latter both in the description of the game and in the strategic behavior it generates. In particular, the Myerson game implements the Raiffa solution whereas our mechanism can be seen as implementing the full Raiffa bargaining process.

Our model consists of two phases. In the first bargaining phase, players make proposals to share one dollar. The exact procedure that is followed in this phase is not critical. Proposals can be simultaneous or sequential in some order. The main point is that the history of pairs of proposals is recorded. If parties fail to reach consensus in a given time, this triggers a mechanism phase in which the two players go through the following procedure. A fair lottery draws one or the other of the last two proposals recorded, at random. The player whose proposal was not chosen has to accept or reject. In case of rejection, the lottery draws, at random, one or the other of the previous proposals and the procedure goes back in this way through the history of the bargaining phase, as long as the proposals are rejected at each step. But it is only if the lottery has gone back through the whole history and if the players have always rejected, that they get their status quo payoff.

It is interesting to note that the way our mechanism operates implements, in a sense, in a bargaining game, an idea of Esteban and Sakovics (2008). Based on this they proposed an axiomatic solution concept, "agreements in the shadow of conflict". Their axiomatic approach assumes that one adds to the usual givens of a bargaining game a disagreement function f(S), specified for each subset of the bargaining set, and which gives each player's worst possible outcome when disagreeing over the set S. By excluding points that

are worse than those determined by the disagreement function, a new set is obtained and the disagreement function applied to the latter and so on until possibly only one point remains. We can view our mechanism as operating by explicitly constructing a type of disagreement function based on players' propositions.

The mechanism triggered at disagreement also provides the players with something reminiscent of an outside option. More precisely, in our setting, the latter is not a fixed function of the players' proposals but a game played in case of disagreement. However, it is a very simple game, which, if played rationally has a unique value. Bargaining with fixed outside options have been studied by Binmore, Shaked and Sutton (1989), Ponsati and Sackovics (1998) and Manzini and Mariotti (2001). Closer to the situation dealt with here are games where the outside option evolves as a function of the propositions. This is precisely the setting that Compte and Jehiel (2004) have shown to generate, under certain conditions, what is referred to in the literature as gradualism: the reluctance of players to make more than minimal concessions or in the extreme case no concessions at all. The main message is that gradualism is generated if concessions increase the outside option of the opponent. This is a discouraging message which invalidates the idea that the presence of a sequence of outside options that increases along the negotiation path could help players move ahead from the status quo. Compte and Jehiel's theoretical study explains, in a general setting, a phenomenon which was already observed in some particular procedures involving outside options depending on propositions. A relevant and well known example is the case of final offers arbitration, in which an arbitrator assigns a solution which is one or the other of the final propositions of the players. Widely used, this arbitration mechanism has nevertheless been shown, in theory and in practice to generate bargaining impasses (see Feuille (1975), Farber (1980), Brams (1983), or Dickinson (2006)). Our model can be seen as a form of final offer arbitration, which, at the cost of what might seem to be a more complex structure,

eliminates some of the problems of the original approach. Under our mechanism, players do not find an agreement immediately. However, convergence to consensus is rapid in a technically precise sense. Intuitively there is not gradualism as defined by Compte and Jehiel (2004), due to the fact that the outside option evolves favorably to a player with its own concessions. Indeed, we replace the outside option which assigns a payment based on proposals by a game. This game, if played rationally, actually has a higher value if a player makes even unilateral concessions.

At a first sight, the extensive recourse to lotteries, required to obtain these results might seem complicated. However, we note that if arbitration comes at a cost no lottery is actually every carried out at equilibrium. The implementation of the lottery is just a threat and the very knowledge of its existence by the players will prevent it from actually having to be used. By comparison, a seemingly simpler model such as the Myerson procedure, viewed as a mechanism, actually begins with a lottery: the one to draw the proposer. Moreover, the use of lotteries in arbitration, even if only putative, is easy to justify by the fact that it ensures neutrality: the solution obtained reflects only the preferences of the bargaining parties and not those of an external arbitrator. The additional structure in our mechanism also gives it some properties which may be appealing for its practical implementation, namely considerable robustness in a setting where the players are boundedly rational in the sense that they do not correctly anticipate the number of negotiation rounds.

We present the basic model in Section 2. Section 3 gives the characterization of the subgame perfect equilibria. We introduce discounting and a cost of implementation of the mechanism in Section 4. Section 5. discusses the performance of the mechanism in the presence of boundedly rational players. We conclude in Section 6.

2. Definitions, notation and setting of the model

In what follows, we will consider the following setting: two parties P^1 and P^2 bargain over the partition of a pie, which is taken to be, without loss of generality, of size 1. The possible ways to split the pie are

$$\{(x, y) : x, y \in [0, 1] \text{ and } x + y \leq 1\}$$

in which x is the share of party P^1 and y that of party P^2. Both parties are assumed to be (weakly) risk averse in the sense that they have concave utility functions, respectively $U^1(x), U^2(y)$ that are strictly increasing on $[0, 1]$. In terms of possible utility pairs $(u, v) \in \mathbb{R}^2$, the compact, convex bargaining set Δ is thus delimited by the points $u = U^1(0), v = U^2(0)$ and the curve $(U^1(x), U^2(1 - x))$, $x \in [0, 1]$. At the status quo, the parties' utilities are $U^1(0), U^2(0)$ respectively. We denote by G^T a sequential game with two phases. In the first negotiation phase, at each round t, $1 \leq t \leq T$, each party makes a proposal. Different cases can be considered, depending on which rules specifying that order are chosen. For example, the two parties can make simultaneous proposals or sequential proposals with a fixed order as in Rubinstein's bargaining game or sequential proposals in a random order as in Myerson's or Sjostrom's bargaining games. The outcome of round t is a pair of proposals (x_t, y_t). There is a finite number of rounds in the negotiation phase, T, which can be very large. Let us denote by h_t the t-history of the negotiation phase, generated by all the proposals up to and including round t: $(x_s)_{s \leq t}$ and $(y_s)_{s \leq t}$. If, at round t, $t \leq T$, $x_t + y_t \leq 1$ we will say that an agreement has been concluded. In this case, the game ends, Party 1 gets x_t and Party 2, y_t (in equilibrium the case $x_t + y_t < 1$ will in fact never arise). The negotiation phase can also end at $t < T$ if at least one player ceases to make further concessions: $x_t = x_{t-1}$, or similarly for y_t.

If the negotiation phase ends at date t without an agreement, a second

phase starts, triggering a mechanism based on the history of the negotiation phase. This mechanism phase has also a maximum of t rounds. At the first round of the mechanism phase, a fair lottery draws randomly one of the last two proposals of the negotiation phase. The party whose proposal has not been chosen by the lottery can reject it and this rejection will then trigger a new lottery between the previous proposals of the negotiation phase. Therefore, as the mechanism proceeds, when proposals are rejected, earlier and earlier pairs of proposals are used.

More precisely, at each round n, $0 \leq n \leq t-1$, a fair lottery can be implemented between the proposals x_{t-n} and y_{t-n} in which each proposal is drawn with probability $\frac{1}{2}$. At each round, when one party's proposal has been drawn, the other party has to accept or reject this outcome. When the outcome is rejected at round n, the mechanism proceeds to the following round in the same manner. At the earliest round n at which the outcome x_{t-n} (or y_{t-n}) is accepted, the game stops. Party 1 gets x_{t-n} $(1-y_{t-n})$ and Party 2 gets $1-x_{t-n}$ (y_{t-n}). If, at round $n=t-1$ the outcome of the lottery is rejected, the game ends and the two party get their status quo payoffs $U^1(0)$ and $U^2(0)$ respectively. We will first consider that there is no discounting. We will consider a case with discounting latter in the paper but we will show that it does not play a determinant role here.

We will denote by $\mathcal{L}_t$ the lottery between x_t and y_t. An outcome of game G^T is defined by a history of proposals h_t, $t \leq T$, in which (x_t, y_t) is the only agreement. When there is no agreement in h_t, it is completed by the outcomes of the lotteries $\mathcal{L}_t, \mathcal{L}_{t-1}, , \mathcal{L}_s$ where $\mathcal{L}_s$ is the first lottery in which the proposal that was drawn was accepted. As we have noted previously, the fact that the lotteries $\mathcal{L}_t, , \mathcal{L}_s$ were implemented necessarily implies that the party whose decision was not drawn rejected his partner's proposal in all the previous lotteries $\mathcal{L}_t, \mathcal{L}_{t-1}, \mathcal{L}_{s+1}$

We will use the convenient notation $U^1(\mathcal{L}_t) = \frac{1}{2}U^1(x_t) + \frac{1}{2}U^1(1-y_t)$.

We note that the recursive structure of the lotteries can be represented in its extensive form as a tree. If the first lottery was implemented after date t, the tree has length t. The first branches from the root lead to two nodes at which x_t and y_t respectively are proposed. Each party intervenes at the nodes where his opponent's proposal is drawn and decides to accept or reject it. If x_t or y_t is rejected at level t, we follow one of the branches to either x_{t-1} or y_{t-1}, which will be one level further from the root. Each node has two branches leading to successor nodes. Whenever the proposal at a node is rejected, each branch leading to a successor node at the level below (i.e. one step further from the root) is chosen with probability $\frac{1}{2}$.

3. Characterizing the equilibria

In this section we characterize the equilibria of game G^T. The proof involves two phases. First, we take the proposals $x = (x_s)_{s\leq t}$ and $y = (y_s)_{s\leq t}$ as given and we consider the strategies for accepting or rejecting proposals in the extensive game that starts with the implementation of the first lottery $\mathcal{L}_t$ which is expected by the players at the end of the negotiation phase at some date t. Subgame perfection of these strategies imposes constraints on the proposals that are accepted. For any proposals $x = (x_s)_{s\leq t}$ and $y = (y_s)_{s\leq t}$, these constraints define the expected payoffs of the parties when a lottery is about to be implemented. This can then be used to characterize the proposal strategies at equilibrium.

We define two recursive functions V_t^1 and V_t^2 which depend on $h_t = (x_s, y_s)_{s\leq t}$, the vector of propositions up to and including date t.

$$V_0^1 = U^1(0), V_0^2 = U^2(0).$$

$$V_t^2 = \frac{1}{2}\left\{ \begin{array}{cc} U^2(y_t) \; if \; U^1(1-y_t) \geq V_{t-1}^1 \\ V_{t-1}^2 \qquad \text{otherwise} \end{array} \right\} + \frac{1}{2}\max[U^2(1-x_t), V_{t-1}^2].$$

$$V_t^1 = \frac{1}{2}\left\{ \begin{array}{cc} U^1(x_t) \; if \; U^2(1-x_t) \geq V_{t-1}^2 \\ V_{t-1}^1 \qquad \text{otherwise} \end{array} \right\} + \frac{1}{2}\max[U^1(1-y_t), V_{t-1}^1]. \quad (1)$$

We can prove the following lemma:

Lemma 3.1. *Suppose that proposals x and y are given and that both parties use subgame perfect strategies when they accept or reject proposals in the mechanism phase of game G^T. Then:*

- *The proposals y_t and x_t are accepted by the parties 1 and 2 respectively if and only if they verify*

$$\begin{aligned} U^1(1-y_t) &\geq V_{t-1}^1 \\ U^2(1-x_t) &\geq V_{t-1}^2 \end{aligned} \quad (2)$$

  where the V_t^i are defined in (1).
- *Party i's expected payoff when facing lottery $\mathcal{L}_t$ is V_t^i.*

Proof. We prove Lemma 3.1 by induction. If $t = 1$, Party 1 must accept any proposal y_1 such that $U^1(1-y_1) \geq V_0^1 = U^1(0)$ since she gets $U^1(0)$ if she refuses, the same is true for party 2. It follows that in a stage of the mechanism phase of the game where lottery $\mathcal{L}_1$ between x_1 and y_1 is about to be implemented, Party 1's expected payoff is V_1^1. Indeed, with probability $\frac{1}{2}$, y_1 is drawn and he accepts this only if $U^1(1-y_1) \geq V_0^1$. With probability $\frac{1}{2}$, x_1 is drawn and it is accepted if and only if $U^2(1-x_1) \geq V_0^2$. We suppose that the statements are true at $t-1$. Suppose that Party 1 rejects y_t. Then lottery $\mathcal{L}_{t-1}$ is implemented and by the induction hypothesis his expected utility is V_{t-1}^1. Thus he should accept y_t if and only if $U^1(1-y_t) \geq V_{t-1}^1$.

An analogous statement is true for party 2. We then show that in a stage of the mechanism phase where the lottery $\mathcal{L}_t$ between x_t and y_t is about to be implemented, Party 1's expected utility is V_t^1. This lottery draws x_t with probability $\frac{1}{2}$ and x_t is accepted if $U^2(1-x_t) \geq V_{t-1}^2$. Otherwise, x_t is refused and the lottery $\mathcal{L}_{t-1}$ is implemented, in which case Party 1's expected utility is V_{t-1}^1 by the induction hypothesis. With probability $\frac{1}{2}$, y_t is drawn. Player 1 accepts if $U^1(1-y_t) \geq V_{t-1}^1$, otherwise he refuses and has, again by the induction hypothesis, an expected utility V_{t-1}^1. Consequently, when the lottery $\mathcal{L}_t$ is about to be implemented, his expected utility is V_t^1 defined in (1). □

The properties stated in lemma 3.1 can now be used to characterize the proposal strategies of the negotiation phase at equilibrium.

Regarding, the notation used in Lemma 3.2 and in its proof, recall that V_t^1 (and similarly V_t^2) is a function of the history of proposals $h_t = (x_s, y_s)_{s \leq t}$. Note that in what follows, we will use the following equivalent notations for $V_t(h_t)$: $V_t(x, y)$, where $(x, y) = (x_s, y_s)_{s \leq t}$ and $V_t(h_{t-1}, x_t, y_t)$.

Lemma 3.2. *If $x(e) = (x_s(e))_{s \leq t}$ and $y(e) = (y_s(e))_{s \leq t}$ are the proposals in a subgame perfect equilibrium, they verify:*

- $x(e) \in argmax_x V_t^1(x, y(e))$ *and* $y(e) \in argmax_y V_t^2(x(e), y)$
- *for every $s \leq t$, given the history h_{s-1}, $x_s(e) \in argmax_{x_s} V_s^1(x_s, h_{s-1})$, and $y_s(e) \in argmax_{y_s} V_s^2(y_s, h_{s-1})$.*

Proof. Suppose that there is an equilibrium in G^T where $x(e)$ and $y(e)$ are the proposals. It follows from the lemma 3.1 that if the parties use subgame perfect strategies when they accept or reject proposals then the expected equilibrium payoffs, which are the expected payoffs when the lottery $\mathcal{L}_t$ is about to be implemented are $V_t^1(x(e), y(e))$ and $V_t^2(x(e), y(e))$. Depending on the proposals, these payoffs are realized either by immediately accepting the proposals in $\mathcal{L}_t$ or by moving further down the tree. Nash equilibrium requires that $x(e) \in argmax_x V_t^1(x, y(e))$ and $y(e) \in argmax_y V_t^2(x(e), y)$ as

claimed.

To show the second part of lemma 3.2, we note some useful properties that are verified by the functions V_t^i for any $1 \leq s \leq t \leq T$ and that can be inferred from the definition of the functions in (1). First, we observe that for any history h_{s-1}, the proposals x_s that maximize V_s^1 do not depend on y_s. Therefore $argmax_{x_s} V_s^1(h_{s-1}, x_s, y_s) = argmax_{x_s} V_s^1(h_{s-1}, x_s)$ and an analogous statement is true for the proposals y_s that maximize V_s^2. Also, we have

$$\begin{aligned} &argmax_{x_s} V_s^1(h_{s-1}, x_s) \subset argmin_{x_s} V_s^2(h_{s-1}, x_s) \\ &argmax_{y_s} V_s^2(h_{s-1}, y_s) \subset argmin_{y_s} V_s^1(h_{s-1}, y_s). \end{aligned} \tag{3}$$

Finally, we note that $max_{x_s} min_{y_s} V_s^1(h_{s-1}, x_s, y_s)$ is a strictly increasing function of V_{s-1}^1 and that $max_{y_s} min_{x_s} V_s^2(h_{s-1}, x_s, y_s)$ is a strictly increasing function of V_{s-1}^2.

Since $V_t^1(x(e), y(e))$ and $V_t^2(x(e), y(e))$ are the expected payoffs, equilibrium requires that

$$\begin{aligned} x_t(e) &\in argmax_{x_t} V_t^1(h_{t-1}, x_t) \\ y_t(e) &\in argmax_{y_t} V_t^2(h_{t-1}, y_t). \end{aligned}$$

The relations (3) imply that

$$\begin{aligned} V_t^1(e) &= max_{x_t} min_{y_t} V_t^1(h_{t-1}, x_t, y_t) \\ V_t^2(e) &= max_{x_t} min_{y_t} V_t^2(h_{t-1}, x_t, y_t). \end{aligned}$$

Thus, $V_t^i(e)$ is a strictly increasing function of V_{t-1}^i. Therefore, at equilibrium, player i's proposal at $t-1$ must maximize V_{t-1}^i given the history of proposals.

This implies that at date $t-1$:

$$x_{t-1}(e) \in argmax_{x_{t-1}} V_t^1(h_{t-2}, x_{t-1})$$
$$y_{t-1}(e) \in argmax_{y_{t-1}} V_t^2(h_{t-2}, y_{t-1}).$$

By reiterating the arguments above, we can show that the second part of lemma 3.2 holds for every $s \leq t$, where t is the end of the negotiation phase. □

The types of proposals that verify the conditions in lemma 3.2 will depend on the players' attitude to risk. We will focus on the case where at least one player is risk averse. When both parties are risk neutral, there is a very large class of equilibria but all of these induce outcomes that give the players the same level of expected utility.

Proposition 1. *Under the assumption that both parties are risk neutral, the equilibria of G^T are such that*

- $x_1(e) = y_1(e) = 1$
- *for all $t > 1$, $x_t(e) \geq \frac{1}{2}$ and $y_t(e) \geq \frac{1}{2}$*
- *If there is a T such that $x_T + y_T \leq 1$, the game ends in agreement.*
- *If at least one player ceases to make concession at a date t, this triggers the lottery $\mathcal{L}_t$. Any proposition inferior to $1/2$ is rejected and when a proposal is rejected the parties go to the next lottery until the lottery $\mathcal{L}_1$ whose result is accepted.*

Clearly, each risk neutral player can guarantee an expected payoff of $\frac{1}{2}$ by demanding 1 in the first period and by always demanding at least $\frac{1}{2}$ after that. If his offers are not accepted, he can return to the first lottery which gives him 1 with probability $\frac{1}{2}$ and thus an expected utility of at least $\frac{1}{2}$ since he is risk neutral. When parties are risk neutral, the mechanism does not always lead to agreement in the bargaining phase. In this case, the players will be assigned an outcome determined by a lottery but for risk neutral

players this entails no loss of utility compared to an agreement on $(\frac{1}{2}, \frac{1}{2})$.

Let us now assume that at least one party is risk averse in the sense that his utility function is strictly concave. We show that under this assumption, for every $t \geq 1$, the proposals x_t and y_t that maximize the expected utilities V_t^1 and V_t^2 given the history of proposals before t are unique. The concessions are partial and gradual and V_t^1 and V_t^2 are strictly increasing in t. The following proposition characterizes the equilibria in the presence of risk aversion.

Proposition 2. *Under the assumption that at least one party is risk averse, the only subgame perfect equilibrium strategies are such that*

- *The negotiation phase continues until T the maximal end date.*
- *The strategies proposed are: $x_1(e) = y_1(e) = 1$ and for $1 < t \leq T$, each party proposes the unique x_t (y_t) that verifies $U^2(1 - x_t) = V_{t-1}^2$ ($U^1(1 - y_t) = V_{t-1}^1$) where the V_t^i are defined by (1) (at equilibrium, $V_t^2 = U^2(\mathcal{L}_t)$ and $V_t^1 = U^1(\mathcal{L}_t)$ for $t \geq 1$).*
- *In the lottery $\mathcal{L}_t$, Party 1 accepts the proposal y_t if and only if $U^1(1 - y_t) \geq V_{t-1}^1$, and similarly for Party 2.*

Proof. To show this, we note that with risk aversion, we have at the equilibrium $V_t^1 = \frac{U^1(x_t)+U^1(1-y_t)}{2}$ and $V_t^2 = \frac{1}{2}[U^2(y_t) + U^2(1 - x_t)]$. This is true at $t = 1$ and can recursively be verified to hold at later dates. For $t = 1$, it is easy to verify that $x_1 = 1$ ($y_1 = 1$) is the unique maximizer of V_1^1 (V_1^2) defined in 1. For an arbitrary t, assuming Party 1 to be risk averse, $U^1(\frac{1}{2}[x_t + (1 - y_t)]) > V_t^1$. The highest acceptable demand Party 2 can make gives party 1 his certainty equivalent of $\mathcal{L}_t$. We verify that this proposal gives Party 2 a strictly greater utility than V_t^2: Indeed, let y_{t+1} be the proposal such that $U^1(1 - y_{t+1}) = V_t^1$. Since Party 1 is risk averse, $V_t^1 = \frac{U^1(x_t)+U^1(1-y_t)}{2} < U^1(\frac{x_t+1-y_t}{2})$. Thus $1 - y_{t+1} < \frac{x_t+1-y_t}{2}$ or equivalently $y_{t+1} > \frac{y_t+1-x_t}{2}$ and $U^2(y_{t+1}) > V_t^2$. The highest acceptable demand that party 1 can make is x_{t+1} such that $U^2(1 - x_{t+1}) = V_t^2$. Since $V_t^2 = \frac{U^2(y_t)+U^2(1-x_t)}{2} \leq U^2(\frac{y_t+1-x_t}{2})$, $1 - x_{t+1} \leq \frac{y_t+1-x_t}{2}$. Since party 1 is risk

averse, $U^1(x_{t+1}) > V_t^1$. Thus, for both parties, the unique way to maximize V_{t+1}^i is to make the smallest possible acceptable offer, leaving the other party his certainty equivalent of $\mathcal{L}_t$. By making such a proposal, the proposing party ensures himself a strictly higher utility than if he does not make any (or a non-acceptable) proposal. The concessions that are made at each period are partial since $x_{t+1} + y_{t+1} > \frac{y_t+1-x_t}{2} + \frac{x_t+1-y_t}{2} = 1$. It also follows from the above that no player will cease to make concessions, triggering the end of the negotiation phase before T. First, consider the case where both players propose $x_t = x_{t-1}$ and $y_t = y_{t-1}$ at some $t < T$. But both players would have an incentive to propose $x_t = x_t(e)$ instead. This proposition gives him strictly higher utility and sub game perfection requires his partner to accept it. If only Player 1 ceased to make concessions, $x_t = x_{t-1}$, similarly he could switch to a strategy providing higher utility by proposing $x_t = x_t(e)$ instead. By doing so, the negotiation phase would no longer end at t but he can clearly find a strategy for the rest of the game that guarantees him $V_t(x_t(e), h_{t-1}) > V_t(x_t, h_{t-1})$. Thus no player has an incentive to end the negotiation phase before T. . □

The first point of Proposition 2 means that there is a unique subgame perfect equilibrium in G^T, for every $T \geq 1$. In each one of these, the strategies for the proposals that are made before T and the decisions about whether to accept them are identical and have a simple structure : each party initially makes a maximal demand $x_1 = 1$ and $y_1 = 1$. In the following periods, each party demands as much as he can, given that he needs to leave the other party the utility that corresponds to his certainty equivalent of the lottery between the previous proposals, x_{t-1} and y_{t-1}. The equilibrium of game G^S where $S < T$ has the same structure as the equilibrium of G^T and in a sense "truncates" the process at an earlier date.

The sequences of proposals $(x_t)_{t\geq 1}$,$(y_t)_{t\geq 1}$ defined recursively in (**??**) are in fact the sequences defining the Raiffa bargaining process (see Raiffa 1953).

Corollary 1. *(Convergence to the Raiffa solution)If we denote by* $(x(R), y(R))$ *the payoffs of the parties in the Raiffa solution, then*

$$\lim_{t\to\infty} x_t(e) = \lim_{t\to\infty} (1 - y_t(e)) = x(R)$$
$$\lim_{t\to\infty} (1 - x_t(e)) = \lim_{t\to\infty} y_t(e) = y(R).$$

As the number of maximal negotiation rounds in our mechanism grows, the equilibrium shares converge to the Raiffa solution.

Note that the structure of our equilibrium is very different from other implementations of the Raiffa solution proposed by Myerson (1991) and Sjöstrom (1991). Myerson (1991) proposes a noncooperative bargaining game in which, at each round, one of the two players is drawn at random to make a proposal. At the equilibrium, solved by backward induction, at each round, each player proposes for his opponent the certainty equivalent of his expected utility of continuing the game. An agreement is reached at the first round, and this converges to the Raiffa solution when the maximum number of rounds tends to infinity. Sjöstrom (1991) uses the same type of framework but with discounting. In our model, at the equilibrium we do not only observe the Raiffa solution but the whole Raiffa sequence. This holds when the procedure used in the negotiation phase of the game is that of Myerson but also with other bargaining procedures, such as for example the Rubinstein's alternating offers procedure. In the latter case, discounting is not necessary for our result but can be introduced. In that case, as we will see in the following section, we retrieve the Raiffa solution at the limit when the discounting vanishes. Of course, when there is discounting, the gradual concessions observed at the equilibrium generate some inefficiency. We will discuss this point in the following two sections. We will see that this inefficiency can easily be limited and we will show the advantages we obtain in compensation, by correcting the non-robustness of the results obtained in bargaining games that rely on backward induction to be solved.

What is crucial in our model is not the precise description of the process of the bargaining phase but what happens in the mechanism phase which has to be triggered in case of disagreement. The mechanism has as its objective to give the parties an incentive to reach consensus, ideally, without actually having to be triggered. In fact, the mechanism achieves this goal when at least one player is risk averse but not otherwise. Formally, even in the case where the parties are risk averse, complete compatibility of propositions is never reached at any finite date t and a lottery is implemented. However, for risk averse parties, this is an artefact of the model, since the parties' proposals are arbitrarily close. In practice, we could easily do without the lottery by assuming that triggering the mechanism carried a small cost (see the following section). In the case of non-risk averse parties, on the other hand, there is a real need for exterior arbitration since the final proposals may be totally incompatible.

4. The effects of adding a discount factor or an arbitration cost

The mechanism we have presented suggests a way in which parties gradually reach an approximate consensus without assuming discounting. Let us now assume that the parties' utilities are discounted by a factor $\delta^1 \leq 1$ and $\delta^2 \leq 1$ respectively so that the utility derived from implementing the lottery at time t and agreeing on $(x_s, 1-x_s)$ with $s \leq t$ is $(\delta^1)^{t-1} U_t^1(x_s)$ and $(\delta^2)^{t-1} U_t^2(1-x_s)$. When the parties are risk neutral, introducing a discount factor has no effect on the equilibrium strategies since players ended the proposition phase as soon as possible even without discounting.

If at least one player is risk averse, we will see that the presence of a discount factor alters the equilibrium proposals but not the general structure of the equilibrium. The recursive sequence of values that define Party i's expected utility at the stage where $\mathcal{L}_t$ is about to be implemented as a function of $(x_s)_{s\leq t}$ and $(y_s)_{s\leq t}$ is now $V_0^i = U^i(0)$ and for $t \geq 1$:

$$V_t^{i,\delta} = \frac{1}{2}\left\{ \begin{array}{c} (\delta^i)^t U^i(x_t) \; if \; (\delta^j)^t U^j(1-x_t) \geq V_{t-1}^{j,\delta} \\ V_{t-1}^{i,\delta} \qquad \text{otherwise} \end{array} \right\} + \frac{1}{2}\max[\left(\delta^i\right)^t U^i(1-y_t), V_{t-1}^{i,\delta}]. \quad (4)$$

Essentially, everything is the same as in the model without discounting. The equilibrium proposals are determined as before. We can represent graphically how the equilibrium solution is affected by a discount factor. (figure 1).

We summarize the effects of discounting on equilibrium in the proposition below.

Proposition 3. *With discounting, under the assumption that at least one party is risk averse, the only subgame perfect equilibrium strategies are such that:*

- *The negotiation phase continues until $\tilde{T}$, which is either the maximal end date T, or a date at which some player ceases to make concessions because the benefits of doing so is outweighed by discounting.*
- *The strategies proposed are: $x_1(e) = y_1(e) = 1$ and for $1 < t \leq T$, each party proposes the unique x_t (y_t) that verifies $U^2(1-x_t) = V_{t-1}^{2,\delta}$ ($U^1(1-y_t) = V_t^{1,\delta}$) where the $V_t^{i,\delta}$ are defined by (4). (If $\tilde{T} < T$ then the strategies are as above except that $x_{\tilde{T}} = x_{\tilde{T}-1}$ or $y_{\tilde{T}} = y_{\tilde{T}-1}$*
- *In the lottery $\mathcal{L}_t$, Party 1 accepts the proposal y_t if and only if $U^1(1-y_t) \geq V_{t-1}^{1,\delta^1}$, and similarly for Party 2.*

Finally, we note that when the discount rates, which are not necessarily identical, go to zero, we recover the Raiffa solution.

Proposition 4. *Let $x^\delta(e), y^\delta(e)$ denote the equilibrium payoffs of the discounted game and as before $x(R), y(R)$ the payoffs of the parties in the Raiffa solution. Under the assumption that U^1 and U^2 are strictly concave and strictly increasing on $[0,1]$, for every $\epsilon > 0$, there is a $T \in N$ and a $\tau > 0$ such that if there are at least T bargaining rounds and if $\delta^i > 1-\tau$ for $i = 1, 2$*

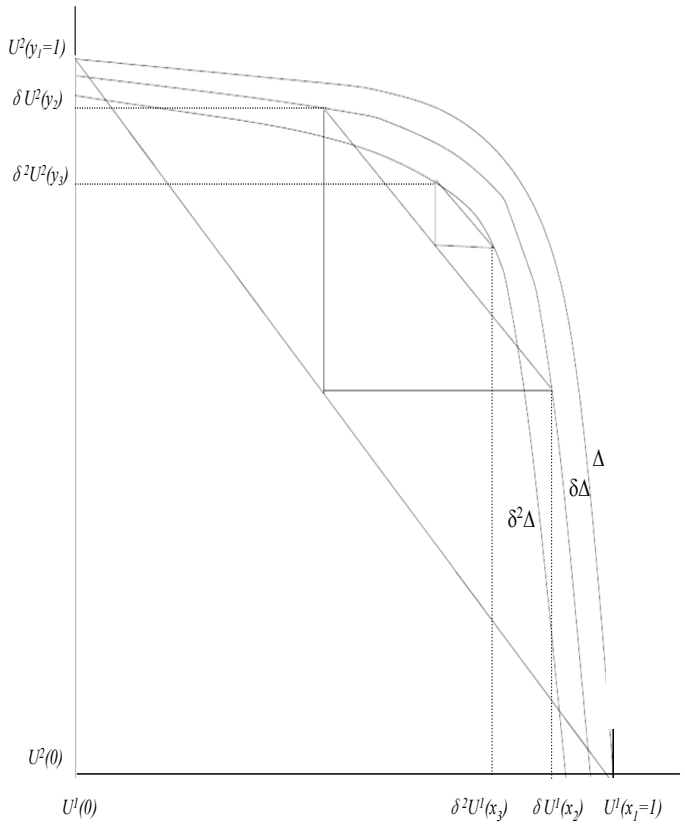


FIGURE 1. *The initial bargaining set* Δ*, and the discounted bargaining sets,* $\delta^1\Delta, \delta^2\Delta, \delta^3\Delta$.... *The last acceptable proposals are made at* $t = 2$*. The black dots show the utility of the last proposals,* $(\delta)^2U^1(x_2)$*,* $(\delta)^2U^2(1-x_2) \in \delta^2\Delta$ *and* $(\delta)^2U^1(1-y_2)$*,* $(\delta)^2U^2(y_2) \in \delta^2\Delta$

then:

$$||(x^{\delta}(e), y^{\delta}(e)) - (x(R), y(R))|| \leq \epsilon$$

Corollary (1) implies that for any fixed $\epsilon > 0$, there is a date T such that in the non-discounted model $||(\frac{U^1(x_T)+U^1(1-y_T)}{2}, \frac{U^2(y_T)+U^2(1-x_T)}{2}) - (x(R), y(R))|| \leq \epsilon$ Let $V_t^{i,\delta}$ denote the value that party i can guarantee in the discounted model. We note that in equilibrium $\frac{U^i(x_T)+U^i(1-y_T)}{2} = V_T^{i,\delta}$. Without detailing the proof of Proposition 4, we note that the hypotheses guarantee that U^1 and U^2 have a continuous inverse so that $(V_t^{1,\delta}, V_t^{2,\delta})$ and (V_t^1, V_t^2) are continuous functions of $\left(V_{t-1}^{1,\delta}, V_{t-1}^{2,\delta}\right)$ and (V_{t-1}^1, V_{t-1}^2) respectively. Since $V_0^i = V_0^{i,\delta}$ it is easily shown that for every $t \leq T$, the difference $|V_t^{i,\delta} - V_t^i|$ can be made arbitrarily small by taking $|V_{t-1}^{i,\delta} - V_{t-1}^i|$ sufficiently small for $i = 1, 2$ and δ^i sufficiently close to 1 for both parties.

Next, we show that introducing a cost C of triggering the mechanism will allow players to reach an agreement without having recourse to it. The second part of the proposition states that the time at which a player's utility from his own proposition and that of his partner differs by at most C, and at which he is thus willing to make a final concession rather than paying the cost, is bounded by a logarithmic function of $1/C$. It is thus a statement about the time required to reach an agreement which shows that even if the cost to trigger the mechanism is very small, players will reach an agreement quickly.

Proposition 5.
- *Let C be a positive cost paid by each player when they have recourse to the mechanism. Then, there is a time $T(C)$ such that an equilibrium exists where both players find an agreement at $T(C)$ and the proposals induced by the equilibrium strategies $(x_t(e))_{t=1}^{t=T(C)}, (y_t(e))_{t=1}^{t=T(C)}$ coincide with those defined in Proposition 2 for $t < T(C)$ and $x_{T(C)} = 1 - y_{T(C)}$ with $x_{T(C)} = x$ for some $x \in [1 - y_{T(C)-1}, x_{T(C)-1}]$.*
- *$T(C) \leq 2 + log_2(\frac{K}{C})$ where K is a constant that depends on the bargaining set.*

Proof. The strategies that induce these propositions coincide with those defined in the case without cost except for the fact that at a certain date $T(C)$ that we will characterize below, the players "agree" on some $x \in [1 - y_{T(C)-1}, x_{T(C)-1}]$ provided that the equilibrium strategies were followed prior to $T(C)$. It can be shown, in the same way as before, that it is neither profitable to deviate from one's strategy for accepting nor from the proposition strategy prior to $T(C)$ since this does not increase the functions V_t^i. We verify that at $T(C)$ it is not profitable to deviate. At equilibrium, there is agreement at $T(C)$. Consider the "worst" case for Player 1 (without loss of generality) where, at equilibrium, he is supposed to make the last concession alone so that $x_{T(C)} = 1 - y_{T(C)-1}(e)$. Suppose that he deviates. It is obvious that any proposition where $x_{T(C)} > x_{T(C)-1}(e)$ would be refused. Therefore the maximal benefit of deviating would be $U_1(x_{T(C)-1}(e)) - U_1(1 - y_{T(C)-1}(e))$ but this benefit will be inferior to the cost C for a sufficiently large end date. Indeed, it is easy to see geometrically that the equilibrium propositions verify the following property: if we consider the Euclidian distance in the utility plane between the points $(U_1(x_T), U_2(1 - x_T))$ and $(U_1(1 - y_T), U_2(y_T))$, it verifies

$$||(U_1(x_T), U_2(1 - x_T)) - (U_1(1 - y_T), U_2(y_T))||_2 \leq \frac{||(U_1(1), U_2(0)) - (U_1(0), U_2(1))||_2}{2^T}.$$

Therefore we can choose $T(C)$ as the smallest t such that

$$t - 1 \geq log_2(\frac{1}{C}||(U_1(1), U_2(0)) - (U_1(0), U_2(1))||_2), \tag{5}$$

which ensures that $max[U_1(x_{T(C)-1}(e)) - U_1(1 - y_{T(C)-1}(e)), U_2(y_{T(C)-1}(e)) - U_2(1 - y_{T(C)-1}(e))] \leq C$. We define the constant K in proposition 5 as $K = ||(U_1(1), U_2(0)) - (U_1(0), U_2(1))||_2)$. □

The results above show that the basic model can satisfactorily incorporate

the introduction of discounting which does not alter the nature of the equilibrium strategies. Moreover, the effect of discounting is continuous in the sense that for sufficiently low discounting, the parties' shares in this equilibrium are close to those of the Raiffa solution. We have also seen that it can be interesting to impose a cost to trigger the mechanism. When the mechanism comes at a small cost, the parties can reach an efficient agreement close to the Raiffa solution without needing to implement any lottery.

5. Robustness with respect to bounded rationality and errors about the number of negotiation rounds

In this section we discuss some properties of our mechanism that could be appealing if the mechanism were actually to be implemented in the presence of parties who are boundedly rational, in the sense that they do not correctly anticipate the number of available negotiation rounds.

This actually stems from a simple property that was stated already in the equilibrium description. If we denote by G^T the game with end date T, the unique equilibrium strategy in games G^T and G^S, with $T < S$ induce the same actions at all dates $t \leq T < S$.

Due to this property the mechanism can perform well even if one or several players misapprehend the number of available rounds, which is something that does not seem unlikely in practice. Suppose some player mistakenly believes that there are fewer rounds than in reality: $T_1 < T$ (it is obvious that if $T_1 > T$, the equilibrium is not altered). Arriving at the date T_1 he necessarily realizes that this is not the end and must change his belief to a $T_2 > T_1$. Possibly he may still not correctly perceive the number of periods so that $T_2 < T$, and so he will change his belief again and so on until the negotiation phase really ends. Both players may make this type of mistake with different errors; in our mechanism, it is easily verified that this has no impact on the equilibrium.

Let T be the actual number of available rounds in the negotiation phase.

Let Player i make a sequence of errors described as above by $(T_k^i)_{k=1}^{k=N}$. This sequence corresponds to the successive beliefs about the number of rounds.

We show recursively that the equilibrium with errors coincides with standard equilibrium until the date n_k, where $(n_k)_k^K$ is an increasing sequence such that $n_K = T$. Define $n_0 = 0$. $h_0 = \emptyset$ coincides with the history in the standard equilibrium.

Define $n_k = min\{n \in N | n > n_{k-1}, n \in (T_k^i)_{k=1}^{k=N} for i = 1 or 2\}$. The induction hypothesis is that both players played the equilibrium proposals until n_{k-1}, so that $h_{n_{k-1}} = \bar{h}_{n_{k-1}}$. Assume that $n_k \in (T_k^i)_{k=1}^{k=N}$. Then Player i plays his equilibrium strategy in G^{n_k} with history $\bar{h}_{n_{k-1}}$. Player $j \neq i$ plays in the game G^t where $t = min\{n \in N | n \geq n_{k-1}, n \in (T_k^j)_{k=1}^{k=N}\}$. He plays the equilibrium strategy in this game given history $\bar{h}_{n_{k-1}}$. The equilibrium strategy at time $t < T$ of a game with end date T is independent of T. These two strategies induce the history $\bar{h}_{n_k}$.

Our mechanism conserves the equilibrium even when used with boundedly rational players who misapprehend the number of available rounds. In comparison, in the Myerson implementation of Raiffa, where the equilibrium propositions are determined by backward induction starting from the last period, knowledge of the exact number of rounds is crucial and mistakes about this number could generate outcomes with absence of consensus. Indeed, this occurs when the player who believes there are fewer remaining rounds is drawn as proposer because his proposals are unacceptable for the player who believes a greater number of rounds remain, leading to a positive probability of a status quo outcome.

6. Conclusion

In the literature on non-cooperative bargaining theory, it has been argued that introducing an outside option as an alternative to the status quo as the outcome, in case of disagreement, can be disruptive for the negotiation, because it generates gradualism or in the worst case no agreement at all

(Compte and Jehiel 2004).

In our paper, the parties who fail to reach an agreement, are committed to follow a specific procedure which, although more complex than resorting to the status quo, leads to an agreement. The novelty of our contribution resides in the choice of the mechanism to be used when there is no agreement in the negotiation phase. The mechanism, which is actually a very simple game, imposes a specific procedure that goes back through the history of proposals of the bargaining phase. This choice of procedure avoids multiplicity of equilibria and the main source of gradualism identified in the analysis of Compte and Jehiel (2004), namely that concessions made by one party only benefit their opponent. For this reason, in our bargaining process, whether the players can opt out, that is, trigger the mechanism at their discretion, or not; does not play a crucial role. The structure of our mechanism rewards even unilateral concessions in the bargaining phase, provided the opponent is rational in the subsequent mechanism phase. A nice property of the equilibrium is that the strategies do not depend on the maximum possible number of bargaining rounds. As a consequence, the mechanism could perform well even in presence of boundedly rational players who do not correctly anticipate the number of negotiation rounds.

Such a mechanism has not been implemented in practice. However it is natural to compare it to ones that are actually in use and with which it shares some features. Arbitration, which similalrly requires from the parties a certain degree of commitment in case of disagreement, is an obvious example. Final Offers Arbitration, in which the arbitrator can only choose between one or the other of the two last incompatible proposals, can be compared to our mechanism. However, there is an important difference since the arbitrator is in our case only a fair lottery and the decision is not binding, leading to the recursivity of our mechanism. The role of risk aversion for bringing par-

ties to an agreement was already identified in the litterature on arbitration. Indeed, Stevens (1966) proposed Final Offer Arbitration in order to correct what was supposed to be a side effect of Conventional Offer Arbitration, in which intermediary solutions can also be imposed: the chilling effect. In the arbitration litterature, this term refers to something similar to gradualism, that is a persistant lack of concessions. Stevens thought that the restricted choice of the arbitrator would allow the mechanism to exploit the parties' risk aversion and to avoid this problem. He appears to have been wrong. It can easily be seen that in the limit case where the arbitrator chooses uniformly at random between the last two proposals, it is even a dominant strategy not to make any concession. Risk aversion, while important, is not in itself sufficient. Our mechanism also involves a fair lottery between proposals, but in addition the recursivity is crucial to exploit the parties' risk aversion.

References

[1] Anbarci, N., Sun, C.-J., 2013. Robustness of intermediate agreements and bargaining solutions. Games and Economic Behavior. 77(1), 367-376.

[2] Binmore, Ken. 1985. "Equilibria in Extensive Games". *The Economic Journal*, 95, Supplement: Conference Papers, 51-59.

[3] Binmore, Ken, Ariel Rubinstein and Asher Wolinsky. 1986. "The Nash Bargaining Solution in Economic Modelling". *RAND Journal of Economics*, 17(2):176-188.

[4] Binmore, Ken, Avner Shaked and John Sutton. 1989. "An Outside Option Experiment". *Quarterly Journal of Economics*, 104(4):753-770.

[5] Brams, Steven.J. 1983. "Equilibrium Strategies for Final Offer Arbitration: there is no median convergence". *Management Science*, 29(8): 927-941.

[6] Compte, Olivier and Philip Jehiel. 2004. "Gradualism in Bargaining and Contribution Games". *The Review of Economic Studies*, 71 (4): 975-1000.

[7] Dickinson, David, L. 2006. "The chilling effect of optimism: The case of final-offer arbitration". *The Journal of Socio-Economics* , 35: 17-30.

[8] Diskin, A., Koppel, M., Samet, D., 2011. Generalized Raiffa solutions. Games and Economic Behavior, 73(2), 452-458.

[9] Esteban, J., Sakovics, J., 2008. A Theory of Agreements in the Shadow of Conflict: The Genesis of Bargaining Power. Theory and Decision. 65(3), 227-252.

[10] Farber, Henry, S. 1980. "An analysis of final-offer arbitration". *The Journal of Conflict Resolution*, 24 (4): 683-705.

[11] Feuille, Peter. 1975. "Final Offer Arbitration and the Chilling Effect". *Industrial Relations*, 14 (3): 302-310.

[12] Kalai, E., Smorodinsky, M., 1975. Other solutions to the Nash bargaining problem. Econometrica 43, 513-518.

[13] Livne, Zvi, A. 1989. "Axiomatic Characterizations of the Raiffa and the Kalai-Smorodinsky Solutions to the Bargaining Problem". *Operations Research*, 37(6): 972-980.

[14] Manzini, P. and Mariotti, M.. 2001, "Perfect Equilibria in a Model of Bargaining with Arbitration", *Games and Economic Behavior* 37, 170-195.

[15] Myerson, Roger. 1991. "Game Theory, Analysis of Conflict", Harvard University Press, Cambridge, Mass.

[16] Nash, John. 1950. "The Bargaining Problem". *Econometrica* ,28:155-152.

[17] Ponsati, Clara and Jozsef Sakovics. 1998. "Rubinstein bargaining with two-sided outside options". *Economic Theory* ,11:667-672.

[18] Raiffa, Howard. 1953. "Arbitration Schemes for Generalized Two-Person Games". In *Contributions to the Theory of Games II.* ed. H.W. Kuhn

and A.W. Tucker. Annals of Mathematical Studies, No 28, Princeton: Princeton University Press, 361-387.

[19] Rubinstein, Ariel. 1982. "Perfect Equilibrium in a Bargaining Model". *Econometrica*, 50, 97-109.

[20] Salonen, H., 1988. Decomposable solutions for N person bargaining games. Europ. J. Polit. Economy. 4, 333-347.

[21] Sjöstrom, Thomas. 1991. "Ståhl's bargaining model". *Economics Letters*, 36: 153-157.

[22] Ståhl Ingolf. 1972. *Bargaining Theory*. Stockholm: Stockholm School of Economics.

[23] Stevens, Carl, M. 1966. "Is Compulsory Arbitration Compatible With Bargaining?" *Industrial Relations*, 5: 38-52.

[24] Trockel, Walter, 2009. "An axiomatization of the sequential Raiffa solution". A manuscript.

[25] Trockel, Walter, 2011. "An exact non-cooperative support for the sequential Raiffa solution". *J. Math. Econ.*, 47(1), 77-83