Loss Aversion and Bargaining *

Jonathan Shalev †

July 23, 1997

Abstract

We consider bargaining situations where two players evaluate outcomes with reference-dependent utility functions, analyzing the effect of differing levels of loss aversion on bargaining outcomes. We find that as with risk aversion, increasing loss aversion for a player leads to worse outcomes for that player in bargaining situations.

An extension of Nash's axioms is used to define a solution for bargaining problems with exogenous reference points. Using this solution concept we endogenize the reference points into the model and find a unique solution giving reference points and outcomes that satisfy two reasonable properties, which we predict would be observed in a steady state.

The resulting solution also emerges in two other approaches, a strategic (non-cooperative) approach using Rubinstein's alternating offers model and a dynamic approach in which we find that even under weak assumptions, outcomes and reference points converge to the steady state solution from any non-equilibrium state.

Keywords: loss aversion, bargaining, reference dependence.

JEL Classification: **C78**.

1 Introduction

Much research has dealt with the connection between risk aversion and Nash's solution to the bargaining problem. In general, Nash's solution predicts that risk aversion is a disadvantage in bargaining over riskless outcomes. Kannai (1977) noted that when bargaining concerns distribution of a divisible commodity between two risk averse individuals, then Nash's solution

*Version 3.11, 19/07/97. This paper is based on a chapter of my Ph.D. thesis at Tel-Aviv University, under the supervision of Dov Samet, to whom I am very grateful for many helpful discussions. A previous version appeared as CORE discussion paper number 9706.

†CORE, 34 voie du Roman Pays, B-1348 Louvain-la-Neuve, Belgium. E-mail: SHALEV@CORE.UCL.AC.BE, Fax: +32-10-474301, Phone: +32-10-478186.

assigns a larger share of the commodity to a bargainer as his utility function becomes less risk averse. Similar results for more general situations can be found in Kihlstrom, Roth and Schmeidler (1981), Roth (1979), and Sobel (1981)[1]. In this paper we analyze bargaining between two *loss* averse individuals. We show that an extension of the Nash solution incorporating loss aversion has a similar characteristic – the solution assigns a larger share of the commodity to a bargainer as her utility function becomes less loss averse.

Loss aversion and reference dependence, as used in Tversky and Kahneman's (1991) prospect theory, refer to the tendency of individuals to attach greater importance to losses (relative to some reference point) than to corresponding gains. Experimental works in both the psychological and the economic literature suggest that people are motivated to minimize loss more than they are motivated to maximize gain (see for example De Dreu, Emans and Van de Vliert (1992), Kahneman and Tversky (1979), Kahneman, Knetsch and Thaler (1990, 1991), Kramer (1989), Taylor (1991), and Tversky and Kaneman (1992)). We feel that these experimental results should be incorporated into relevant areas of theoretical research. As Rabin (1996) points out, "Reference dependence deserves to be, and is gradually becoming, an important part of economic modeling."

A natural extension to Nash's (1950, 1953) classical bargaining model is to incorporate loss aversion, as the psychological elements inherent in loss-aversion play an important part in bargaining. For example, Bazerman, Magliozzi and Neale (1985) showed that in bargaining experiments subjects were more likely to reach agreement when the outcomes were framed as gains than when they were framed as losses. They suggest that because of loss aversion, subjects were more willing to concede a portion of their prospective gains than to lose an asset that they already possess. Similar results were found by Neale and Bazerman (1985) and Neale, Huber and Northcraft (1987). In such experimental situations designed to test framing effects, the reference points are created exogenously and manipulated by the examiner to demonstrate their effects on bargaining. In contrast to this[2], we are interested in predicting reference points endogenously. We regard reference points as representing expectations based on past experience, norms of fairness and social customs. We assume, as Binmore (1994, p.63) notes, that "we seem to have a built-in urge to imitate the behavior of those around us, and the capacity to learn to like what we are accustomed to do." The question of how gain and loss

[1]However, Roth and Rothblum (1982) present a situation with bargaining over risky outcomes where this result does not necessarily hold.

[2]For a comparison of methods used by psychologists and economists, see Hogarth and Reder (1986).

frames are determined in various situations is also addressed by Kahneman (1992), where he surmises that "some of the messages that negotiators exchange are attempts by each side to communicate its reference point and to affect the reference point of the other side."

We analyze two-player bargaining situations, assuming that bargainers evaluate outcomes using reference-dependent utility functions characterized by a level of loss aversion. We first approach the problem with an axiomatic solution concept based upon Nash's (1950) solution[3], which remains one of the most important results in bargaining theory. We start by solving the case where the reference points are given exogenously (Section 2), using an extension of Nash's axioms. In Section 3 we endogenize the reference points into the model, in an attempt to answer the question "where do reference points come from in a bargaining situation?". We suggest that the reference points reflect aspirations and *expectations*[4]. These are formed from previous experiences and from knowledge of outcomes reached by others in similar situations, and they may be *influenced* by the appearance, attitude and behavior of one's bargaining partner. If this is so, then after an encounter where one's expectations were not fulfilled, an individual may approach a similar situation in the future with lower expectations. Conversely, achieving more than one expected may increase expectations for the future. To demonstrate one's opponent's influence, if an opponent radiates self-confidence one may not expect to do as well as in a similar situation against a more timid partner. Since such external characteristics are to some extent under the players' control, it follows that they have some control over the reference point of their bargaining partners. We develop a solution concept using the assumption that reference points[5] should be self-supporting (should fulfill expectations) and stable (neither player can gain by influencing her bargaining partner's reference point). This assumption leads to a unique solution for any bargaining problem extended to include the bargainers' levels of loss aversion. A corollary is that loss aversion is a disadvantage in bargaining situations. Our results are a generalization of Nash's bargaining solution, since when both bargainers have equal levels of loss aversion our solution coincides with the Nash solution.

We complement these results with two other approaches. In Section 4 we add loss aversion to Rubinstein's (1982) alternating offers model. The non-cooperative game corresponding to such a bargaining situation has a unique subgame perfect equilibrium point for any given (com-

[3]The choice of Nash's solution is slightly arbitrary, since although it is the most widely used solution, there exist many other axiomatic solutions. A similar extension of the Kalai-Smorodinsky (1975) solution to include loss aversion gives similar results, as will be noted in Section 3.

[4]The word expectation is used throughout this paper with the psychological connotation denoting anticipation, and not with the statistical meaning.

[5]When the context is clear, we sometimes refer to pairs of reference points as reference points.

mon) discount factor. The limit of these points as the discount factor tends to one is precisely the axiomatic solution for the same underlying bargaining situation with loss aversion. As in the axiomatic approach, this solution is an answer to the question: what reference points might we expect to find in a steady state. In contrast, Section 5 provides an answer to the question: what happens to reference points when the system is not in such a steady state. We combine axiomatic and strategic approaches in a model where axiomatic bargaining is repeated over time, and under reasonable assumptions we find that both the reference points and the outcomes converge over time to the same solution reached in the first two approaches. This result serves as a further justification for the calculations in Section 3 pertaining to a "steady state".

2 The Extended Bargaining Model

As in Nash (1950), the basic elements of a bargaining problem are a set of possible allocations that can be agreed upon, one of which is designated as a default or disagreement outcome. We denote by X the set of possible agreements, and by τ the default outcome. We assume that X is a convex set[6], as is the case in many applications, such as those in which the bargaining is over a divisible commodity.

In the spirit of prospect theory, we assume that outcomes are evaluated by each player with respect to a reference point. We assume also the existence of an *underlying utility function*, $u_i : X \longrightarrow \mathbb{R}$, for each player i, which translates outcomes into real numbers. Such a function can capture the risk aversion aspects of a player's preferences, but not reference dependence, and therefore not loss aversion. The function U_i (defined in the next paragraph) takes as inputs the underlying utility of an outcome and a reference level, and expresses the reference dependence and loss aversion aspects of a player's preferences. It has some important general properties. It is continuous everywhere, differentiable whenever $u_i(\xi) \neq r_i$, and all one-sided derivatives exist at points where $u_i(\xi) = r_i$. The magnitude of loss aversion at any reference level r_i is given by comparing the left and right derivatives of U_i with respect to the first parameter, at the point (r_i, r_i). Our specific choice of U_i has a constant level of loss aversion, λ_i, for each player i, as we are interested in the effects of heterogenous loss aversion on bargaining outcomes.

[6]For any lottery over outcomes, there is a certain outcome in X such that each player is indifferent between the lottery and the certain outcome.

The *utility of player i* with a reference point $r_i \in \mathbb{R}$ from an outcome $\xi \in X$ is given by

$$U_i(\xi, r_i) = \begin{cases} u_i(\xi) & \text{if } u_i(\xi) \geq r_i \\ u_i(\xi) - \lambda_i(r_i - u_i(\xi)) & \text{if } u_i(\xi) < r_i \end{cases} \tag{1}$$

For lotteries over outcomes, if x is a finite-support lottery giving outcomes $\xi_1, \ldots, \xi_s \in X$ with respective probabilities $p_1, \ldots, p_s$, then $U_i(x, r_i) = \sum_{k=1}^{s} p_k U_i(\xi_k, r_i)$. The constant $\lambda_i \in \mathbb{R}_+$ is called player i's *loss-aversion coefficient* and summarizes the loss aversion of player i. The case $\lambda_i = 0$ represents no loss aversion, while higher values of λ_i signify higher levels of loss aversion. The utility function given by (1) is similar to the value function found experimentally by Tversky and Kahneman (1992) for monetary prospects[7]. Note the deliberate distinction between λ_i and u_i, which are assumed fixed, as part of the utility function, and r_i, which is a parameter, depicting one's reference level, or anticipated utility.

Using the underlying utilities of the players we can transform a convex set of available outcomes X into a convex set S consisting of the pairs of underlying utilities of the players for any possible contract. A disagreement outcome τ can similarly be transformed into a pair of utilities d[8]. Thus, given a set X, an outcome τ, and a pair of utility functions U_1, U_2, we can construct an extended bargaining problem (S, d, λ). The set B^* of extended bargaining problems is defined as

$$B^* = \{(S, d, \lambda) | S \subseteq \mathbb{R}^2,\ S \text{ convex},\ d \in S, \lambda \in \mathbb{R}^2\},$$

where S represents the underlying utilities of the outcomes for the players, d represents the underlying utilities of the disagreement outcome, and λ represents the loss-aversion coefficients of the players. This extends the set of Nash bargaining problems, given by $B = \{(S, d) | S \subseteq \mathbb{R}^2,\ S \text{ convex},\ d \in S\}$ to include the loss aversion characteristics of the players. We assume for now that each of the players in an extended bargaining problem has an exogenous reference point, representing her expectations regarding the outcome. These reference points may be based on their experience, their knowledge, and their perception of the present situation. We extend a bargaining problem in B^* with a pair of reference points $r = (r_1, r, 2)$, and construct the set of extended bargaining problems with exogenous reference points: $B^{**} = \{(S, d, \lambda, r) | S \subseteq$

[7] Tversky and Kahneman found that the value function (when the reference point is zero) has the approximate form x^α for $x \geq 0$ and $-\lambda(-x)^\alpha$ for $x < 0$. They found the median values of α and λ to be 0.88 and 2.25 respectively.

[8] As in Nash (1950) we assume throughout that d is dominated by at least one point in S.

$\mathbb{R}^2,\ S$ convex, $d \in S, \lambda \in \mathbb{R}^2, r \in \mathbb{R}^2\}$. Since any element of B^{**} contains both the reference points and the loss-aversion coefficients, it can be transformed to a Nash bargaining problem using formula (1).

Using Nash's axioms, together with a representation axiom to ensure consistency, we derive a solution for B^{**}. We first present the notation used in the analysis. For an element $b = (S, d, \lambda, r) \in B^{**}$, the utility of an individual i from an outcome with underlying utilities $x = (x_1, x_2) \in S$ is given, using (1), by

$$U_i^b(x_i) = \begin{cases} x_i & \text{if } x_i \geq r_i \\ x_i - \lambda_i(r_i - x_i) & \text{if } x_i < r_i \end{cases} \tag{2}$$

The notation for pairs of utilities is given by

$$U^b(x) = (U_1^b(x_1), U_2^b(x_2)), \tag{3}$$

which we extend to sets by defining, for any set $A \subseteq \mathbb{R}^2$, $U^b(A) = \{U^b(x) | x \in A\}$. Since U_i^b can be regarded as a one to one function from $\mathbb{R}$ onto $\mathbb{R}$, an inverse function exists and we denote it by $B_i^b : \mathbb{R} \longrightarrow \mathbb{R}$, where

$$B_i^b(x_i) = \begin{cases} x_i & \text{if } x_i \geq r_i \\ x_i + \frac{\lambda_i}{1+\lambda_i}(r_i - x_i) & \text{if } x_i < r_i \end{cases} \tag{4}$$

The corresponding function for the two players is given by

$$B^b(x) = (B_1^b(x_1), B_2^b(x_2)). \tag{5}$$

From (3) and (5), it is clear that if $b = (S, d, \lambda, r) \in B^{**}$ and $x \in S$, then $B^b(U^b(x)) = x$.

We define a transformation taking an element $b = (S, d, \lambda, r) \in B^{**}$ to a Nash bargaining problem in B (evaluating the utilities with respect to the reference points) by

$$N(b) = (U^b(S), U^b(d)). \tag{6}$$

Points in S which are (weakly) above r for both players are unchanged. Other points $x \in S$ are transformed to points with a lower coordinate for each player i for whom $x_i < r_i$.

The *extended Nash solution* for B^{**} is the natural one given by the following algorithm: Transform the problem using (6), then solve using Nash's solution, and finally calculate the corresponding point in the initial problem according to (5), which is the solution. The rest of this section gives a direct axiomatization of this solution.

We denote our solution function by $\varphi : B^{**} \longrightarrow I\!R^2$ satisfying $\varphi(S,d,\lambda,r) \in S$. The Nash solution is denoted by $\varphi_N : B \longrightarrow I\!R^2$. The following axioms are used to characterize φ. The first four are extensions of Nash's axioms and the fifth ensures consistency when two problems in B^{**} transfer to the same element of B.

Pareto Optimality (PAR): *The solution is not weakly dominated by any point in S except itself.*

Symmetry (SYM): *If S,d,λ and r are symmetrical in the plane, then the solution assigns the same outcome to each player.*

Invariance (INV): *The solution is invariant with respect to a positive linear transformation of S,d and r.*

Independence of Irrelevant Alternatives (IIA): *If the solution of (S,d,λ,r) is x^*, $T \subseteq S$ and $x^* \in T$, then the solution of (T,d,λ,r) is also x^*.*

Representation Invariance (REP): *If two elements of B^{**} both give the same set of utility pairs (using transformation (6)), then the evaluations of the solution points of the two problems give the same utilities to the players. Formally, if $b = (S,d,\lambda,r)$, $b' = (S',d',\lambda',r')$ and $N(b) = N(b')$ then $U^b(\varphi(b)) = U^{b'}(\varphi(b'))$.*

Theorem 2.1 *There exists a unique solution function $\varphi : B^{**} \longrightarrow I\!R^2$ satisfying $\varphi(S,d,\lambda,r) \in S$ and the five axioms, and it is given by (for $b = (S,d,\lambda,r)$)*

$$\varphi(b) =$$
$$B^b\left(argmax_{x\in S}(U_1^b(x_1) - U_1^b(d_1))(U_2^b(x_2) - U_2^b(d_2))\right) =$$
$$B^b\left(\varphi_N(N(b))\right).$$

All non-trivial proofs are given in the appendix.

Note that the solution given by Theorem 2.1 for problems in B^{**}, is just a first step to obtaining a solution for problems in B^{*}. It is no more than a straightforward application of Nash's solution, assuming exogenous reference points and a specific form of reference dependence and loss aversion. The fact that the reference points are usually *not* given exogenously is the motivation for the next three sections.

3 Endogenization of Reference Points

The question addressed in this section is "What is a *suitable* reference point for an extended bargaining problem?" We investigate two criteria for an answer to this question. The first criterion is that the reference point pair should also be the solution to the extended bargaining problem. Such a reference point will be called *self-supporting*. The set of self-supporting reference points for an element of B^{*} is always non-empty, and is a closed segment of the Pareto frontier of the outcome set (Theorem 3.1). The second criterion is the *stability* of the reference point. The notion of stability refers to the assumption that a player can (by her behavior, appearance, remarks about her own reference point, etc.) affect the reference point of her opponent. A reference point is not stable if either player prefers the (axiomatic) solution of a problem differing only in her opponent's reference point to the solution of the original problem. The set of stable self-supporting reference points for any extended bargaining problem contains exactly one point (Theorem 3.2). Corollories of the two theorems in this section show that loss aversion is a disadvantage in bargaining.

3.1 The Psychology of Reference Points

An important assumption we make about reference points is that one can manipulate the reference point of one's bargaining partner, but that one cannot do so to one's own reference point. This is akin to assuming that one might get a higher grade by cheating successfully on an exam, but this would affect the perception of one's knowledge or ability only in the eyes of the examiner. One way of manipulating the opponent's reference point is by stating (or misstating) one's own reference point (see Kahneman (1992)). This may affect the opponent's reference point, but does not affect one's own, as self deception is not as simple as deceiving others. [9]

[9] This may seem to go against the notion of complete information that is implicitly assumed. A partial justification is that at the solution, neither player chooses to change the other's reference point, and no strategic behavior is necessary. To adapt the model to incorporate incomplete information would introduce complications that would obscure the main points.

3.2 Self-Supporting Outcomes

When two players approach a bargaining problem, each comes with certain expectations. These may be realistic or exaggerated. The pair of expectations (reference points) could be compatible or they might not be satisfiable by any feasible contract. Each player might also have some idea about what she expects the other player's reference point to be. If the reference point pair of an element of B^{**} is equal to the extended Nash solution, it is called a *self-supporting* outcome (or self-supporting reference point) of the corresponding element of B^*. Formally, for any $(S,d,\lambda) \in B^*$ we define the set of self-supporting outcomes of (S,d,λ) by

$$\text{Self}(S,d,\lambda) = \{x \in S | \varphi(S,d,\lambda,x) = x\}.$$

The following theorem characterizes the set of self-supporting outcomes for any given extended bargaining problem.

Theorem 3.1 *For any extended bargaining problem (S,d,λ), the set Self(S,d,λ) is a closed non-empty segment of the Pareto frontier of S (The Pareto frontier of S is denoted by Par(S)). Furthermore,*

$$\begin{aligned} &Self(S,d,\lambda) = \\ &= \Big\{ x \in Par(S) \mid \frac{(x_2-d_2)}{(x_1-d_1)} \frac{1}{(1+\lambda_1)} \leq -a \leq \frac{(x_2-d_2)}{(x_1-d_1)}(1+\lambda_2) \\ &\qquad \textit{for some } a < 0 \textit{ such that the line through } x \textit{ with slope } a \textit{ is tangent to } S \Big\} \end{aligned}$$

Figure 1 gives a graphical example of such a set.

Corollary 3.1 *If $b = (S,d,(\lambda_1,\lambda_2))$ and $b' = (S,d,(\lambda_1',\lambda_2))$ are elements of B^*, and $\lambda_1' \geq \lambda_1$, then any self-supporting outcome of b is a self-supporting outcome of b', and all points that are self-supporting outcomes of b' and not of b are worse for player 1 (and better for player 2) than any self-supporting outcome of b.*

Corollary 3.2 *The Nash solution of $(S,d) \in B$ is a self-supporting outcome of $(S,d,\lambda) \in B^*$ for any $\lambda \in \mathbb{R}^2_+$.*

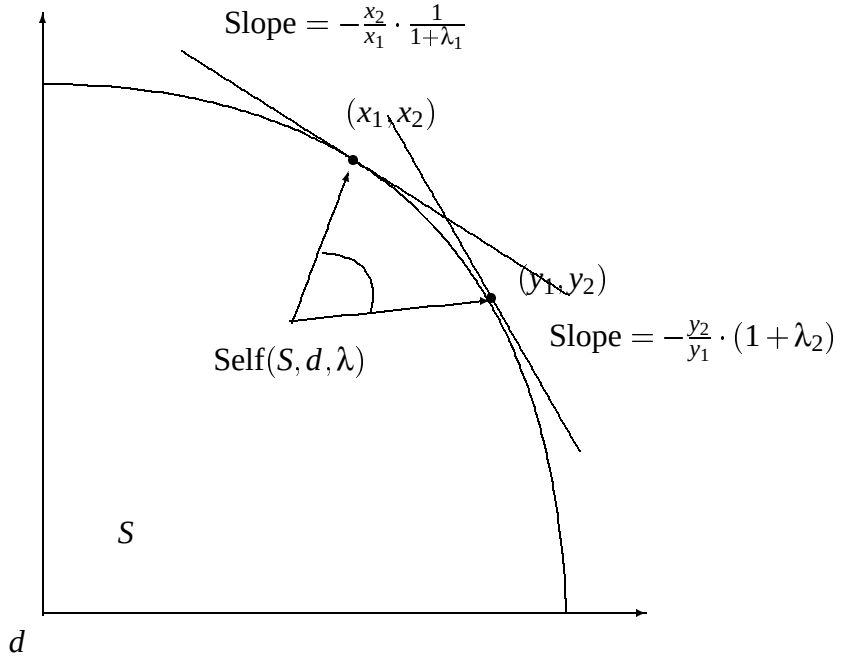

Figure 1: **Example - set of self-supporting reference points.**

Corollary 3.2 follows from the theorem, using Lemma 6.1 on Page 22. Corollary 3.1 shows that for any extended bargaining problem (S, d, λ), increasing λ_i leads to a (weakly) worse set of results for player i that are self-supporting outcomes.

Remark: When the axiomatic solution concept is the Kalai-Smorodinsky (1975) solution, there exists a unique self-supporting outcome pair. If a player becomes more loss averse, the self-supporting outcome is (weakly) worse for her. If neither player is loss averse, the self-supporting outcome is equal to the Kalai-Smorodinsky solution.

3.3 Stable Reference Point Pairs

As mentioned previously, each player has an effect on the formation and the value of the other player's reference point. The concept of stability stems from the notion that a player's appearance and behavior can affect the other player's reference point. If the solution of $b \in B^{**}$ is (weakly) preferred by each player to the solution of any element of B^{**} differing only in her opponent's reference point, we call the reference point of b a *stable* point of the corresponding element of B^{*}. If a reference point is stable, neither player can improve her own outcome by changing her opponent's reference point. Given the assumption of a player's ability to affect her

opponent's reference point, this is a natural form of equilibrium. The definition of stability does not limit the change to the opponent's reference point, but from a self-supporting outcome that is not stable, one of the players can improve her outcome by changing her opponents reference point even by an infinitesimal amount[10]. Formally, for an extended bargaining problem (S,d,λ) we define the set

$$Stab(S,d,\lambda) = \{x \in S \mid \varphi_i(S,d,\lambda,x) \geq \varphi_i(S,d,\lambda,(x_i,x'_{-i})), \forall x'_{-i} \in I\!R,\ i = 1,2\}$$

where the subscript $-i$ refers to player $3-i$, player i's bargaining partner. The following theorem characterizes the set of stable and self-supporting outcomes[11] (which is a single point) for any element of B^*.

Theorem 3.2 *For any extended bargaining problem (S,d,λ), the set of stable self-supporting reference point pairs contains exactly one element, which is given by*

$$\begin{aligned} &Stab(S,d,\lambda) \cap Self(S,d,\lambda) = \\ &= \left\{x \in Par(S) \mid \exists \text{ tangent to } S \text{ through } x \text{ with slope } -\frac{(x_2-d_2)}{(x_1-d_1)}\frac{(1+\lambda_2)}{(1+\lambda_1)}\right\} \end{aligned} \tag{7}$$

Figure 2 gives a graphical example of such a set.

Corollary 3.3 *If $b = (S,d,(\lambda_1,\lambda_2))$ and $b' = (S,d,(\lambda'_1,\lambda_2))$ are elements of B^*, and $\lambda'_1 \geq \lambda_1$, then the stable self-supporting outcome of b' is (weakly) worse for player 1 (and better for player 2) than that of b.*

Corollary 3.4 *The stable self-supporting outcome of any $(S,d,\lambda) \in B^*$ with $\lambda_1 = \lambda_2$ is the Nash solution of (S,d).*

Corollary 3.3 shows that for any extended bargaining problem (S,d,λ), increasing λ_i leads to a (weakly) worse result for player i in the stable self-supporting outcome. The outcome is strictly worse whenever the Pareto frontier of S is smooth and λ_i strictly increases. An intuitive reason for this phenomenon is that with a reference point on the Pareto frontier of S, as a player

[10] This can be shown using Lemma 5.3.

[11] We call a stable reference point that is a self-supporting outcome a stable self-supporting outcome or a stable self-supporting reference point.

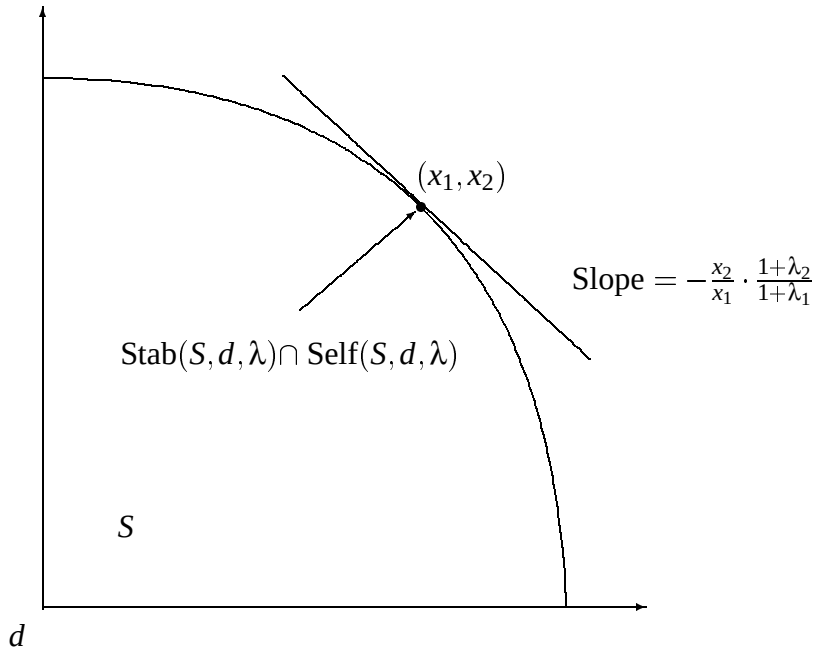

Figure 2: **Example - set of stable self-supporting reference point pairs.**

becomes more loss averse, the disagreement point becomes more unattractive. Thus she is less willing to risk breakdown of the bargaining and will therefore concede more to reduce the risk of receiving the disagreement outcome. From Corollary 3.4, for (S,d,λ) with $\lambda_1 = \lambda_2$, the stable self-supporting outcome coincides with the Nash solution of (S,d). Thus, the stable and self-supporting outcome is a generalization of the Nash bargaining solution.

Remark: When the axiomatic solution concept is the Kalai-Smorodinsky (1975) solution, if either of the players is loss averse and the bargaining problem is not trivial, then there is no stable reference-point pair.

4 The Strategic Approach

In Rubinstein's (1982) seminal paper, a model is presented where two players have to reach an agreement on the partition of a pie of size one. The procedure consists of the players making alternating offers, and the pie is divided if an offer is accepted. Rubinstein characterizes the (subgame-)perfect equilibrium partitions (PEP) for various classes of preference relations.

We generalize the pie from Rubinstein's model, similarly to Osborne and Rubinstein (1990, pp.73-76), by allowing a general convex set of outcomes in $\mathbb{R}^2_+$. Given an extended bargaining

problem $b = (S, (0,0), \lambda)$, we construct a game form identical to that of Rubinstein's alternating offers game, with offers being outcomes on the Pareto frontier of S. Acceptance of a point $x = (x_1, x_2) \in Par(S)$ at period t gives an outcome denoted (x,t). An infinite stream of rejections leads to the outcome $((0,0),\infty)$. The evaluation of outcomes by the players uses a form of time preference modified to include the players' level of loss-aversion.

To incorporate reference dependence and loss aversion into this model, a number of important questions need to be addressed. What are the reference points? How should they change during the bargaining process? What is the final utility? How should it depend on the history of offers and counteroffers? A reasonable assumption to make is that the utility should depend on the final offer and a reference point. A possible reference point could be *the highest offer made to the player during the bargaining procedure*. Surely, rejecting an offer of x and later accepting a lower offer would give a feeling of loss. Using such a reference point causes the game to be non-stationary, which makes it extremely difficult to analyze and find the subgame perfect equilibria. We therefore chose a different approach, even though the motivation is less appealing.

We take the notion of time preference, and modify it to encompass two separate items: objective discounting and loss-aversion. The objective discounting describes how the feasible outcomes change over time. We assume that as each time period passes, each feasible outcome shrinks by a factor of δ ($\delta < 1$). This discount factor is common to the two players. For example, it could signify the interest rate on money, or the reduced desirability of consumption goods as they get older. Loss-aversion describes how each player suffers additional disutility as time passes from the fact that agreement was not reached at the previous period, and the value of the outcomes diminishes. We assume loss-aversion evaluation (as given by Equation (2)), where the reference point is the value of the outcome in the previous period, and the realized outcome is the value in the present period (if agreement is reached at the first period, no losses are entailed, either by objective discounting or loss-aversion.) Thus, we assume that loss aversion is equivalent to higher impatience. This is a plausible assumption, as higher loss aversion makes a player more eager to reach agreement earlier, as she is more sensitive to the (objective) shrinking of the pie. Formally, we have the following recursive function describing how player i evaluates an outcome $((x_1,x_2),t)$.

$$V_i(x,t) = \begin{cases} \delta V_i(x,t-1) - \lambda_i(V_i(x,t-1) - \delta V_i(x,t-1)) & \text{if } t > 1 \\ x_i & \text{if } t = 1 \end{cases}$$

where λ_i is player i's loss-aversion coefficient.

Expanding the recursion, we have the following formula:

$$V_i(x,t) = x_i(\delta + \delta\lambda_i - \lambda_i)^{t-1} \tag{8}$$

which is equivalent to Rubinstein's sub-family of preferences with fixed discount factors, where player i's discount factor δ_i is given here by

$$\delta_i = \delta + \delta\lambda_i - \lambda_i, \quad \text{for } i \in \{1,2\}. \tag{9}$$

For the simple case where $S = \{(x_1,x_2)|x_1 + x_2 \leq 1\}$, there exists a unique PEP, giving player 1 the portion $\frac{1+\lambda_2}{1+\delta(1+\lambda_1+\lambda_2)+(\delta-1)(\lambda_1\lambda_2)}$, agreement being reached at the first period. If we take the limit as δ tends to 1, the unique PEP gives player 1 the portion $\frac{1+\lambda_2}{2+\lambda_1+\lambda_2}$. For the corresponding extended bargaining problem $(S,(0,0),\lambda)$ this point is exactly the stable self-supporting outcome derived in Section 3.3. We now show that this is no coincidence, and that this equivalence occurs for any extended bargaining problem $(S,(0,0),\lambda)$.

Using the method of Rubinstein (1982), we derive the unique subgame perfect equilibrium outcome for the case of a general S (assuming without loss of generality that $d = (0,0)$), with the evaluation of outcomes using time preference and loss-aversion. For any δ such that $\frac{\lambda_i}{1+\lambda_i} < \delta < 1$ for each player i[12]. , define the players' evaluations of outcomes according to (8). Denote the alternating offers game with these utility evalutions by G_δ. Since S is a convex set, the Pareto frontier of S can be described by a continuous, strictly monotonic decreasing, concave function f, such that $x_1 = f(x_2)$ iff (x_1,x_2) is on the Pareto frontier of S. Adapting Rubinstein's method to our set of feasible outcomes, the set of subgame perfect equilibrium outcomes of G_δ is determined by the following set:

$$\Delta_\delta = \left\{ (x_\delta, y_\delta) | \begin{array}{l} y_\delta = \delta_1 x_\delta \\ f(x_\delta) = \delta_2 f(y_\delta) \end{array} \right\}$$

For any element $(x_\delta, y_\delta) \in \Delta_\delta$, x_δ signifies an outcome that player 1 can obtain in a subgame perfect equilibrium of G_δ when she makes the first offer (and y_δ is an outcome that player 1

[12]Since for $\delta \leq \frac{\lambda_i}{1+\lambda_i}$ the player would only care about the first-period payoff, and we are interested in the limit as $\delta \longrightarrow 1$, we assume without loss of generality that $\delta > \frac{\lambda_i}{1+\lambda_i}$ for each player i.

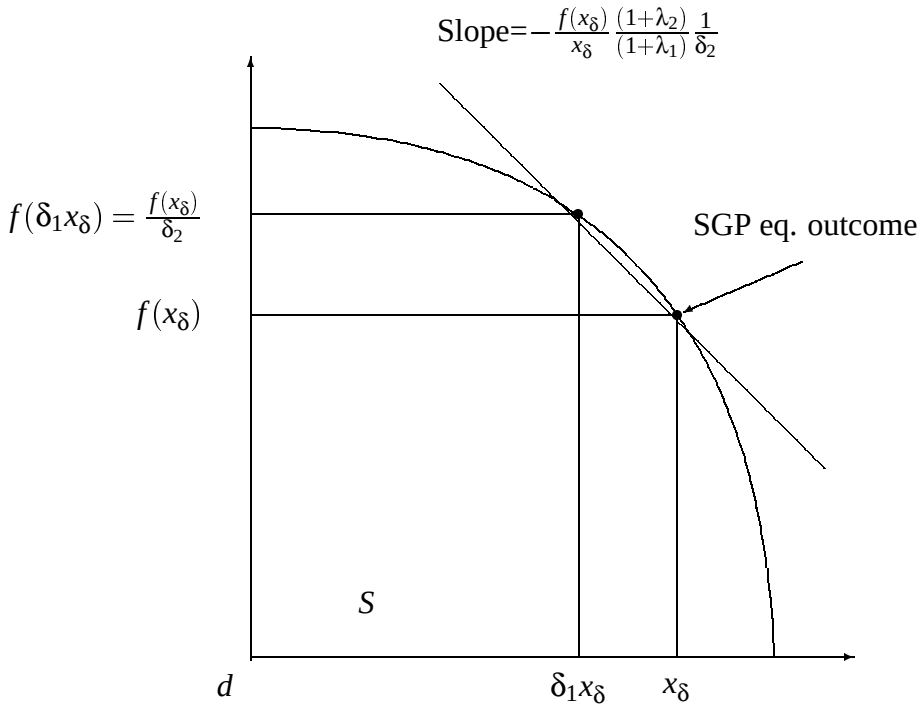

Figure 3: **The point $(x_\delta, f(x_\delta))$, the subgame perfect equilibrium outcome of G_δ, solves the equation $f(x_\delta)=\delta_2 f(\delta_1 x_\delta)$.**

obtains in a subgame perfect equilibrium of G_δ when player 2 makes the first offer). Since for our method of evaluating payoffs, for any extended bargaining problem and any δ, this set is always a singleton, it follows that the unique subgame perfect equilibrium outcome of G_δ is at the point $(x_\delta, f(x_\delta))$ that satisfies

$$f(x_\delta)=\delta_2 f(\delta_1 x_\delta), \tag{10}$$

where δ_1 and δ_2 are given by Equation (9). A graphical example of this is given in Figure 3.

It is simple to show that $x^* = lim_{\delta\to 1} x_\delta$ exists, and that this is the limit of the subgame perfect equilibrium outcomes of G_δ as δ tends to one. Now, for any $\delta < 1$,

$$\frac{f(x_\delta(\delta+\delta\lambda_1-\lambda_1))-f(x_\delta)}{x_\delta(\delta+\delta\lambda_1-\lambda_1)-x_\delta} = \frac{f(x_\delta)\frac{1}{\delta+\delta\lambda_2-\lambda_2}-f(x_\delta)}{x_\delta(\delta+\delta\lambda_1-\lambda_1-1)} \quad (11)$$

$$= \frac{f(x_\delta)(1-\delta)(1+\lambda_2)}{x_\delta(\delta-1)(1+\lambda_1)(\delta+\delta\lambda_2-\lambda_2)} \quad (12)$$

$$= -\frac{f(x_\delta)}{x_\delta}\frac{(1+\lambda_2)}{(1+\lambda_1)}\frac{1}{(\delta+\delta\lambda_2-\lambda_2)} \quad (13)$$

where we use equation (10) to derive equation (11). Thus, taking the limit as δ tends to one, we have that

$$lim_{\delta\to 1}\frac{f(x_\delta)(\delta+\delta\lambda_1-\lambda_1)-f(x_\delta)}{x_\delta(\delta+\delta\lambda_1-\lambda_1)-x_\delta} = -\frac{f(x^*)}{x^*}\frac{(1+\lambda_2)}{(1+\lambda_1)}.$$

Therefore, the line through $(x^*, f(x^*))$ with slope $-\frac{f(x^*)}{x^*}\frac{(1+\lambda_2)}{(1+\lambda_1)}$ is tangent to S. Hence, from Theorem 3.2, x^* is the stable self-supporting solution. To summarize, we have the following theorem.

Theorem 4.1 *If G_δ is a Rubinstein alternating offers game derived from $b = (S,(0,0),\lambda) \in B^*$ using the utility evaluation of Equation (8), and $1 > \delta > \frac{\lambda_i}{1+\lambda_i}$, $i = 1,2$, then*

$$\lim_{\delta\to 1} SGP(G_\delta) = Sel\, f(b) \cap Stab(b)$$

where $SGP(G_\delta)$ is the (unique) subgame perfect equilibrium of the game G_δ.

The result of this section is that the unique subgame perfect equilibrium outcome to the alternating offers game when objective discounting is negligible, is equal to the unique stable self-supporting outcome derived in Section 3. As in Nash (1950, 1953), both the strategic and the axiomatic methods lead to the same solution.

5 Evolution of Reference Points

The concept of stable self-supporting reference points that was examined in the previous sections was a static concept. We did not treat the question of how these reference points come about. To understand how such steady-states evolve, it is necessary to examine states that are not in equilibrium, where the reference points are not self-supporting or not stable. In this section we approach the question of how stable, self-supporting reference points are reached in a dynamic model. We do so with a non-cooperative (multi-period) dynamic game.

The game is denoted by $\Gamma(b, r^0, \delta)$ and is characterized by its three parameters as follows: $b = (S, d, \lambda) \in B^*$ is an extended bargaining problem. $r^0 \in I\!R^2$ is an initial reference point. The number $0 < \delta < 1$ is the discount factor used by the players to aggregate their per-period payoffs.

In the game $\Gamma(b, r^0, \delta)$ the extended bargaining problem b is solved at each period $t = \{0, 1, 2, ...\}$ (with different reference points) according to the extended Nash solution for elements of B^{**} (derived in Section 2.) The reference point for period $t = 0$ is given exogenously by r^0. No choices are made by the players at this period, and the outcome for the first round is thus given by $x^0 = \varphi(S, d, \lambda, r^0)$. The reference point changes from period to period according to the history of the game and the actions of the players. At period $t \geq 1$, the set of actions available to each player is $A^t = [-\frac{1}{t}, \frac{1}{t}]$. If the action chosen at period t by each player i is $\alpha_i^t \in A^t$ then the reference points used to solve the bargaining problem for period t are $r_i^t = x_i^{t-1} + \alpha_{-i}^t$. Thus, the reference point for each player is determined by her outcome in the previous period[13] (experience) and the action of her opponent (influence). The action sets shrink as time passes, reflecting the fact that as players get to know more about each other they can less easily influence their opponents. The stage payoff to player i at stage $t \geq 1$ is $x_i^t = \varphi_i(S, d, \lambda, r^t)$. The total payoff to player i is the discounted sum of her stage payoffs, i.e. $\sum_{t=0}^{\infty} \delta^t x_i^t$. Our main result (Theorem 5.1) is that if the players use undominated strategies in $\Gamma(b, r^0, \delta)$, then the sequence of outcomes converges to the stable self-supporting outcome of the extended bargaining problem b.

We first present a number of properties of the stage game that we use to show our convergence results. The first result is the interesting property that the extended Nash solution for any

[13] The same results would hold if we took a weighted average of a number of past periods, with most weight on the latest periods.

$(S,d,\lambda,r) \in B^{**}$ is a self-supporting outcome of (S,d,λ).

As in Section 4 we regard the Pareto frontier of S as a function f, and for notational simplicity we assume that f is differentiable (the results hold even if f is not differentiable, but the notation in the proof is more cumbersome).

Lemma 5.1 *The extended Nash solution of any extended bargaining problem with a reference point $(S,d,\lambda,r) \in B^{**}$ is a self-supporting outcome of the corresponding extended bargaining problem (S,d,λ), i.e.*

$$\varphi(S,d,\lambda,r) \in \mathit{Self}(S,d,\lambda).$$

The next lemma shows that for any extended bargaining problem with reference points, if each player's reference point is part of a self-supporting pair of reference points, and the reference point pair is a feasible point, then neither player gets less than her reference point at the extended Nash solution.

Lemma 5.2 *For any $(S,d,\lambda) \in B^*$, if $r \in S$, and both $(r_1, f(r_1))$ and $(f^{-1}(r_2), r_2)$ are self-supporting reference points of (S,d,λ), then $\varphi_i(S,d,\lambda,r) \geq r_i$ for $i \in \{1,2\}$.*

The third and final lemma of this section deals with extended bargaining problems with a reference point pair, where the reference point pair is feasible and not strictly dominated by the stable self-supporting reference point pair. In this case one player's reference point may be above (her element of) the stable self-supporting outcome, while the other player's reference point is below. The lemma states that if this is the case, then the extended Nash solution for this bargaining problem gives the player with the reference point higher than the stable self-supporting outcome not more than the value of her reference point. Thus, from Lemma 5.2, if both the reference points are part of self-supporting reference point pairs, the solution gives the player whose reference point is above the stable self-supporting reference point exactly her reference point (and the other gets at least her reference point). See Figure 4 for a graphical example of this case.

Lemma 5.3 *Fix $(S,d,\lambda) \in B^*$ and let x^* denote the stable self-supporting outcome of (S,d,λ). If $r \in S$ and for some $i \in \{1,2\}$ both $r_i > x_i^*$ and $r_{-i} < x_{-i}^*$, then $\varphi_i(S,d,\lambda,r) \leq r_i$.*

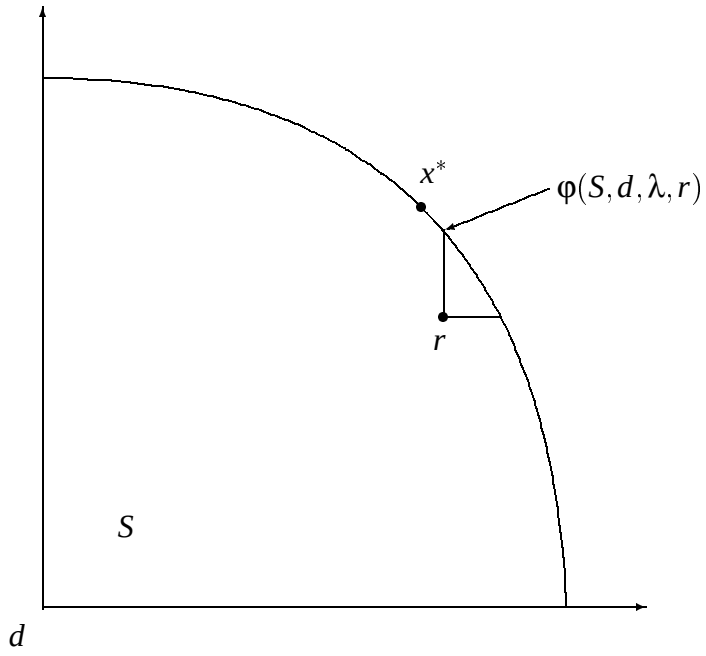

Figure 4: **If x^* does not dominate r, the solution is no better than r for the player i s.t. $r_i > x_i^*$.**

The main theorem of this section is a convergence result – if players use undominated strategies in the dynamic bargaining game then the outcomes (and the reference points) converge to the stable self-supporting outcome of the underlying extended bargaining problem.

Fix a repeated bargaining game $\Gamma(b, r^0, \delta)$ with $b = (S, d, \lambda)$, and denote the stable self-supporting outcome of (S, d, λ) by x^*.

For $t > 1$ define

$$C^t = \{x \in Par(S) | 0 \leq x_i - x_i^* \leq \frac{1}{t} \ for \ i = 1 \ or \ i = 2\}$$

$$D^t = \{x \in Par(S) | \exists y \in C^t \ \ s.t. \ \ 0 \leq y_i - x_i \leq \frac{1}{t} \ for \ i = 1 \ or \ i = 2\}$$

See Figure 5 for a graphical example of these sets.

Lemma 5.4 *For $\Gamma(b, r^0, \delta)$, for all $t \geq 1$,*

1. $x^t \in C^t \Longrightarrow x^{t+1} \in D^t$.

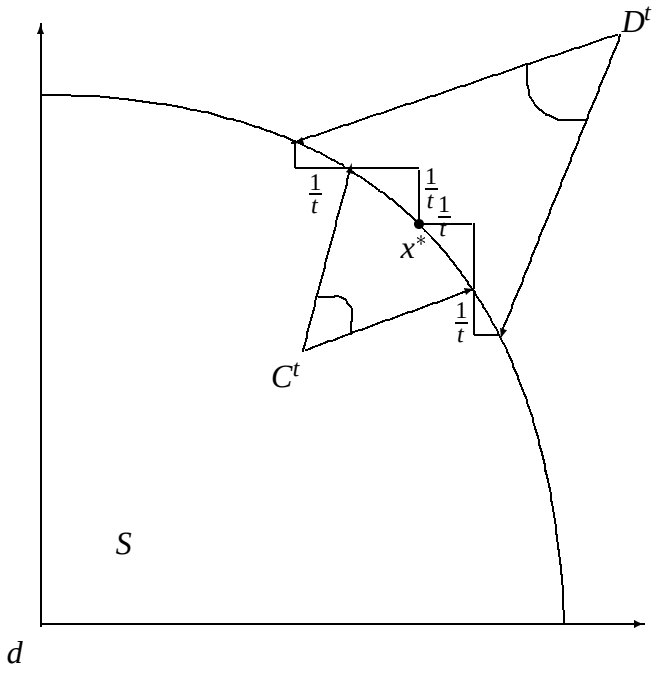

Figure 5: **A graphical example of the sets C^t and D^t.**

2. $x^t \notin C^t \Longrightarrow |x^{t+1} - x^*| \leq |x^t - x^*|$.

Proof: Immediate from Lemmas 5.1, 5.2 and 5.3.

Theorem 5.1 *In any repeated bargaining game $\Gamma(b, r^0, \delta)$, if the players play undominated pure strategies, then $x^t \to x^*$, where x^* is the stable self-supporting outcome of b.*

Thus, even under very weak assumptions on the strategies of the players (not using dominated strategies), we see that when bargaining situations are repeated, the reference points of the players will converge to the stable self-supporting reference point pair that exists in a "steady state" equilibrium.

6 Appendix: Proofs of the Theorems

Proof of Theorem 2.1:

1. For $b = (S, d, \lambda, r) \in B^{**}$ define $\varphi(b) = B^b(\varphi_N(N(b)))$.

2. Take $b=(S,d,\lambda,r), b'=(S',d',\lambda',r') \in B^{**}$. If $N(b)=N(b')=(S'',d'')$ then

$$U^b(\varphi(b)) = U^b(B^b(\varphi_N(S'',d''))) = \varphi_N(S'',d'')$$

and

$$U^{b'}(\varphi(b')) = U^{b'}(B^{b'}(\varphi_N(S'',d''))) = \varphi_N(S'',d'')$$

Therefore, REP is satisfied by φ.

3. SYM is satisfied by φ, since U^b (and therefore N) and B^b preserve symmetry, and φ_N (the Nash solution in B) of a symmetrical bargaining problem is symmetric.

4. PAR is satisfied by φ, since for any $b=(S,d,\lambda,r) \in B^{**}$, U^b and B^b preserve the Pareto frontier (i.e. if x is on the Pareto frontier of S, then $U^b(x)$ is on the Pareto frontier of $U^b(S)$), and the Nash solution to $N(b)$ is a Pareto optimal point in $U^b(S)$.

5. IIA also holds for φ, since both U^b and B^b are one to one functions, and φ_N satisfies IIA in B.

6. Take $\alpha_1,\beta_1,\alpha_2,\beta_2 \in \mathbb{R}$ with $\alpha_1,\alpha_2 > 0$. Define $t : \mathbb{R}^2 \longrightarrow \mathbb{R}^2$, a positive linear transformation by $t(x_1,x_2) = (\alpha_1 x_1 + \beta_1, \alpha_2 x_2 + \beta_2)$. Define t accordingly for sets, and define for $b=(S,d,\lambda,r) \in B^{**}$, t(N(b))=(t(U(S)),t(U(d))) It is simple to show that

   (a) $U^{t(b)}(t(x)) = t(U^b(x))$ for any $x \in \mathbb{R}^2, b \in B^{**}$.

   (b) $B^{t(b)}(t(x)) = t(B^b(x))$ for any $x \in \mathbb{R}^2, b \in B^{**}$.

   Therefore, since φ_N satisfies INV on B,

$$\begin{aligned}
&\varphi(t(b)) = \\
&= B^{t(b)}(\varphi_N(N(t(b)))) = \\
&= B^{t(b)}(\varphi_N(t(N(b)))) = \\
&= B^{t(b)}(t(\varphi_N(N(b)))) = \\
&= t(B^b(\varphi_N(N(b)))) = \\
&= t(\varphi(b))
\end{aligned}$$

   Thus φ satisfies INV.

7. Let φ' be a different solution satisfying the five axioms. Since φ' is different from the solution φ defined in this proof, there exists an element $b \in B^{**}$ such that $x^* = \varphi'(b) \neq \varphi(b)$. Since for any $b=(S,d,\lambda,r)$, U^b is a one to one transformation,

$$U^b(x^*) \neq U^b(B^b(\varphi_N(N(b)))) = \varphi_N(N(b)) \tag{14}$$

   Denote $b^0 = (N(b),(0,0),(0,0)) \in B^{**}$. From REP, since $N(b)=N(b^0)$, it is true that $U^b(\varphi'(b)) = U^{b^0}(\varphi'(b^0)) = \varphi_N(N(b))$, where the last equality is from the first four axioms on the subset of B^{**} where $\lambda = r = (0,0)$, since this subset is isomorphic to B, where the four axioms give us the Nash solution. Thus, $U^b(x^*) = \varphi_N(N(b))$, in contradiction with (14). Therefore, the solution is unique.

■(Theorem 2.1)

Proof of Theorem 3.1: The proof is based on the following three lemmas.

Lemma 6.1 *For any* $(S,d) \in B$, $x = \varphi_N(S,d) \iff$ *the line through* x *with slope* $-\frac{x_2-d_2}{x_1-d_1}$ *is tangent to* S.

This lemma, characterizing the slope of the tangent to S at the Nash solution of $(S,d) \in B$, is essentially the same as Lemma VII.2.4 in Owen (1982, page 132).

Definitions: We use the following notation to describe loss-aversion evaluation with respect to varying reference points. The utility of the point $x_i \in \mathbb{R}$ to player i with a loss-aversion coefficient $\lambda_i > 0$, with respect to a reference point $r_i \in \mathbb{R}$ is given by $L_i(x_i,\lambda_i,r_i) = \begin{cases} x_i & \text{if } x_i \geq r_i \\ x_i - \lambda_i(r_i - x_i) & \text{if } x_i < r_i \end{cases}$. We extend the definitions to pairs of players and sets of outcome pairs by $L(x,\lambda,r) = (L_1(x_1,\lambda_1,r_1), L_2(x_2,\lambda_2,r_2))$ and $L(A,\lambda,r) = \{L(x,\lambda,r) | x \in A\}$. We denote the inverse function, that takes points back from the loss aversion evaluation to the underlying utilities (given loss-aversion coefficients and reference points) by L_i^{-1}, and it is given by:

$$L_i^{-1}(x_i,\lambda_i,r_i) = \begin{cases} x_i & \text{if } x_i \geq r_i \\ x_i + \frac{\lambda_i}{1+\lambda_i}(r_i - x_i) & \text{if } x_i < r_i \end{cases},$$

with the corresponding extensions to pairs and sets.

Lemma 6.2 *Given a convex set* $S \subseteq \mathbb{R}^2$ *and* $x \in Par(S)$, *the line through* x *with slope* a *is tangent to* $S \iff \forall c \in [a(1+\lambda_2), \frac{a}{1+\lambda_1}]$, *the line through* x *with slope* c *is tangent to the set* $L(S,\lambda,x)$.

Proof:

$\Longrightarrow$: For any $a < 0$ the function $L(\cdot,\lambda,x)$ transforms the half-plane under the line through x with slope a into the intersection of the half planes under the lines through x with slopes $a(1+\lambda_2)$ and $\frac{a}{1+\lambda_1}$. Thus, since all of S is under the line with slope a, all of $L(S,\lambda,x)$ is under both the other lines, and therefore under their intersection.

$\Longleftarrow$: Since $L^{-1}(\cdot,\lambda,x)$ returns this intersection of the two half planes to the half plane under the line with slope a, this direction is also true.

■(Lemma 6.2)

Lemma 6.3 *Given a closed convex set* $S \subseteq \mathbb{R}^2$ *and* $x \in Par(S)$, *the line through* x *with slope* c *is tangent to* $L(S,\lambda,x) \iff \exists a \in [c(1+\lambda_1), \frac{c}{1+\lambda_2}]$, *such that the line through* x *with slope* a *is tangent to* S.

Proof:

$\Longrightarrow$: Since S is a closed convex set, the set of slopes of tangents to S through x is a closed convex set in $\mathbb{R}$. Denote it by $A = [a^-, a^+]$ (a^- and a^+ may be $-\infty$). From Lemma 6.2, the set of slopes of tangents to $L(S,\lambda,x)$ at x is $[a^-(1+\lambda_2), \frac{a^+}{1+\lambda_1}]$. Thus, if there exists a tangent to $L(S,\lambda,x)$ through x with slope c, then $c \in [a^-(a+\lambda_2), \frac{a^+}{1+\lambda_1}]$, i.e. $a^-(1+\lambda_2) \leq c \leq \frac{a^+}{1+\lambda_1}$. Since $c < 0$, $c(1+\lambda_1) \leq \frac{c}{1+\lambda_2}$. Thus $[a^-, a^+] \cap [c(1+\lambda_1), \frac{c}{1+\lambda_2}]$ is non-empty, and any point in this intersection gives us the slope of a tangent to S through x.

$\Longleftarrow$: Immediate from Lemma 6.2 and the fact that for any $a, c < 0$, $\lambda_1, \lambda_2 > 0$, it is true that $a \in [c(1+\lambda_1), \frac{c}{1+\lambda_2}] \iff c \in [a(1+\lambda_2), \frac{a}{1+\lambda_1}]$.

■(Lemma 6.3)

We now proceed to prove Theorem 3.1.
Since the axiom INV holds for our solution concept φ, we can assume without loss of generality that $d = (0,0)$. Thus, we now prove that $x \in Sel\, f(S,(0,0),\lambda) \iff$ there exists a tangent to S at x with slope a such that $a \in [-\frac{x_2}{x_1}(1+\lambda_2), -\frac{x_2}{x_1}\frac{1}{(1+\lambda_1)}]$.

$\Longrightarrow$:

$$x \in \text{Self}(S,(0,0),\lambda) \Longrightarrow \tag{15}$$

$$x = \varphi(S,(0,0),\lambda,x) \Longrightarrow \tag{16}$$

$$\text{the line through } x \text{ with slope } -\frac{x_2+\lambda_2 x_2}{x_1+\lambda_1 x_1} = -\frac{x_2}{x_1}\frac{(1+\lambda_2)}{(1+\lambda_1)}$$
$$\text{is tangent to } L(S,\lambda,x) \Longrightarrow \tag{17}$$

$$\text{there exists a tangent to } S \text{ through } x$$
$$\text{with slope } a \in [-\frac{x_2}{x_1}(1+\lambda_2), -\frac{x_2}{x_1}\frac{1}{(1+\lambda_1)}], \tag{18}$$

where (17) is from Lemma 6.1, and (18) is from Lemma 6.3.

$\Longleftarrow$:

$$\text{There exists a tangent to } S \text{ through } x \text{ with slope}$$
$$a \in [-\frac{x_2}{x_1}(1+\lambda_2), -\frac{x_2}{x_1}\frac{1}{(1+\lambda_1)}] \Longrightarrow \tag{19}$$

$$\text{All lines through } x \text{ with slopes in } [a(1+\lambda_2), \tfrac{a}{1+\lambda_1}]$$
$$\text{are tangent to } L(S,\lambda,x) \Longrightarrow \tag{20}$$

$$\text{The line with slope } -\tfrac{x_2}{x_1}\tfrac{(1+\lambda_2)}{(1+\lambda_1)} \text{ through } x \text{ is tangent to } U(S,\lambda,x) \Longrightarrow \tag{21}$$

$$x \text{ is the Nash solution to } (L(S,\lambda,x),(-\lambda_1 x_1,-\lambda_2 x_2)) \Longrightarrow \tag{22}$$

$$x = \varphi(S,(0,0),\lambda,x) \Longrightarrow \tag{23}$$

$$x \in \text{Self}(S,(0,0),\lambda), \tag{24}$$

where (20) is from Lemma 6.2, and (22) is from Lemma 6.1.

It remains to show that the set $\text{Self}(S,d,\lambda)$ is non-empty, closed and connected.

If the Pareto frontier of S, denoted by $Par(S)$, does not extend to the axes, extend it with lines parallel to the axes. Denote this extended set P.

For $k>0$ define $f_k:[0,1]\longrightarrow P$ by $f_k(\alpha)=x$ if $x_1\cdot\alpha = x_2\cdot(1-\alpha)\cdot k$. It is easy to see that f_k is one to one, onto P and continuous, for any $k>0$.

Define $g: P \longrightarrow 2^{[0,1]}$ by $g(x)=\{\alpha|(\alpha,1-\alpha) \text{ is normal to } S \text{ at } x\}$. Thus, g is upper semicontinuous.

Define $h_k:[0,1]\longrightarrow 2^{[0,1]}$ by $h_k(\alpha)=g(f_k(\alpha))$. h_k is upper semicontinuous since f_k is continuous and g is upper semicontinuous.

Thus, from Kakutani's fixed point theorem, h_k has a fixed point α_k^*, such that $\alpha_k^* \in h_k(\alpha_k^*)$. It is easy to see that $f_k(\alpha_k^*)\in Par(S)$.

Therefore the vector $(\alpha^*,1-\alpha^*)$ is normal to S at a point x which satisfies $x_1\alpha^* = x_2(1-\alpha^*)k$, i.e. $k\frac{x_2}{x_1}=\frac{\alpha^*}{1-\alpha^*}$, which is equivalent to the fact that the line through x with slope $-k\frac{x_2}{x_1}$ is tangent to S.

For any $k\in[\frac{1}{1+\lambda_1},1+\lambda_2]$ this gives us a point in $\text{Self}(S,(0,0),\lambda)$, which is therefore non-empty. Since $\frac{x_2}{x_1}$ is strictly decreasing as we traverse $Par(S)$ to the right, and the slopes of the tangents to $Par(S)$ are non-increasing in this direction, the set $\text{Self}(S,d,\lambda)$ is a closed segment of $Par(S)$. ■(Theorem 3.1)

Proof of Theorem 3.2:
We start with two lemmas. The first lemma shows that a player can never gain by *increasing* her opponent's reference point from a self-supporting reference point.

Lemma 6.4 *Assume* $(S,d,\lambda)\in B^*$. *If* $x^*\in Sel\,f(S,d,\lambda)$, *and* $r'=(x_i^*,x_{-i}^*+\varepsilon)$ *for* $\varepsilon>0$, *then* $\varphi_i(S,d,\lambda,r')\le\varphi_i(S,d,\lambda,x^*)$.

Proof:

1. Denote, for $x\in S$ and $r\in\mathbb{R}^2$,

$$m(x,r)=(L_1(x_1,\lambda_1,r_1)-L_1(d_1,\lambda_1,r_1))(L_2(x_2,\lambda_2,r_2)-L_2(d_2,\lambda_2,r_2))$$

2. From Theorem 2.1 we know that $\varphi(S,d,\lambda,x^*) = argmax_{x\in S} m(x,x^*)$ and $\varphi(S,d,\lambda,r') = argmax_{x\in S} m(x,r')$.

3. For any point $x \in Par(S)$ with $x_i \geq x_i^*$ (and therefore $x_{-i} \leq x_{i-}^*$), it is true that $m(x,r') = m(x,x^*)$.

4. Therefore, since $argmax_{x\in S} m(x,r') \in Par(S)$ from Theorem 2.1, the argmax must be achieved at some point $x \in Par(S)$ for which $x_i \leq x_i^*$. Thus, since $\varphi_i(S,d,\lambda,x^*) = x_i^*$ from our assumption that $x^* \in Sel\, f(S,d,\lambda)$, it is true that $\varphi_i(S,d,\lambda,r') \leq \varphi_i(S,d,\lambda,x^*)$.

■(Lemma 6.4)

The following lemma is a consequence of Lemma 6.1, since S is a convex set. It deals with points on the Pareto frontier of S that are different from the Nash solution to (S,d).

Lemma 6.5 *For any $(S,d) \in B$, $x \in Par(S)$, $i \in \{1,2\}$, the line through x with slope $-\frac{x_2-d_2}{x_1-d_1}$ intersects the interior of S at a point x' such that $x_i' > x_i \iff \varphi_{Ni}(S,d) > x_i$, where φ_{Ni} denotes the Nash solution outcome for player i.*

The following observation deals with the Pareto frontier of $L(S,\lambda,r')$ around x^* when $r' = (x_i^*, x_{-i}^* - \varepsilon)$. It states that for points giving more to player i than x^*, the Pareto frontier of $L(S,\lambda,r')$ is equal to the Pareto frontier of $L(S,\lambda,x^*)$, and for points giving less to player i than x^* (and close enough to x^*), the Pareto frontier of $L(S,\lambda,r')$ is identical to that of S.

Observation: If $(S,d,\lambda) \in B^*,\ x^* \in Par(S),\ r' = (x_i^*, x_{-i}^* - \varepsilon),\ \varepsilon > 0,\ x \in Par(S)$ then

1. $x_{-i}^* \leq x_{-i} \Longrightarrow L(x,\lambda,r') = L(x,\lambda,x^*)$.

2. $x_{-i}^* - \varepsilon \leq x_{-i} \leq x_{-i}^* \Longrightarrow L(x,\lambda,r') = x$.

We now proceed with the proof of Theorem 3.2:
As in the proof of Theorem 3.1, we assume without loss of generality that $d = (0,0)$.

Take $(S,(0,0),\lambda) \in B^*$ and $x^* \in \text{Self}(S,(0,0),\lambda)$. From the above observation and Lemma 6.5, the condition for non-stability of x^* arising from the possibility of player 1 gaining from reducing player 2's reference point is:

$$\exists \varepsilon > 0 \text{ such that } -\frac{x_2^* + \lambda_2(x_2^* - \varepsilon)}{x_1^* + \lambda_1 x_1^*} > a,$$

for $a < 0$ satisfying that the line through x^* with slope a is tangent to S.

Similarly, the condition for non-stability of x^* arising from the possibility of player 2 gaining from reducing player 1's reference point is:

$$\exists \varepsilon > 0 \text{ such that } -\frac{x_2^* + \lambda_2 x_2^*}{x_1^* + \lambda_1(x_1^* - \varepsilon)} < a,$$

for $a < 0$ satisfying that the line through x^* with slope a is tangent to S.

For x^* to be stable, we need both these conditions to be false for any $\varepsilon > 0$, which is equivalent to

$$-\frac{x_2^*(1+\lambda_2)}{x_1^*(1+\lambda_1)} = a$$

for $a < 0$ satisfying that the line through x^* with slope a is tangent to S.

Such a point exists from the proof of existence in Theorem 3.1, with $k = \frac{1+\lambda_2}{1+\lambda_1}$. It is unique, since $-\frac{x_2}{x_1}$ is strictly increasing as we traverse the Pareto frontier of S to the right, while the slope of the tangents to S is non-increasing as we move in the same direction. ■(Theorem 3.2)

Proof of Lemma 5.1:
We assume without loss of generality that $d = (0,0)$. We deal with two cases, $r \in S$ and $r \notin S$.
Case 1: $r \in S$.
According to Theorem 2.1, the solution $x = \varphi(S,d,\lambda,r)$ is on the Pareto frontier of S, and maximizes the function $m(x,r) = m_1(x,r) \cdot m_2(x,r)$, where

$$m_i(x,r) = \begin{cases} x_i(1+\lambda_i) & \text{if } x_i \le r_i \\ x_i + \lambda_i r_i & \text{if } x_i \ge r_i \end{cases}$$

We divide the Pareto frontier into three parts, A: $x_1 < r_1$, B: $x_1 \ge r_1$ and $x_2 \ge r_2$, and C: $x_2 < r_2$.

For $x \in A$, $m(x,r) = x_1(1+\lambda_1)(x_2+\lambda_2 r_2)$. If the solution is in A, it is at a point x satisfying $-f'(x_1) = \frac{x_2+\lambda_2 r_2}{x_1}$ and therefore from Theorem 3.1 x is self-supporting (since $\frac{x_2}{x_1(1+\lambda_1)} \le \frac{x_2+\lambda_2 r_2}{x_1} \le \frac{x_2(1+\lambda_2)}{x_1}$).

For $x \in C$, $m(x,r) = x_2(1+\lambda_2)(x_1+\lambda_1 r_1)$. If the solution is in C, it is at a point x satisfying $-f'(x_1) = \frac{x_2}{x_1+\lambda_1 r_1}$ and therefore from Theorem 3.1 x is self-supporting.

For $x \in B$, $m(x,r) = (x_2+\lambda_2 r_2)(x_1+\lambda_1 r_1)$. If the solution is in the relative interior of B, it is at a point x satisfying $-f'(x_1) = \frac{x_2+\lambda_2 r_2}{x_1+\lambda_1 r_1}$ and therefore from Theorem 3.1 x is self-supporting.

We are left with the endpoints of B. At $x = (r_1, f(r_1))$, x maximizes $m(x,r)$ if $-f'(x_1) \le \frac{x_2(1+\lambda_2)}{x_1}$ and $-f'(x_1) \ge \frac{x_2+\lambda_2 r_2}{x_1+\lambda_1 r_1}$ Thus $\frac{x_2}{x_1(1+\lambda_1)} \le \frac{x_2+\lambda_2 r_2}{x_1(1+\lambda_1)} = \frac{x_2+\lambda_2 r_2}{x_1+\lambda_1 r_1} \le -f'(x_1) \le \frac{x_2+\lambda_2 r_2}{x_1} \le \frac{x_2(1+\lambda_2)}{x_1}$ and $x \in \text{Self}(S,d,\lambda)$ from Theorem 3.1. The other endpoint is treated analogously.
Case 2: $r \notin S$.
Here too we divide the Pareto frontier into three parts, A: $x_2 > r_2$, B: $x_1 \le r_1$ and $x_2 \le r_2$, and C: $x_1 > r_1$. Parts A and C are treated as in case 1. For $x \in B$, $m(x,r) = x_1(1+\lambda_1)x_2(1+\lambda_2)$. If the solution is in this area, it is therefore the Nash solution, which satisfies $-f'(x_1) = \frac{x_2}{x_1}$ from Lemma 6.1, and is therefore self-supporting from Theorem 3.1. ■ (Lemma 5.1)

Proof of Lemma 5.2: We assume without loss of generality that $d = (0,0)$. Denote by A the subset of the Pareto frontier of S where $x_1 < r_1$, by C the subset where $x_2 < r_2$, and by B the

rest of the Pareto frontier of S. We show that $\frac{\partial m(x,r)}{\partial x_1} > 0$ for $x \in A$, and $\frac{\partial m(x,r)}{\partial x_1} < 0$ for $x \in C$, and therefore the solution outcome is in B, therefore satisfying our requirements.

For $x \in A$, $m(x,r) = x_1(1+\lambda_1)(x_2+\lambda_2 r_2)$. Therefore, for $x \in A$,

$$\frac{\partial m(x,r)}{\partial x_1} > 0 \iff -f'(x_1) < \frac{x_2+\lambda_2 r_2}{x_1}$$

which is satisfied since $x_1 < r_1$, $x_2 > r_2$, and $(r_1, f(r_1))$ is self-supporting. For $x \in C$ the proof is analogous. ■(Lemma 5.2)

Proof of Lemma 5.3: We assume without loss of generality that $d = (0,0)$ and that $r_1 > x_1^*$, $r_2 < x_2^*$. We show that for any point on the Pareto frontier of S, if $x_1 > r_1$ then $\frac{\partial m(x,r)}{\partial x_1}(x) < 0$.

For $x_1 > r_1$,

$$\frac{\partial m(x,r)}{\partial x_1} < 0 \iff -f'(x_1) > \frac{x_2}{x_1+\lambda_1 r_1}$$

which is true from Theorem 3.1 as $x_1 > r_1 > x_1^*$. ■(Lemma 5.3)

Proof of Theorem 5.1: Fix a repeated bargaining game $\Gamma(b, r^0, \delta)$ with $b = (S, d, \lambda)$, and denote the stable self-supporting reference point pair of (S, d, λ) by x^*. Fix $\varepsilon > 0$.

Since $d(D^t) \to 0$,[14] $\exists T_1$ *s.t.* $d(D^{T_1}) < \varepsilon$. Take σ_1 and σ_2 undominated pure strategies of players 1 and 2 respectively. If $x^{T_1} \in D^{T_1}$ then for all $t > T_1$, $|x^t - x^*| < \varepsilon$ from Lemma 5.4. If $x^{T_1} \notin D^{T_1}$ then assume w.l.o.g. that $x_2^{T_1} < x_2^*$. If $\forall t > T_1$, $x_2^t \notin D^{T_1}$ then σ_2 is dominated by the strategy σ_2' which is given by "Play as σ_2 unless $t > T_1$ and $x_2^t \notin D^{T_1}$, in which case play $\alpha_2^t = -\frac{1}{t}$". The strategy σ_2' dominates σ_2, as after period T_1 the payoff is at least as large in every period, and will reach D^{T_1} after a finite time and will then strictly dominate the payoff from σ_2 (since $\sum \frac{1}{t} = \infty$), in contradiction to our assumption that σ_2 is undominated. Therefore if σ_2 is undominated $\exists T_2 > T_1$ s.t. $x^{T_2} \in D^{T_1}$, and $|x^t - x^*| < \varepsilon$ for all $t > T_2$ from Lemma 5.4. ■(Theorem 5.1)

References

[1] Bazerman, M. H., T. Magliozzi and M. A. Neale (1985): "Integrative Bargaining in a Competitive Market," *Organizational Behavior and Human Decision Processes* **35** 294-313.

[2] Binmore, K. (1994): *Playing Fair: Game Theory and the Social Contract,* The MIT Press, Cambridge, MA.

[14] $d(A)$ denotes the diameter of the set A.

[3] De Dreu, C. K. W., B. J. M. Emans and E. Van de Vliert (1992): "Frames of Reference and Cooperative Social Decision Making," *European Journal of Social Psychology* **22** 297-302.

[4] Hogarth, R. M., and M. W. Reder (1986): "Introduction: Perspectives from Economics and Psychology," in *Rational Choice: The Contrast between Economics and Psychology*, eds. Hogarth and Reder, The University of Chicago Press, Chicago, IL.

[5] Kahneman, D. (1992): "Reference Points, Anchors, Norms, and Mixed Feelings," *Organizational Behavior and Human Decision Processes*, **51** 296-312.

[6] Kahneman, D., J. L. Knetsch and R. H. Thaler (1990): "Experimental Tests of the Endowment Effect and and the Coase Theorem," *Journal of Political Economy* **98** (6) 1325-1348.

[7] Kahneman, D., J. L. Knetsch and R. H. Thaler (1991): "The Endowment Effect, Loss Aversion and Status Quo Bias," *Journal of Economic Perspectives* **5** (1) 193-206.

[8] Kahneman, D. and A. Tversky (1979): "Prospect Theory: An Analysis of Decision Under Risk," *Econometrica* **47** 263-291.

[9] Kalai, E. and M. Smorodinsky (1975): "Other Solutions to Nash's Bargaining Problem," *Econometrica* **43** 513-518.

[10] Kannai, Y. (1977): "Concavifiability and Constructions of Concave Utility Functions," *Journal of Mathematical Economics* **4** 1-56.

[11] Kihlstrom, R. E., A. E. Roth, and D. Schmeidler (1981): "Risk Aversion and Nash's Solution to the Bargaining Problem," *Game Theory and Mathematical Economics*, ed. by O. Moeschlin and D. Pallaschke. Amsterdam: North-Holland.

[12] Kramer, R. M. (1989): "Windows of Vulnerability or Cognitive Illusions: Cognitive Processes and the Nuclear Arms Race," *Journal of Experimental Social Psychology* **25** 79-100.

[13] Nash, J. F. (1950): "The Bargaining Problem," *Econometrica* **18** 155-162.

[14] Nash, J. F. (1953): "Two-Person Cooperative Games," *Econometrica* **21** 128-140.

[15] Neale, M. A. and M. H. Bazerman (1985): "The Effects of Framing and Negotiator Confidence on Bargaining Behaviors and Outcomes," *Academy of Management Journal*," **28** (1) 34-39.

[16] Neale, M. A., V. L. Huber and G. B. Northcraft (1987): "The Framing of Negotiations: Contextual Versus Task Frames," *Organizational Behavior and Human Decision Processes*, **39** 228-241.

[17] Osborne, M. J. and A. Rubinstein (1990): "Bargaining and Markets," Academic Press, San Diego, CA.

[18] Owen, G. (1982): "Game Theory, Second Edition," Acadamic Press, New York.

[19] Rabin, M. (1996): "Psychology and Economics," *mimeo*, University of California at Berkeley.

[20] Roth, A. E. (1979): *Axiomatic Models of Bargaining*, Berlin and New York: Springer.

[21] Roth, A. E., and U. G. Rothblum (1982): "Risk Aversion and Nash's Solution for Bargaining Games with Risky Outcomes," *Econometrica* **50** (3) 639-647.

[22] Rubinstein, A. (1982): "Perfect Equilibrium in a Bargaining Model," *Econometrica* **50** (1) 97-109.

[23] Sobel, J. (1981): "Distortion of Utilities and the Bargaining Problem," *Econometrica* **49** 597-620.

[24] Taylor, S. E. (1991): "Asymmetrical Effects of Positive and Negative Events: The Mobilization-Minimization Hypothesis," *Psychological Bulletin* **110** 67-85.

[25] Tversky, A. and D. Kahneman (1991): "Loss Aversion in Riskless Choice," *Quarterly Journal of Economics* **106** (4) 1039-1061.

[26] Tversky, A. and D. Kahneman (1992): "Advances in Prospect Theory: Cumulative Representation of Uncertainty," *Journal of Risk and Uncertainty,* **5** 297-323.