

ON THE INTERPRETATION OF THE NASH BARGAINING SOLUTION AND ITS EXTENSION TO NON-EXPECTED UTILITY PREFERENCES¹

BY ARIEL RUBINSTEIN, ZVI SAFRA, AND WILLIAM THOMSON

The paper reexamines the foundations of the axiomatic Nash bargaining theory. More specifically it questions the interpretation of the Nash bargaining solution and extends it to a family of non-expected utility preferences.

A bargaining problem is presented as $\langle X, D, \geq_1, \geq_2 \rangle$ where X is a set of feasible agreements (described in physical terms), D is the disagreement event and \geq_1 and \geq_2 are preferences defined on the space of lotteries in which the prizes are the elements in X and D . The (ordinal)-Nash bargaining solution is defined as an agreement y^* satisfying for all $p \in [0, 1]$ and for all $x \in X$: if $px >_1 y^*$ then $py^* \geq_2 x$ and if $px >_2 y^*$ then $py^* \geq_1 x$ where px is the lottery which gives x with probability p and D with probability $1 - p$.

Revisions of the Pareto, Symmetry, and IIA Axioms characterize the (ordinal)-Nash bargaining solution. In the expected utility case this definition is equivalent to that of the Nash bargaining solution. However, this definition is to be preferred since it allows a statement of the Nash bargaining solution in everyday language and makes possible its natural extension to a wider set of preferences. It also reveals the logic behind some of the more interesting results of the Nash bargaining solution such as the comparative statics of risk aversion and the connection between the Nash bargaining solution and strategic models.

KEYWORDS: Nash bargaining solution, non-expected utility, interpretation.

1. THE NASH BARGAINING SET-UP

THIS PAPER REEXAMINES the foundations of the axiomatic bargaining theory as formulated by Nash (1950). More specifically, it questions the standard interpretation of the Nash bargaining solution and extends its scope to a family of non-expected utility preferences.

Let us review the basic elements of Nash's (two-person) bargaining theory. The bargaining problem consists of the "feasible set", S , and a "disagreement point", d . Each element of S gives the utility levels reached by the two agents at one (or more) of the possible agreements. The utilities are understood to be *von Neumann-Morgenstern utilities* in that they are derived from preferences over lotteries which satisfy the expected utility assumptions. A bargaining solution is a function which assigns a unique pair of utility levels to each problem $\langle S, d \rangle$ taken from some domain (usually containing all problems $\langle S, d \rangle$ where S is compact and convex and contains a point that strictly dominates d).

¹This paper is a combination of two projects. It contains a previously unpublished paper coauthored by the first and the third authors entitled "On the Interpretation of the Nash Bargaining Solution" and includes new material regarding the extension of the Nash solution to non-expected utility preferences prepared by the first and the second authors.

We wish to thank the editor and two referees of this journal for very valuable comments.

Nash showed that there is a unique solution satisfying the following four axioms:

(IAT) INVARIANCE TO POSITIVE AFFINE TRANSFORMATIONS: *The solution is invariant to all independent person-by-person, positive affine transformations of utilities.*

(SYM) SYMMETRY: *If the set S is symmetric with respect to the main diagonal and if $d_1 = d_2$, then the solution should assign equal utilities to both players.*

(PAR) PARETO OPTIMALITY: *There is no point in S which Pareto-dominates the solution outcome.*

(IIA) INDEPENDENCE OF IRRELEVANT ALTERNATIVES: *If $S \subset T$ and $f(T, d) \in S$, then $f(T, d) = f(S, d)$.*

The unique solution satisfying the above four axioms is the Nash solution, i.e., the function N defined by

$$N(S, d) = \operatorname{argmax} \{(u_1 - d_1)(u_2 - d_2) \mid (u_1, u_2) \in S \text{ and } u_i \geq d_i \text{ for both } i\}.$$

The very simplicity of this formula is in itself an attractive feature and is responsible for the widespread application of the solution. However, the solution lacks a straightforward interpretation since the meaning of a product of two von Neumann-Morgenstern utility numbers is unclear. One of the main goals of this paper is to provide a more attractive definition. We will therefore start by restating Nash's model in terms of preferences. This will prove useful not only in re-interpreting the axioms and the Nash solution, but also in exposing the logic behind some of the more interesting results of the Nash bargaining solution such as the comparative statics of risk aversion and the connection between the Nash bargaining solution and strategic models. Furthermore, in recent years there has been a growing interest in non-expected utility theories of decision making under uncertainty, since they explain a wide range of behavior patterns and experimental results that are inconsistent with expected utility theory. The statement of the model in terms of preferences allows us to extend the definition of the Nash solution to a significant class of non-expected utility preferences. As such, we view this paper as a part of the big project of examining the significance of the expected utility assumptions for the foundations of game theory.

Let us start by restating the Nash bargaining model.

2. STARTING FROM PREFERENCES

In our search for a more natural definition of the Nash solution we switch from utility language to alternatives-preferences language. A Nash problem,

$\langle S, d \rangle$, is a condensed version of a more natural problem, $\langle X, D, \succsim_1, \succsim_2 \rangle$ where X is a set of feasible alternatives (described in physical terms), D is the disagreement alternative, and \succsim_1 and \succsim_2 are preferences defined on the space of lotteries (probability measures with finite support) in which the prizes are D and the elements of X .² We think about the feasible alternatives as being deterministic outcomes. From now on, we refer to a quadruple $\langle X, D, \succsim_1, \succsim_2 \rangle$ as a *problem*. We restrict the set of problems as follows:

- (i) X is a compact set (in some topological space).
- (ii) The preferences \succsim_1 and \succsim_2 are continuous.
- (iii) For all $x \in X$ and both i , $x \succsim_i D$ and there exists at least one $x \in X$ such that $x \succ_i D$ for both i .
- (iv) The problem is *convex* in the sense that for all $x, y \in X$ and for all $\alpha \in [0, 1]$ there is $z \in X$ such that both players are indifferent between z and $\alpha x + (1 - \alpha)y$, the lottery which gives x with probability α and y with probability $1 - \alpha$.

To simplify the presentation we make two additional assumptions:

- (v) There are no two alternatives x and x' such that $x \sim_i x'$ for both i .³
- (vi) For each i , there is a unique best agreement $B_i \in X$, which satisfies $B_i \sim_j D$ (and thus is strongly Pareto optimal).

Specifying the domain of a solution is a delicate issue. A certain pair $\langle S, d \rangle$ can be derived from *many* different quadruples $\langle X, D, \succsim_1, \succsim_2 \rangle$. Nash's space of pairs $\langle S, d \rangle$ may be spanned by either varying preferences over a fixed set of alternatives or by keeping preferences on a universe of possible alternatives fixed and varying the set of alternatives. Nash's hidden assumption is that all problems that produce the same pair $\langle S, d \rangle$ have the same payoff vector. The significance of the space choice lies in its use within the axioms. Recall that whereas two of the axioms, PAR and SYM, refer to problems in isolation, the other two, IAT and IIA, require some relationship between the solution outcomes for any two problems which are related in a certain manner. The larger the domain of a solution, the more restrictive are the axioms imposed on it. In particular, if all possible quadruples $\langle X, D, \succsim_1, \succsim_2 \rangle$ are included in the domain, axioms pertaining to pairs relate the solution outcomes of problems that may be different either in terms of sets of alternatives, in terms of preference relations, or both. Conversely, if we "overrestrict" the domain, the axioms lose some of their power and may cease to characterize a unique solution. Thus, the choice of domain has to be related to the consideration that leads to the imposition of the IAT and IIA axioms in the first place. Therefore, let us reexamine the justification of those two axioms.

The textbook justification of IAT (see for example Luce and Raiffa (1957)) is that the choice of utility representations should not affect the solution. The

² It is assumed implicitly by this formulation that compound lotteries are reduced to simple lotteries by the usual multiplication rule.

³ Without this assumption there could be a multiplicity of agreements, all of which are equivalent in terms of the players' preferences. Eliminating the assumption only requires rephrasing of several definitions and results in terms of equivalent classes of agreements.

von Neumann-Morgenstern utility representations are unique up to positive affine transformations. Rescaling the utilities should not, therefore, affect the alternative predicted by the solution. As such, IAT appears to be a minimal requirement. However, if we allow the domain to include problems with different sets of alternatives, the axiom loses its innocence. Consider, for instance, two risk-neutral expected utility maximizers who must divide a prize worth one dollar for bargainer 1 and one thousand dollars for bargainer 2. IAT requires that Player 2 receives the same share of the prize as he would were the prize worth only 1 dollar for bargainer 2. Thus, when IAT is applied to a solution with a domain comprising a varied set of alternatives, it is more demanding than when it is justified as giving meaning to the preferences.

The textbook interpretation of IIA is that it relates the solution outcomes of problems with different sets of alternatives. If a^* is the solution outcome of the problem $\langle T, d \rangle$ and is a member of a set S which is a subset of T , then a^* is also the solution outcome of $\langle S, d \rangle$. As has often been emphasized, this justification of IIA fits in with a normative theory, where the solution concept is intended to reflect the social desirability of an alternative. In this paper we consider bargaining as a strategic interaction of two self-interested bargainers and in this context IIA is questionable (see Binmore (1987a)).

Our resolution of the modelling dilemma regarding the choice of domain consists of the fixing of a set of alternatives and varying the bargainers' preferences. Given a pair of preferences on X , a bargaining solution is henceforth a function which assigns a unique element of X to every pair of preferences on the set of lotteries on X within some domain.

The above definition makes IAT redundant, while availing us of a more attractive interpretation of IIA. We use the letter F to denote a solution and the symbol px for the lottery which gives x with probability p and D with probability $1 - p$.

IIA: Let $F(\geq_1, \geq_2) = y^$ and let \geq'_i be a preference which agrees with \geq_i on the set of deterministic agreements, X , such that: (i) for all x such that $x \geq_i y^*$, if $px \sim_i y^*$, then $px \leq'_i y^*$; and (ii) for all x such that $x \leq_i y^*$, if $x \sim_i qy^*$, then $x \sim'_i qy^*$. Then $F(\geq_i, \geq_j) = F(\geq'_i, \geq_j)$.*

The switch of agent i 's preference from \geq_i to \geq'_i reflects his increased apprehension towards the risk of demanding alternatives which are *better* than the outcome y^* . Though player i still prefers x to y^* he is less willing to risk demanding x . The axiom captures an intuition that the bargaining solution outcome y^* should be defensible against possible objections. The change in player i 's preference described in the axiom, makes player i "less eager" to object and does not change the intensity of player j 's objections. Thus, the change in the preference "should not" change the bargaining outcome.

We denote the axiom as IIA because of its connection with Nash's Independence of Irrelevant Alternatives Axiom. However, as we have already emphasized, our axiom refers to a comparison between two problems where the

preferences vary, while Nash's axiom refers to two problems, in one of which "irrelevant alternatives" are dropped.

The above definition of IIA is related to the second part of what Binmore (1987c) calls IIA which states: "... we require that, if both players' utilities for all feasible trades except the solution outcome of the no-trade point decrease, then the solution (outcome) remains unchanged."

Before redefining the Nash solution let us restate PAR and SYM. There is no difficulty in defining PAR in terms of preferences.

PAR: *If $F(\geq_1, \geq_2) = y^*$, then there is no $x \in X$ such that for all i , $x \geq_i y^*$ with a strong preference on the part of at least one i .*

In order to formulate the symmetry axiom we first define a *symmetric problem* as one for which there exists a *symmetry function* $\phi: X \cup \{D\} \rightarrow X \cup \{D\}$ which satisfies $\phi(x) = y$, if and only if $\phi(y) = x$ and $\phi(D) = D$ such that for all lotteries L_1 and L_2 , $L_1 \geq_i L_2$, if and only if $\phi(L_1) \geq_j \phi(L_2)$, where $\phi[pa + (1 - p)b]$ is the lottery $p\phi(a) + (1 - p)\phi(b)$.

SYM: *If $\langle X, D, \geq_1, \geq_2 \rangle$ is a symmetric problem with a symmetry function ϕ and the preferences \geq_1 and \geq_2 satisfy the expected utility assumptions, then $F(\geq_1, \geq_2)$ is a fixed point of the symmetry function ϕ .*

To see that the above formulation of SYM is essentially the same as Nash's original symmetry axiom, consider a symmetric problem $\langle X, D, \geq_1, \geq_2 \rangle$ where the preferences satisfy the expected utility assumptions. To every $x \in X$ there are two numbers p_1 and p_2 such that $x \sim_1 p_1 B_1 + (1 - p_1) B_2$. The function ϕ must assign to x an agreement in X such that $\phi(x) \sim_j p_2 B_2 + (1 - p_2) B_1$. Thus, in terms of von Neumann-Morgenstern utilities, if we choose $u_i(D) = 0$ and $u_i(B_i) = 1$ we receive ϕ which transforms an agreement x with a pair of utilities $(u_1(x), u_2(x)) = (a, b)$ into $\phi(x)$ satisfying $(u_1(\phi(x)), u_2(\phi(x))) = (b, a)$. It follows that for symmetric problems with preferences satisfying the expected utility assumptions there is a unique symmetry function ϕ with a unique Pareto-optimal fixed point. One should note that a point x^* is a fixed point of ϕ if and only if $u_1(x^*) = u_2(x^*)$.

Before we redefine the Nash solution, recall the common definition of the Nash solution as applied to the problem $\langle X, D, \geq_1, \geq_2 \rangle$ where \geq_1 and \geq_2 are expected utility preferences represented by the utility functions u_1 and u_2 . The (utility)-Nash bargaining solution of $\langle X, D, \geq_1, \geq_2 \rangle$ is

$$\operatorname{argmax} \{ (u_1(x) - u_1(D))(u_2(x) - u_2(D)) \mid x \in X \}.$$

COMMENT: Several other papers have also dealt explicitly with the underlying physical structure of the set of agreements (see Binmore (1987c), Chun and Thomson (1988), Roemer (1986), and Roemer (1988)). In contrast to our model, these papers present the set of agreements as one of the variables in the domain

of the bargaining solution. Usually, an algebraic structure is imposed on the set of agreements and some of the axioms utilize this structure. Roemer (1990) is particularly relevant to our discussion. He argues that by fixing von Neumann-Morgenstern preferences and varying the set X , “the natural generalization of the Nash axioms hardly restricts the bargaining solution.” Thus, Roemer’s work may be considered as additional support for our choice of domain. Notice, however that Roemer’s result relies on a weak generalization of Nash’s Symmetry Axiom.

3. REDEFINING THE NASH BARGAINING SOLUTION

In this section we look for an alternative definition of the Nash solution, one which only uses the terms “alternative,” “disagreement,” and “preference” and avoids the term “utility.” After translating Nash’s axioms into the alternatives-preferences language, one may think that the conjunction of the axioms is a proper definition of the Nash solution. However, as already emphasized, while PAR and SYM refer to each problem in isolation, the other axiom, IIA, requires consistency between the solution outcomes of *different* problems. We look for a definition that specifies an outcome of a particular problem directly in terms of the problem, without referring to its consistency with outcomes suggested by the solution in other problems. The definition of the Nash solution which uses utility language fulfills this requirement since it is applied to each problem in isolation. However, as already stated, it requires a noninterpretable maximization and is applicable only in the expected utility case.

DEFINITION: An *(Ordinal)-Nash solution* outcome for the problem $\langle X, D, \geq_1, \geq_2 \rangle$ is an alternative y^* , such that for all $p \in [0, 1]$ and for all $x \in X$ and i , if $px >_i y^*$, then $py^* \geq_j x$.

The definition of the (ordinal)-Nash solution has a clear interpretation. Assume that the players perceive that whenever they raise an objection to an alternative, they face a risk that the negotiations will end in disagreement. If player i appeals against y^* by suggesting an alternative x , which is to his advantage, notwithstanding the risk that the appeal could cause disagreement with some probability $1 - p$, then it is optimal for player j to insist on the original alternative y^* even though his insistence may cause disagreement with the same probability $1 - p$.

First, we wish to verify that for the expected utility case the ordinal definition coincides with the definition of the (utility)-Nash solution. The following proposition shows the equivalence of the definitions for all classes of problems (not just those satisfying assumptions (i)–(v)).

PROPOSITION 1: *Let $\langle X, D, \geq_1, \geq_2 \rangle$ be a problem (not necessarily satisfying assumptions (i)–(v)) such that the preferences are expected utility preferences. The (ordinal)-Nash solution outcome is then well defined and is equal to y^* if and only if the (utility)-Nash solution outcome is well defined and is equal to y^* .*

PROOF: Let u_1 and u_2 be utility representations of \succsim_1 and \succsim_2 . We show that y^* is an (ordinal)-Nash solution outcome of $\langle X, D, \succsim_1, \succsim_2 \rangle$ if and only if $y^* = \operatorname{argmax}\{(u_1(x) - u_1(D))(u_2(x) - u_2(D)) \mid x \in X\}$.

Without loss of generality we can assume that $u_i(D) = 0$ for both i . The alternative y^* satisfies $u_1(y^*)u_2(y^*) \geq u_1(x)u_2(x)$ for all x if and only if for both i , for all $x \succ_i y^*$ for all $p \leq 1$, if $p > u_i(y^*)/u_i(x)$ then $p \geq u_j(x)/u_j(y^*)$ if and only if for all i, x and $p \leq 1$ satisfying $pu_i(x) > u_i(y^*)$, it is also true that $pu_j(y^*) \geq u_j(x)$ if and only if for all i, x , and $p \leq 1$, $px \succ_i y^*$ implies $py^* \succ_j x$.
Q.E.D.

REMARK: It is appropriate at this point to mention here the connection between the above definition of the Nash solution and previous related ideas. Zeuthen (1930) was the first to suggest a theory in which negotiators bear in mind the risk of a breakdown in the negotiations. He describes a dynamic process which reflects what he calls "psychological considerations." The set X is taken to have a finite number of Pareto-optimal alternatives. At each stage, given the two current positions x_1 and x_2 , one of the players with the lowest $1 - p_i$ (where p_i is defined by $p_i x_i + (1 - p_i)D \sim_i x_j$) makes a "slight" concession. Thus, a concession is made by a player whose position is associated with a product of utilities which is no greater than that of the other player's position. By its definition this process must converge and it was Harsanyi (1956) who showed, through an argument similar to the above proof, that the process converges to the Nash bargaining solution outcome. Similarly, Aumann and Kurz (1977) define a measure of a player's boldness as the maximum probability which makes him willing to take the risk of losing the entire gain against an additional small gain. They observe that the Nash solution outcome is the point at which both players are equally bold.

4. THE FAMILY OF PREFERENCES

We will endeavor to show that the revisions of the axioms PAR, SYM, and IIA characterize the (ordinal)-Nash solution for a family of preferences wider than those which satisfy the expected utility assumptions. The interest in preferences which violate the independence axiom⁴ is now well-established.⁵ This section contains a preliminary discussion of the family of preferences for which we are able to extend the Nash theory.

We consider preferences which satisfy the following assumptions:

DOM (FIRST ORDER STOCHASTIC DOMINANCE): *If $x \succ_i D$ and $p > q$, then $px + ry \succ_i qx + ry$ (where $px + ry$ is the lottery which gives x, y , and D with probabilities p, r , and $1 - p - r$ accordingly).*

⁴ If $L_1 \succcurlyeq L_2$, then for all p and for all L , $pL_1 + (1 - p)L \succcurlyeq pL_2 + (1 - p)L$.

⁵ For surveys of the literature on non-expected utility preferences see, for example, Machina (1987), Fishburn (1989), or Karni and Schmeidler (1990).

Q (QUASI CONCAVITY): *If $L' >_i L$, then $\alpha L' + (1 - \alpha)L >_i L$.*

CCE (CONDITIONAL CERTAINTY EQUIVALENCE): *If $x \sim_i L'$, then $\alpha x + (1 - \alpha)z \geq_i \alpha L' + (1 - \alpha)z$ for any z such that $z \geq_i y$ for all y in the support of L' .*

H (HOMOGENEITY): *If $x \geq_i L$, then $\alpha x \geq_i \alpha L$, and if $x \sim_i L$, then $\alpha x \sim_i \alpha L$.*

Conditions DOM and Q are quite standard. DOM requires that a shift of probability from an inferior prize to a superior one makes the lottery more attractive. Notice that Q in this context is no more plausible than quasiconvexity. However, for our purposes we do not need the more restrictive assumption that $L' \sim_i L$ implies $\alpha L' + (1 - \alpha)L >_i L$. Thus, we allow for preferences which satisfy both quasiconcavity and quasiconvexity (as do preferences which satisfy the expected utility assumptions). Condition CCE states that if x is the certainty equivalent of the lottery L' , then whenever L' is an element in a compound lottery with a prize z which is better than all prizes in the support of L' , the individual is willing to exchange L' for x . In other words, the decision maker is at least as risk-averse when there is a chance to win a superior prize (receiving a prize z with probability $1 - \alpha$) as when there is no such chance. A player willing to insure himself by trading L' for some sure outcome would be willing to do so for a less favorable outcome if more uncertainty existed. Condition CCE is satisfied whenever the preference satisfies the fanning out property defined by Machina (1987). An example of a preference which satisfies DOM, Q, and CCE is the preference defined over the lotteries on the interval $[1, 2]$ and represented by the expectation of the product wv divided by the expectation of w where $w(x) = x^t$ ($t < 0$) and $v(x) = \lambda x$ ($\lambda > 0$).⁶ This preference belongs to the weighted utility class (see Chew (1983)).

Finally, condition H is a strong restriction on the set of preferences. It permits arbitrary preference on the set of lotteries over X ; however, if this preference is represented by a functional V , then condition H is satisfied if the decision maker decomposes a lottery L into $\alpha L' + (1 - \alpha)D$, where L' is a lottery over X , and evaluates the lottery L as $\alpha V(L')$. (Notice, however, that if the preference is represented by the function $f(\alpha)V(L')$ for nonlinear function f , it does not satisfy H.)

Before we proceed to the main proposition we need some notation and two lemmas. For $x, y \in X$, define $x \gg y$ if there is a bargainer i and a number p such that both $px >_i y$ and $x >_j py$. In other words, $x \gg y$ if one of the players, player i , can "appeal" successfully against y by suggesting x . That is, he prefers x to y , even when taking into account a probability $1 - p$ of breakdown, and his opponent, player j , who also fears that his insistence on y will cause a

⁶ Q is satisfied since this preference is both quasiconcave and quasiconvex. CCE is satisfied since local risk aversion increases when a stochastically dominating shift is made. DOM is satisfied since every local utility function is increasing.

breakdown with probability $1 - p$, prefers agreeing on x . If p and q satisfy $px \sim_1 y$ and $qy \sim_2 x$, then (assuming DOM and continuity) $x \gg y$ if and only if $p < q$. Thus if $x \gg y$, then not $[y \gg x]$. Using the new terminology we can say that the alternative y^* is the Nash solution outcome if there is no x such that $x \gg y^*$. Under expected utility $x \gg y$ if and only if $u_1(x)u_2(x) > u_1(y)u_2(y)$ where for both i , u_i is taken to represent \geq_i with $u_i(D) = 0$.

LEMMA 1: Consider a pair of preferences (\geq_1, \geq_2) which satisfy conditions DOM, Q, and H. Let y, x, z be Pareto-optimal alternatives in X satisfying $z >_1 x >_1 y$. If not $[x \gg z]$ then $x \gg y$.

PROOF: Assume the contrary. Let p and q satisfy $pz \sim_1 x$ and $qy \sim_2 x$. Since not $[x \gg z]$ then $z \geq_2 px$ and since not $[x \gg y]$ then $y \geq_1 qx$. By the continuity of \geq_1 there exists t such that $x \sim_1 tz + (1 - t)y$. Using H, $qpz \sim_1 qx$ and by Q, $(1 - q(1 - t))pqz + q(1 - t)y \geq_1 qx$. Now, by H again $qx \sim_1 qtz + q(1 - t)y$ and by DOM we conclude that $(1 - q(1 - t))p \geq t$. Similarly, by H $pqy \sim_2 px$ and by Q $px \leq_2 ptz + (1 - pt)pqy$. By the convexity of the bargaining problem $tz + (1 - t)y \leq_2 x$ which implies by H that $ptz + p(1 - t)y \leq_2 px$. Now, by DOM $p(1 - t) \leq (1 - pt)pq$ and it is simple to verify the contradiction with $(1 - q(1 - t))p \geq t$. Q.E.D.

LEMMA 2: Consider a pair of preferences (\geq_1, \geq_2) which satisfy conditions DOM, Q, and CCE. Let y^* be a Nash solution outcome of (\geq_1, \geq_2) . If p and x satisfy $x \sim_2 py^*$, then $[1/(2 - p)]x \leq_1 y^*$.

PROOF: Assume the contrary that $[1/(2 - p)]x >_1 y^*$. We will find α and t such that $\alpha[tx + (1 - t)y^*] >_1 y^*$ and $[tx + (1 - t)y^*] >_2 \alpha y^*$. Since by convexity there is $w \in X$ such that $w \geq_i tx + (1 - t)y^*$ for both i , then $\alpha w >_1 y^*$ and $w >_2 \alpha y^*$, a contradiction to y^* being a Nash solution outcome. By property CCE, $tx + (1 - t)y^* \geq_2 tpy^* + (1 - t)y^* = (tp + 1 - t)y^*$ and thus, by the continuity of the preferences and setting $\alpha = tp + 1 - t$, it is sufficient to find t such that $[tp + (1 - t)][tx + (1 - t)y^*] >_1 y^*$. Now, for all α ,

$$\begin{aligned} \alpha[tx + (1 - t)y^*] &= \alpha tx + \alpha(1 - t)y^* \\ &= (1 - \alpha(1 - t))[(\alpha t)/(1 - \alpha(1 - t))x] \\ &\quad + \alpha(1 - t)y^*. \end{aligned}$$

By Q, it is left to show that $(\alpha t)/(1 - \alpha(1 - t))x >_1 y^*$ for some t . This can be shown by using L'hospital's rule to verify that $\lim_{t \rightarrow 0} (\alpha t)/(1 - \alpha(1 - t)) = 1/(2 - p)$ and by our assumption $[1/(2 - p)]x >_1 y^*$. Q.E.D.

REMARK: Notice that in the proof of Lemma 2 we do not use H and use only a weak version of Q which requires that if $L >_i x$ then $\alpha L + (1 - \alpha)x >_i x$. We refer to this property as WQ. Also notice that Lemma 1 can be derived without

condition H, by strengthening condition CCE to CCE*⁷ (a condition which implies WQ):

CONDITION CCE*: *If $x \succcurlyeq_i L'$ then $\alpha x + (1 - \alpha)L \succcurlyeq_i \alpha L' + (1 - \alpha)L$ for every lottery L .*

Thus, both sets of conditions {DOM, Q, CCE, H} and {DOM, CCE*} imply the two lemmas. Since the proof of the main proposition will use the conditions on the preferences only in the sense that they imply the two lemmas, it follows that the analysis in the next section is valid for both sets of assumptions. The reason for emphasizing the first set of conditions will be discussed later.

5. THE NASH THEOREM

We are now ready to prove a theorem analogous to that of Nash (1950). The theorem refers to a more general family of preferences and to the ordinal definition.

PROPOSITION 2: (a) *The Nash solution is well defined for all problems in which the preferences satisfy conditions DOM, Q, and H.*

(b) *Let B be any domain of problems in which the Nash solution is well defined. Then the Nash solution satisfies PAR, SYM, and IIA.*

(c) *Let B be any domain of problems containing all problems ($\succcurlyeq_1, \succcurlyeq_2$) where (\succcurlyeq_i) _{$i=1,2$} are expected utility preferences and in which the preferences satisfy DOM, Q, CCE, and H. Then, the Nash solution is the unique solution on B which satisfies PAR, SYM, and IIA.*

A special case of Proposition 2 is as follows.

PROPOSITION 2*: *For the domain of problems which includes only all expected utility preferences, the (ordinal) Nash solution is well defined and is the unique solution which satisfies PAR, SYM, and IIA.*

⁷ Assume the contrary. Let p and q satisfy $pz \sim_1 x$ and $qy \sim_2 x$; then, $z \succcurlyeq_2 px$ and $y \succcurlyeq_1 qx$. By the continuity of \succcurlyeq_1 , there is a number t such that $tz + (1 - t)y \sim_1 x$. However, $pz \leq_1 x$ and $qx \leq_1 y$, together with CCE*, imply that

$$(t + (1 - t)qp)z = tz + (1 - t)qpz \leq_1 tz + (1 - t)qx \leq_1 tz + (1 - t)y \sim_1 x \sim_1 pz$$

and by condition DOM it follows that $t + (1 - t)qp \leq p$.

On the other hand, the convexity of X implies that $tz + (1 - t)y \leq_2 x$ and using condition CCE* twice more we get

$$(tpq + (1 - t))y = tpqy + (1 - t)y \leq_2 tpx + (1 - t)y \leq_2 tz + (1 - t)y \leq_2 x \sim_2 qy$$

which by DOM implies that $tpq + (1 - t) \leq q$. Summing this inequality with the previous one we arrive at $1 + pq \leq p + q$ which implies $1 - p \leq q(1 - p)$ which is a contradiction given that $q \neq 1$ and $p \neq 1$.

PROOF: (a) This part is proved in three stages:

A Nash solution outcome exists: If this were not so, there would exist a problem such that for all $x \in X$ there is a y such that $y \gg x$. By Lemma 1, for any Pareto-optimal x , it is impossible for both y and z to satisfy $y >_1 x$, $y \gg x$, $z >_2 x$, and $z \gg x$. Thus we can split the Pareto frontier of X into two exclusive sets, V_1 and V_2 , where $V_i = \{x \mid \text{there exists } y \text{ such that } y >_i x \text{ and } y \gg x\}$. Obviously $B_i \in V_j$. Thus V_1 and V_2 are nonempty open sets which cover the compact Pareto frontier which is a contradiction.

A Nash solution outcome is always Pareto optimal. Let y^* be a Nash solution outcome for the pair (\geq_1, \geq_2) . If there exists an x which satisfies $x >_i y^*$ and $x \geq_j y^*$ then for large enough p , $px >_i y^*$ but $x \geq_j y^* >_j py^*$, in contradiction to y^* being a Nash solution outcome for the pair (\geq_1, \geq_2) .

Uniqueness of a Nash solution outcome. Assume that both y and z are Nash solution outcomes of a particular problem and $z >_1 y$. Thus, neither $z \gg y$ nor $y \gg z$. By the convexity of the problem there exists a Pareto-optimal x such that $x \geq_i .5y + .5z$ for both i . It follows (using Q) that $z >_1 x >_1 y$. Since not $[x \gg z]$, Lemma 1 implies that $z \gg y$, in contradiction to y being a Nash solution outcome.

(b) We have already seen that the Nash solution satisfies PAR. Next we show that it satisfies SYM and IIA.

The Nash solution satisfies Symmetry. Assume (\geq_1, \geq_2) is a symmetric problem and ϕ a symmetry correspondence. If $\phi(z) \gg \phi(y)$ then $z \gg y$. Thus, if y^* is a Nash solution outcome then so is $\phi(y^*)$. Therefore $\phi(y^*) = y^*$.

The Nash solution satisfies IIA. Let y^* be the Nash solution outcome for (\geq_1, \geq_2) and let \geq'_i be a preference which preserves \geq_i on X , such that for all $x \geq_i y^*$, if $px \sim_i y^*$, then $px \leq'_i y^*$ and for all $x \leq_i y^*$, if $x \sim_i qy^*$, then $x \sim'_i qy^*$. Then for any p and x for which $px >'_i y^*$ we have $px >_i y^*$ and $py^* \geq_j x$. For any p and x for which $px >_j y^*$, we have in addition $py^* \geq_i x$ and $py^* \geq'_i x$. Thus, y^* is also a Nash solution outcome for the pair (\geq'_1, \geq_2) .

(c) Let y^* be the Nash solution outcome of (\geq_1, \geq_2) . We construct \geq'_1 and \geq'_2 to be expected utility preferences preserving the orderings of \geq_1 and \geq_2 on X defined as follows: For any Pareto-optimal x ,

- (1) if $x <_1 y^*$: $x \sim_1 py^*$ implies $x \sim'_1 py^*$ and $y^* \sim'_2 [1/(2-p)]x$;
- (2) if $x >_1 y^*$: $x \sim_2 py^*$ implies $x \sim'_2 py^*$ and $y^* \sim'_1 [1/(2-p)]x$.

Thus, we can choose von Neumann-Morgenstern utility functions u_1 and u_2 to represent \geq'_1 and \geq'_2 respectively, such that $u_i(D) = 0$ and $u_i(y^*) = 1$ and such that, for all Pareto-optimal x , $u_1(x) + u_2(x) = 2$. The problem (\geq'_1, \geq'_2) is convex and by defining $\phi(x)$ so that $pB_1 \sim'_1 x$ implies $pB_2 \sim'_2 \phi(x)$, we obtain a symmetric problem. From $B_1 \sim_2 Oy^*$ it follows that $1/2B_1 \sim'_1 y^*$ and from $B_2 \sim_1 Oy^*$ it follows that $1/2B_2 \sim'_2 y^*$ which implies $\phi(y^*) = y^*$. By SYM, $F(\geq'_1, \geq'_2) = y^*$. The problems (\geq'_1, \geq'_2) and (\geq_1, \geq_2) satisfy the hypothesis of IIA since by Lemma 2 if $x \sim_i py^*$ then $[1/(2-p)]x \leq_j y^*$. Therefore $F(\geq_1, \geq_2) = F(\geq'_1, \geq'_2) = y^*$. Q.E.D.

COMMENT: We have already noted that conditions DOM and CCE* imply Lemmas 1 and 2 and thus are alternative conditions for Proposition 2 to hold. Though we find DOM and CCE* more appealing conditions than DOM, Q, CCE, and H, we hesitate to adopt them as the conditions in the proposition since we do not have examples of *convex* problems in which the preferences satisfy DOM and CCE*.

COMMENT: In the standard Nash framework Roth (1977) has shown that replacing PAR with Strong Individual Rationality (by which the bargaining outcome is strictly better than the disagreement point) is sufficient to derive the Nash solution. It is interesting to note that in the current setting IIA (as defined here) combined with SYM and strong individual rationality do not characterize the Nash solution. Denote the Nash solution by $N(\succsim_1, \succsim_2)$. The function which assigns to (\succsim_1, \succsim_2) the outcome $x \in X$, which satisfies $x \sim_i \alpha N(\succsim_1, \succsim_2)$, is a proper counterexample where $0 < \alpha < 1$.

COMMENT: The following is an example which demonstrates the role of condition H in deriving the existence of a Nash solution. Assume that $X = [0, 1]$ where $x \in X$ is interpreted as the partition of a dollar where player 1 receives x and player 2 receives $1 - x$, while in D both get 0. Assume that player 2 is an expected value maximizer and that player 1 decomposes a lottery L into $\alpha L' + (1 - \alpha)D$, where L' is a lottery on X , and evaluates L as $f(\alpha)V(L')$ where $V(L')$ is the expected value of L' and f is an increasing function with $f(0) = 0$, $f(1) = 1$, and $f'(1) = 0$. The problem does not have a Nash solution. Obviously the partition $x = 1$ is not a Nash solution since choosing $p > 0$ so that $f(p) < 1/2$ would result in $px <_1 1/2$ and $p(1/2) >_2 1 - x$. For any $x < 1$ there is $y > x$ and p such that $f(p)y > x$ and $(1 - y) = p(1 - x)$, i.e., there exists $y > x$ such that $g(y) = f[(1 - y)/(1 - x)]y > x$ which follows from $g(x) = x$ and $g'(x) = f(1) - [x/(1 - x)]f'(1) = 1$. Intuitively, player 1 does not care about taking small risks and thus for any x , he is able to set $y > x$ and p so that he prefers py to x while player 2 is better off accepting y than remaining with the lottery px .

COMMENT: A different extension of the Nash bargaining solution to a family of non-expected utility preferences (requiring smoothness) is suggested in Safra and Zilcha (1990). This extension is based on the dynamic scenario that bargainers sequentially bargain over "small" pieces of the surplus. At each stage they agree on the regular Nash bargaining solution with respect to local expected utility approximations of the bargainers' preferences.

6. INSIGHTS ON SOME OF THE BASIC RESULTS OF BARGAINING THEORY

Comparative Statics of Risk Aversion

We show here that with respect to the Nash bargaining solution, the more risk-averse the player, the worse his outcome in the bargaining. We say that \succsim'_1 is more risk-averse than \succsim_1 if for any lottery L where $x \succsim_1 L$ for some $x \in X$, then $x \succ'_1 L$.

PROPOSITION 3: If \succsim'_1 is more risk-averse than \succsim_1 , then $N(\succsim'_1, \succsim_2) = x^* \leq_1 y^* = N(\succsim_1, \succsim_2)$ (where $N(\succsim_1, \succsim_2)$ is the Nash solution to the problem (\succsim_1, \succsim_2)).

PROOF: Assume that $x^* >_1 y^*$. Choose Pareto-optimal z such that $x^* >_1 z >_1 y^*$. Since $y^* = N(\succsim_1, \succsim_2)$ not $[z \gg y^*]$ and by Lemma 1 $z \gg x^*$; that is, there is a number p such that $pz >_2 x^*$ and $z >_1 px^*$. Since \succsim'_1 is more risk averse than \succsim_1 , it is also true that $z >'_1 px^*$, a contradiction to x^* being the Nash solution outcome for $(\succsim'_1, \succsim_2)$. Q.E.D.

The Connection with the Strategic Alternating Offers Model

The ordinal definition of the Nash solution makes very clear the relationship between the strategic alternating offers model of Rubinstein (1982) and the Nash solution (see also Roth (1989)). The connection between these two different models was first pointed out by Binmore (1987b). See also Binmore, Rubinstein, and Wolinsky (1986).

Recall the version of the infinite alternating offers model where at the end of each period there is a probability $1 - p > 0$ of breakdown. For the expected utility case the model has a unique subgame perfect equilibrium characterized by two alternatives $x^*(p)$ and $y^*(p)$ satisfying $px^*(p) \sim_1 y^*(p)$ and $py^*(p) \sim_2 x^*(p)$. Player 1 always offers $x^*(p)$ and accepts any alternative y such that $y \geq_1 y^*(p)$, while player 2 always offers $y^*(p)$ and accepts any alternative x such that $x \geq_2 x^*(p)$.

PROPOSITION 4: The alternating offers equilibrium outcomes, $x^*(p)$ and $y^*(p)$, converge to $N(\succsim_1, \succsim_2)$ where $p \rightarrow 1$.

PROOF: There is no $x >_1 x^*(p)$ such that $x \gg x^*(p)$ since if there was, the fact that not $[x^*(p) \gg y^*(p)]$ together with Lemma 1 imply that $x^*(p) \gg x$. Similarly there is no $x >_2 y^*(p)$ such that $x \gg y^*(p)$. Therefore for all p , $x^*(p) \geq_1 N(\succsim_1, \succsim_2) \geq_1 y^*(p)$. For any subsequence (p_n) converging to 1 such that $x^*(p_n)$ and $y^*(p_n)$ converge to x^* and y^* respectively, it has to be true that $x^* \sim_i y^*$ for both i and thus $x^* = y^*$. Furthermore, $x^* \geq_1 N(\succsim_1, \succsim_2) \geq_1 y^*$ and thus the sequences converge to $N(\succsim_1, \succsim_2)$. Q.E.D.

Exact Implementation of the Nash Bargaining Solution

Howard (1988) suggests a game which implements the Nash solution exactly. The following is a variation of Howard's game:

- Phase 1: Player 1 announces $y \in X$.
- Phase 2: Player 2 announces $x \in X$ and $p \in [0, 1]$.
- Phase 3: Nature makes a choice:
 - With probability $1 - p$ the game terminates with D .
 - With probability p the game continues.
- Phase 4: Player 1 chooses between x and the lottery py .

If the preferences satisfy condition H then the only subgame perfect equilibrium outcome of this game is the Nash solution outcome y^* . To see this, note the following:

(i) In all subgames starting in Phase 2 after player 1's announcement y , player 2 announces a number p and an outcome x , such that $py \sim_1 x$. If $py <_1 x$ then player 1 chooses x and player 2 will do better by increasing p (and decreasing the probability that the game ends in Phase 3). If $py >_1 x$, then player 1 chooses py and the lottery p^2y is worse for player 2 than y (which he can guarantee by choosing $x = y$ and $p = 1$).

(ii) After player 1 announces y^* then, in equilibrium, player 2 chooses $x = y^*$ and $p = 1$ since if there is a subgame perfect equilibrium in which player 2 chooses $(x, p) \neq (y^*, 1)$ it has to be that $x \sim_1 py^*$ and $px \geq_2 y^*$. But since y^* is the Nash solution and $x \neq y^*$ then $x \sim_1 py^*$ implies $y^* >_2 px$.

(iii) We will show that if player 1 announces $y >_1 y^*$, the subgame perfect equilibrium outcome must be worse for player 1 than y^* . Since y is not a Nash solution, there exists q and z such that $qz >_2 y$ and $z \sim_1 qy$. In equilibrium if player 1 announces y , player 2 announces z and q . If the outcome after y is better for 1 than y^* , it implies that $qz \geq_1 y^*$. Choose x such that $z >_1 x >_1 y^*$. Let $py \sim_1 x$. Since player 2 chooses (z, q) , it has to be that $qz \geq_2 px$ and by H, $z \geq_2 (p/q)x$. By H, $(p/q)z \sim_1 py$ and thus $(p/q)z \sim_1 x$. Thus, not $[x \gg z]$ and, by Lemma 1, $x \gg y^*$ contradicting the fact that y^* is a Nash solution.

The Kalai-Smorodinsky Solution

The meaning of other solutions can also be clarified by the use of preference language. Consider, for example, the Kalai-Smorodinsky (1975) solution. Translating their definition from utility language to preference language implies that the Kalai-Smorodinsky solution to the problem (X, D, \geq_1, \geq_2) is the Pareto-optimal alternative y^* for which there is a $p \in [0, 1]$ such that both $pB_2 \sim_2 y^*$ and $pB_1 \sim_1 y^*$.

In their axiomatic characterization Kalai and Smorodinsky used an axiom which implies the monotonicity axiom which requires that if $S \supset T$ and if both S and T have the same "ideal point" (the vector whose i th component is agent i 's maximal utility over all x which Pareto-dominates D), then the solution outcome of S Pareto dominates that of T . A straightforward implication of the axiom is the following: if the solution for S is Pareto optimal in T and if S and T have the same ideal point, then $R = S \cup T$ has the same ideal point as S and has the solution to S as a Pareto-optimal point. The solution to R has to dominate both solutions to S and T and therefore the solutions to S and T must coincide. Translating into preference language, we obtain the following:

MON: Assume $F(\geq_1, \geq_2) = y^*$ and assume that, for both i , \geq'_i is a preference relation such that if $p_i B_i + (1 - p_i) B_2 \sim_i y^*$, then $p_i B_1 + (1 - p_i) B_2 \sim'_i y^*$. Then, $F(\geq'_1, \geq'_2) = y^*$.

It is easy to verify that the only solution which satisfies PAR, MON, and SYM is the Kalai-Smorodinsky solution. The result regarding the sensitivity of this solution to risk aversion also follows easily from the ordinal definition.

7. CONCLUSION

In this paper we have provided and analyzed a more verbal interpretation of the Nash bargaining solution. We interpret the solution as a convention which assigns to very bargaining problem an outcome with the following property: if it is worthwhile for one of the players to make a demand for an improvement upon the convention, combined with actions which may cause a breakdown of the negotiations, then it is optimal for the other player to reject the demand and to insist on following the convention even if he takes into account the existence of the hazard conditions.

Although this paper has dealt almost exclusively with the Nash solution, it has a more general message regarding the methodology of formal models in economics. The Nash bargaining theory is typically defined in utility language which allows the use of geometrical presentations and facilitates analysis; however, the parametric presentation results in an unnatural statement of the solution and axioms. The judgment and interpretation of the axioms and bargaining solution is thus made more difficult. The difficulties are even more severe when "technical" assumptions (such as continuity and differentiability) are made. The switch to the alternative-preferences language allows a more natural statement of the Nash solution. It enables us to extend the definition to non-expected utility preferences and helps us to better understand certain well-known results. This transition may also prove beneficial in other areas of economics and game theory.

*Dept. of Economics, Tel Aviv University, Tel Aviv, Israel 69978,
Grad. School of Business Administration, Tel Aviv University, Tel Aviv, Israel 69978,*

and

Dept. of Economics, Rochester University, Rochester, NY 14627, U.S.A.

Manuscript received September, 1990; final revision received December, 1991.

REFERENCES

- AUMANN, R., AND M. KURZ (1977): "Power and Taxes," *Econometrica*, 45, 1137–1161.
 BINMORE, K. (1987a): "Nash Bargaining Theory I," in *The Economics of Bargaining*, ed. by K. Binmore and P. Dasgupta. Oxford: Blackwell, pp. 27–46.
 ——— (1987b): "Nash Bargaining Theory II," in *The Economics of Bargaining*, ed. by K. Binmore and P. Dasgupta. Oxford: Blackwell, pp. 61–76.
 ——— (1987c): "Nash Bargaining Theory III," in *The Economics of Bargaining*, ed. by K. Binmore and P. Dasgupta. Oxford: Blackwell, pp. 239–257.
 BINMORE, K., A. RUBINSTEIN, AND A. WOLINSKY (1986): "The Nash Bargaining Solution in Economic Modelling," *The Rand Journal of Economics*, 17, 176–188.

- CHEW, S. H. (1983): "A Generalization of the Quasilinear Mean with Applications to the Measurement of Income Inequality and Decision Theory Resolving the Allais Paradox," *Econometrica*, 51, 1061-1092.
- CHUN, Y., AND W. THOMSON (1988): "Monotonicity Properties of Bargaining Solutions when Applied to Economics," *Mathematical Social Sciences*, 15, 11-27.
- FISHBURN, P. (1989): *Non-Linear Preference and Utility Theory*. Baltimore: Johns Hopkins University Press.
- HARSANYI, J. (1956): "Approaches To the Bargaining Problem Before and After the Theory of Games: A Critical Discussion of Zeuthen's, Hicks', and Nash's Theories," *Econometrica*, 24, 144-157.
- HOWARD, JOHN (1988): "A Social Choice Rule and its Implementation in Perfect Equilibrium," London School of Economics, D.P. 88/172.
- KALAI, E., AND M. SMORODINSKY (1975): "Other Solutions to Nash's Bargaining Problem," *Econometrica*, 43, 513-518.
- KARNI, E., AND D. SCHMEIDLER (1990): "Utility Theory and Uncertainty," in *Handbook of Mathematical Economics (Volume IV)*, ed. by W. Hildenbrand and H. Sonnenschein. New York: North Holland Pub. Co. (forthcoming).
- LUCE, R. D., AND H. RAIFFA (1957): *Games and Decisions*. New York: Wiley.
- MACHINA, M. (1987): "Choice Under Uncertainty: Problems Solved and Unsolved," *Journal of Economic Perspectives*, 1, 121-154.
- NASH, J. (1950): "The Bargaining Problem," *Econometrica*, 28, 155-152.
- ROEMER, J. (1986): "The Mismatch of Bargaining Theory and Distributive Justice," *Ethics*, 96, 88-110.
- (1988): "Axiomatic Bargaining Theory in Economic Environments," *Journal of Economic Theory*, 45, 1-31.
- (1990): "Welfarism and Axiomatic Bargaining Theory," *Recherches Economiques de Louvain*, 56, 287-301.
- ROTH, A. (1977): "Individual Rationality and Nash's Solution to the Bargaining Problem," *Mathematics of Operations Research*, 2, 64-65.
- (1989): "Risk Aversion and the Relationship between Nash's Solution and Subgame Perfect Equilibrium of Sequential Bargaining," *Journal of Risk and Uncertainty*, 2, 354-363.
- RUBINSTEIN, A. (1982): "Perfect Equilibrium in a Bargaining Model," *Econometrica*, 50, 97-110.
- SAFRA, Z., AND I. ZILCHA (1990): "Bargaining Solutions without the Expected Utility Hypothesis," forthcoming in *Games and Economic Behavior*.
- ZEUTHEN, F. (1930): *Problems of Monopoly and Economic Welfare*. London: Routledge and Kegan Paul, 1930.