

Roth, Alvin E. *Axiomatic Models of Bargaining*, , Springer-Verlag, 1979.
http://kuznets.fas.harvard.edu/~aroth/Axiomatic_Models_of_Bargaining.pdf

Axiomatic Models of Bargaining

By Alvin E. Roth

Suggested citation:

Roth, A. E. *Axiomatic Models of Bargaining*, Lecture Notes in Economics and Mathematical Systems #170,
Springer Verlag, 1979. http://kuznets.fas.harvard.edu/~aroth/Axiomatic_Models_of_Bargaining.pdf

N.B. the page numbers differ slightly from the original printed version.

Preface

The problem to be considered here is the one faced by bargainers who must reach a consensus--i.e., a unanimous decision. Specifically, we will be considering n -person games in which there is a set of feasible alternatives, any one of which can be the outcome of bargaining if it is agreed to by all the bargainers. In the event that no unanimous agreement is reached, some pre-specified disagreement outcome will be the result. Thus, in games of this type, each player has a veto over any alternative other than the disagreement outcome.

There are several reasons for studying games of this type. First, many negotiating situations, particularly those involving only two bargainers (i.e., when $n = 2$), are conducted under essentially these rules. Also, bargaining games of this type often occur as components of more complex processes. In addition, the simplicity of bargaining games makes them an excellent vehicle for studying the effect of any assumptions which are made in their analysis. The effect of many of the assumptions which are made in the analysis of more complex cooperative games can more easily be discerned in studying bargaining games.

The various models of bargaining considered here will be studied axiomatically. That is, each model will be studied by specifying a set of properties which serve to characterize it uniquely.

Only conventional mathematical notation will be used throughout. Thus R^n will denote n -dimensional Euclidean space, $S = \{x|B\}$ will mean that S is the set of elements x such that condition B holds, and $a, b \in S$ will denote that a and b are elements of S . For n -tuples $x, y \in R^n$, $x \geq y$ means that $x_i \geq y_i$ for each $i = 1, \dots, n$; $\sum_{i=1}^n x_i$ denotes the sum $x_1 + x_2 + \dots + x_n$, and $\prod_{i=1}^n (x_i - y_i)$ denotes the product $(x_1 - y_1)(x_2 - y_2) \dots (x_n - y_n)$.

A preliminary version of this material was assembled in the Spring of 1978, when I delivered a series of lectures on this topic at Stanford University. At that time I was a guest of the Institute for Mathematical Studies in the Social Sciences and the Graduate School of Business at Stanford, and it is a pleasure to acknowledge the support which I received from them, as well as the encouragement which I

Roth, Alvin E. *Axiomatic Models of Bargaining*, Springer-Verlag, 1979.
http://kuznets.fas.harvard.edu/~aroth/Axiomatic_Models_of_Bargaining.pdf

received from my colleagues at Stanford, particularly Mordecai Kurz and Robert Wilson. I also benefitted from many discussions with those who attended these lectures, and particularly with J. Cave, H. Imai, P. Milgrom and M. Osborne.

My subsequent work on this material has been supported by a grant from the National Science Foundation to the University of Illinois. I also owe a considerable debt to my co-workers in this field, particularly John Harsanyi, Ehud Kalai, and Roger Myerson, who have helped me to better understand both their work and my own.

My greatest debt, in this endeavor as in others, is to my wife Emilie, to whom this book is dedicated.

Alvin E. Roth
Professor
Department of Business Administration
and Department of Economics
University of Illinois
Urbana, Illinois 61801

Table of Contents

Page

Part I: Nash's Model of Bargaining

Section

A.	Introduction	1
B.	The Formal Model and Axiomatic Derivation	4
	Nash's Theorem	8
	Individual Rationality	12
	Symmetry and Asymmetry	15
C.	Probabilistic Models	20
	Bargaining as a Non-Cooperative Game	20
	Bargaining as a Single Player Decision Problem	25
	A Model of Negotiation	28
D.	Risk Posture	35
	Comparative Risk Aversion	38
	Boldness and Fear of Ruin	49
	Strategic Risk Posture and the Utility of Bargaining	52

Part II: Other Models of Bargaining

Section

A.	A Critical Evaluation of the Independence Properties	61
	Independence of Equivalent Utility Representations	61
	Independence of Irrelevant Alternatives	67
B.	Ordinal Models of Bargaining	68
C.	Interpersonal Models of Bargaining	78
	Proportional Solutions	78
	Ordinal Interpersonal Comparisons	92
D.	"Irrelevant" Alternatives	98
	An Individually Monotonic Solution	98
	Dependence on the Ideal Point	107

Appendix:	Summary of the Principal Properties and Results	110
------------------	--	------------

Bibliography	116
---------------------	------------

AXIOMATIC MODELS OF BARGAINING

by

Alvin E. Roth

Part I: Nash's Model of Bargaining

A. Introduction

One of the simplest yet most fruitful paradigms in cooperative game theory is the pure bargaining problem, in which a group of two or more participants is faced with a set of feasible outcomes, any one of which will be the result if it is specified by the unanimous agreement of all the participants. In the event that no unanimous agreement is reached, a given disagreement outcome is the result. If there are feasible outcomes which all the participants prefer to the disagreement outcome, then there is an incentive to reach an agreement; however, so long as at least two of the participants differ over which outcome is most preferable, there is a need for bargaining and negotiation over which outcome should be agreed upon. Each participant has the ability to veto any outcome different than the disagreement outcome, since unanimity is required for any other result.

The approach to the problem which we shall consider here was first taken in 1950 by John Nash, who concentrated on the two-person bargaining problem. He presented a framework which permitted a unique feasible outcome to be selected as the "solution" of a given bargaining problem. This was in contrast to earlier approaches within the game-theoretic tradition: the von Neumann-Morgenstern [1944] solution to the bargaining problem coincides with Edgeworth's [1881] "contract curve", and is equal to the entire set of individually rational, Pareto optimal¹ outcomes.

Nash's model of bargaining draws heavily on the theory of rational individual choice behavior as initiated by von Neumann and Morgenstern. For situations in which a single individual must choose from a set of feasible alternatives, von

¹These terms will be defined shortly.

Neumann and Morgenstern defined rational behavior as behavior which can be modelled by assuming that the individual has a consistent preference relation over alternatives, and always chooses the most preferred feasible alternative. They considered situations in which the feasible alternatives might involve chance events, and showed that a simple set of consistency conditions on the preferences makes the process of rational choice equivalent to picking the alternative which maximizes the expected value of some real-valued "utility function." The theory of such utility functions has generated a large literature of its own; we give here a brief summary of the theory as presented by Herstein and Milnor [1953].

Expected Utility Theory

A set M is a mixture set if for any elements $a, b \in M$, and for any number $p \in [0,1]$, we can associate another element of M denoted by $[pa; (1-p)b]$ called a lottery between a and b . (This lottery can be interpreted as the event which results in alternative a with probability p , and which with probability $(1-p)$ results in alternative b .) We assume that lotteries have the following properties for all $a, b \in M$:

$$[1a; 0b] = a, [pa; (1-p)b] = [(1-p)b; pa], \text{ and}$$

$$[q[pa; (1-p)b]; (1-q)b] = [pqa; (1-pq)b].$$

The first line is self-explanatory, while the second line states that a compound lottery, whose alternatives may themselves be lotteries, is equivalent to a simple lottery which results in each alternative with the appropriate probabilities.

A preference on M is defined to be a binary relation R such that for any $a, b \in M$ either aRb or bRa must hold, and if aRb and bRc then aRc . We write aPb if aRb but not bRa , and aIb if aRb and bRa . (The relation R is to be interpreted as representing the preferences of an individual; so that aRb means he likes a at least as well as b , aPb means he (strictly) prefers a to b , and aIb means he is indifferent between a and b .) See Richter [1966] for a discussion of what kind of choice behavior implies the existence of such a preference relation. A real valued function u defined on a mixture set M is a utility function

for the preference R if it is order preserving (i.e., if for all $a, b \in M$, $u(a) > u(b)$ if and only if aPb), and if

$$u([pa; (1-p)b]) = pu(a) + (1-p)u(b).$$

That is, the utility of a is higher than the utility of b if and only if a is preferred to b ; and the utility of a lottery is equal to its expected utility.

If R is a preference ordering on a mixture set M , then the following conditions insure that a utility function exists:

For any $a, b, c \in M$, the sets $\{p | [pa; (1-p)b]Rc\}$

and $\{p | cR[pa; (1-p)b]\}$ are closed.

If $a, a' \in M$ and aIa' then for any $b \in M$,

$$[\frac{1}{2}a; \frac{1}{2}b] I [\frac{1}{2}a'; \frac{1}{2}b].$$

The first condition states that the preference relation is continuous as a function of probability, so that if, for instance, a given alternative c is preferred to some lottery, then the preference will not be reversed if an arbitrarily small change is made in the probabilities which define the lottery. The second condition says that if an individual is indifferent between two alternatives a and a' , then he is also indifferent between two lotteries whose only difference is that alternative a' in one lottery is substituted for alternative a in the other.

The utility function is unique up to an order-preserving linear transformation, or in the language of measurement theory, it is uniquely defined up to an interval scale (cf. Krantz, Luce, Suppes, and Tversky [1974]). That is, if u is a utility function for a given preference relation, then v is a utility function reflecting the same preference relation if and only if there exist real numbers c and d with $c > 0$ such that $v = cu + d$. That is, the origin (zero point) and scale (unit) chosen for a utility function are arbitrary, and if $v = cu + d$ then the utility function v and u convey the same information about an individual's preferences, just as Fahrenheit and Celsius temperature scales convey the same information about temperature.

B. The Formal Model and Axiomatic Derivation

Because the rules of the bargaining problem permit the final outcome to be determined only by the coalition of all the participants acting together, or by the individual participants acting alone, the special case of bargaining among two participants shares many of the properties of the general case of bargaining among n participants, for $n \geq 2$. That is, even when n is greater than two, so that intermediate coalitions exist which contain more than one participant but fewer than n , these coalitions cannot by themselves secure any outcome but the disagreement outcome; i.e., they cannot secure any outcome which is not available to their members acting individually. So the effect of intermediate coalitions on the bargaining process is of generally minor importance.¹

Partly for this reason, most of the literature on the bargaining problem has concentrated on the special case when $n = 2$. This reduces complicated notation to a minimum, and for many questions no new techniques would be needed, or new phenomena observed, in the general case. However there are some differences between the special case $n = 2$ and the case of more than two bargainers. For this reason, we shall begin the formal development for the general case of n bargainers.

Following Nash, we assume that each player² has preferences over the feasible outcomes which are represented by a von Neumann-Morgenstern utility function. (Such a representation has, of course, an arbitrary origin and scale.) A particular outcome can then be represented as an n -tuple of real numbers, where the i -th component is the utility of player i for the outcome in question. The set of all feasible outcomes can then be represented as a subset S of R^n (n dimensional Euclidean space). If we assume that the players may agree to randomize between outcomes if they so choose, then the set S will be convex, since von Neumann-Morgenstern utility functions evaluate lotteries at their expected utility.

¹This is not to say that, in formulating a particular bargaining problem, there may not be difficulties concerning the aggregation of individuals. For instance, in various contexts it might be appropriate to consider the participants in the bargaining process to be households, firms, unions, or political parties, rather than the individuals who make up these various organizations.

²We use the terms "player", "participant", and "bargainer" interchangeably.

So a bargaining game is described by a set $N = \{1, \dots, n\}$ of players and a pair (S, d) where S is a compact³ convex subset of R^n representing the feasible utility payoffs⁴ to the players, and d is an element of S corresponding to the disagreement outcome. For simplicity, we will further assume that there is at least one point s in S such that $d < s$ (i.e., $d_i < s_i$ for $i = 1, \dots, n$). This confines our attention to those bargaining problems which offer each player some potential reward for reaching an agreement. Denote by B the set of all such bargaining games.⁵ Unless otherwise stated, we will henceforth assume that the players know all the details of the game, including the preferences of the other players.

Nash defined a solution to the bargaining problem to be a function $f: B \rightarrow R^n$ such that $f(S, d)$ is an element of S for any (S, d) in B . That is, a solution is a rule which assigns to each bargaining game a feasible utility payoff of the game. Thus a solution can be interpreted as a model of the bargaining process.⁶ When no confusion will result, the outcome $f(S, d)$ will sometimes be referred to as the solution of the game (S, d) , even though, strictly speaking, the term 'solution' refers to the function f defined over all games, rather than to the outcome which it selects in a particular game.

Nash proposed that a solution should possess the following four properties.

³A set S is compact in R^n if and only if it is closed and bounded. For instance, if S is generated by allowing randomization of a finite set of alternatives, then it is compact.

⁴We will refer interchangeably to "outcomes", "payoff vectors", and simply "points" in S .

⁵Henceforth all bargaining games will be assumed to be in the class B , unless otherwise noted.

⁶Alternatively, a solution can be interpreted as an arbitration procedure; i.e., a rule which tells an arbitrator what outcome to select. So long as the arbitration procedure is intended to reflect the relative advantages which the game gives to the players, this interpretation need not be at odds with the interpretation of a solution as a model of the bargaining process. But, of course, a (descriptive) bargaining model can be tested empirically, while a (prescriptive) arbitration procedure cannot.

Property 1. Independence of Equivalent Utility Representations: For any bargaining game (S, d) and real numbers a_i and b_i for $i = 1, \dots, n$ such that each $a_i > 0$, let the bargaining game (S', d') be defined by $S' = \{y \in R^n \mid \text{there exists an } x \text{ in } S \text{ such that } y_i = a_i x_i + b_i \text{ for } i = 1, \dots, n\}$ and $d'_i = a_i d_i + b_i$ for $i = 1, \dots, n$. Then $f_i(S', d') = a_i f_i(S, d) + b_i$ for $i = 1, \dots, n$.

This property reflects the amount of information contained in a von Neumann-Morgenstern utility function. If a given player's preferences over the possible outcomes are represented by a utility function u_i , then they are also represented by any utility function $v_i = a_i u_i + b_i$, where a_i and b_i are real numbers, and a_i is positive. The rationale for requiring property 1 is that, since u_i and v_i summarize the same information, the solution f should yield the same underlying outcome if some player's utility function is changed from u_i to v_i . Of course the utility payoff which the model assigns to the player in question at any given underlying outcome q changes from $u_i(q)$ to $a_i u_i(q) + b_i$ (just as changing from Celsius to Fahrenheit temperature changes the temperature of boiling water from 100 to 212).

Property 2. Symmetry: Suppose that (S, d) is a symmetric bargaining game—i.e., suppose that $d_1 = d_2 = \dots = d_n$, and that if x is contained in S , then so is every permutation of x . Then

$$f_1(S, d) = f_2(S, d) = \dots = f_n(S, d).$$

Like the previous property, this one requires that the solution depend only on information contained in the model. In particular, if (S, d) is a symmetric bargaining game, then it gives no information which distinguishes one player from another. The property of symmetry requires that the solution should not distinguish between the players if the model does not.

Property 3. Independence of Irrelevant Alternatives: Suppose that (S, d) and (T, d) are bargaining games such that T contains S , and $f(T, d)$ is an element of S . Then $f(S, d) = f(T, d)$.

This property says that the outcome of bargaining, as identified by the solution f , depends only on the relationship of the outcome to the disagreement point, and does not depend on other alternatives in the feasible set. This property imposes a very different sort of requirement than do Properties 1 and 2, and it can be interpreted as expressing the kind of bargaining which Nash's solution is intended to model. In particular, Property 3 models a bargaining process which can proceed by first narrowing down the original set T of feasible alternatives to some smaller set S , without changing the outcome.

Nash [1950] motivates this property of a solution by saying: "If two rational individuals would agree that $f(T,d)$ would be a fair bargain if T were the set of possible bargains, then they should be willing to make an agreement, of lesser restrictiveness, not to attempt to arrive at any bargains represented by points outside of the set S if S contained $f(T,d)$. If S were contained in T , this would reduce their situation to one with S as their set of possibilities. Hence $f(S,d)$ should equal $f(T,d)$." Another way to state the same relationship is to say that, if a set S of outcomes is enlarged to the set T , then either $f(T,d) = f(S,d)$ or $f(T,d) \notin S$. That is, either the solution selects one of the new alternatives, or else it selects the same alternative as when S was the feasible set.⁷

Property 4. Pareto Optimality: For any bargaining game (S,d) , if x and y are elements of S such that $y \succ x$, then $f(S,d) \neq x$.

This property can be thought of as requiring that the players collectively should behave in a rational way, since it specifies that the solution will select an outcome such that no other feasible outcome is preferred by all of the players. The outcome must be a maximal element of the "social preference" defined by the intersection of all the individual preferences.

⁷ An alternative interpretation of the property is that it models a bargaining process in which only certain kinds of arguments are admissible, or effective. In particular, suppose that, given a fixed disagreement outcome, only binary comparisons between alternate outcomes are permitted, the object being to select the feasible outcome which is in some sense the "best". Then the "best" outcome in a set (i.e., the winner of all the pairwise comparisons) would also be the "best" on any subset of which it was a member, as required by property 3.

The subset of these "collectively rational" outcomes contained in a given set S is called the Pareto optimal⁸ subset, $P(S) \equiv \{x \in S \mid \text{there exists no } y \text{ in } S \text{ for which } y > x\}$. The property of Pareto optimality requires that, for any bargaining game (S, d) , the solution f should always select an outcome contained in $P(S)$.

Nash [1950] showed that the four properties stated above define a unique solution to the bargaining problem. In particular, he proved the following remarkable theorem.

Theorem 1 (Nash's theorem): There is a unique solution possessing Properties 1-4. It is the function $f = F$ defined by $F(S, d) = x$ such that $x \geq d$ and $\prod_{i=1}^n (x_i - d_i) > \prod_{i=1}^n (y_i - d_i)$ for all y in S such that $y \geq d$ and $y \neq x$.

Thus the Nash solution is a function which selects the unique outcome which maximizes the geometric average⁹ of the gains which the players realize by reaching an agreement instead of settling for the disagreement outcome.

Summary of the proof: Before presenting the proof in detail, we can summarize the main ideas as follows. It is easily shown that the solution F described in the theorem is well-defined and possesses the four required properties. It is then necessary to show that any solution f which possesses these four properties must be identical to F .

For any symmetric bargaining game, f and F must coincide, since there is a unique Pareto optimal outcome all of whose components are equal, and Properties 2 and 4 require a solution to pick this outcome. To verify that f and F coincide

⁸After the Italian economist V. Pareto.

⁹Strictly speaking, the geometric average is $\prod_{i=1}^n (x_i - d_i)^{1/n}$, so that we

are actually considering the n -th power of the geometric average. But the difference is of no consequence, since the geometric average and its n -th power are maximized by the same point.

on arbitrary bargaining games, Property 1 implies that it is sufficient to consider games (S', d') normalized so that $F(S', d') = (1, \dots, 1)$ and $d' = (0, \dots, 0)$. But the outcome $(1, \dots, 1)$ is Pareto optimal in a symmetric set A which contains S' . Consequently $f(A, d') = (1, \dots, 1) = f(S', d')$, where the last equality follows from Property 3.

So the mechanism of the proof is that Properties 2 and 4 determine the outcome chosen by a solution in symmetric games, and Properties 1 and 3 permit every game to be treated as a symmetric game.

Proof: The solution described in the theorem is in fact well-defined. That is, there actually is an outcome x in S which maximizes the geometric average of the gains, since the geometric average is a continuous function, and thus it achieves its maximum on the set S , since S is a compact set. Furthermore, this maximum is achieved at a unique point in S . This is because if two different points x and x' in S such that $x, x' \geq d$ yield the same geometric average, then their mean (which is also contained in S , since S is convex) yields a strictly higher geometric average, so that the maximum is not achieved at either x or x' . So the Nash solution is well-defined; we now must show that it possesses the properties required by the theorem.

To see that F is independent of the particular origin and scale chosen to represent each player's utility function, let (S', d') be a bargaining game derived from an arbitrary bargaining game (S, d) by a change of utility representations, as in the statement of Property 1. Then a point y in the set S' has coordinates $y_i = a_i x_i + b_i$, where $x = (x_1, \dots, x_n)$ is the corresponding point in the set S . So $\Pi(y_i - d'_i) = \Pi(a_i x_i + b_i - a_i d_i - b_i) = \Pi a_i \Pi(x_i - d_i)$. That is, the relevant geometric average evaluated at an outcome y in the game (S', d') is simply a positive constant ($c = \Pi a_i$) times the geometric average of the corresponding point in (S, d) . This constant doesn't affect where the geometric average achieves its maximum, so $F_i(S', d') = a_i F_i(S, d) + b_i$, as required.

To see that F possesses Property 2, let (S, d) be a symmetric bargaining game with $x = F(S, d)$. Let y be any permutation of x ; (e.g., let $y_i = x_j$ and $y_j = x_i$ for

some components i and j). Then y is an element of S at which the average gains are equal to those at x . But, as noted previously, the maximum average is achieved at a unique point in S , so $y = x$. Since y was an arbitrary permutation, it follows that $x_i = x_j$ for all i, j , as required.

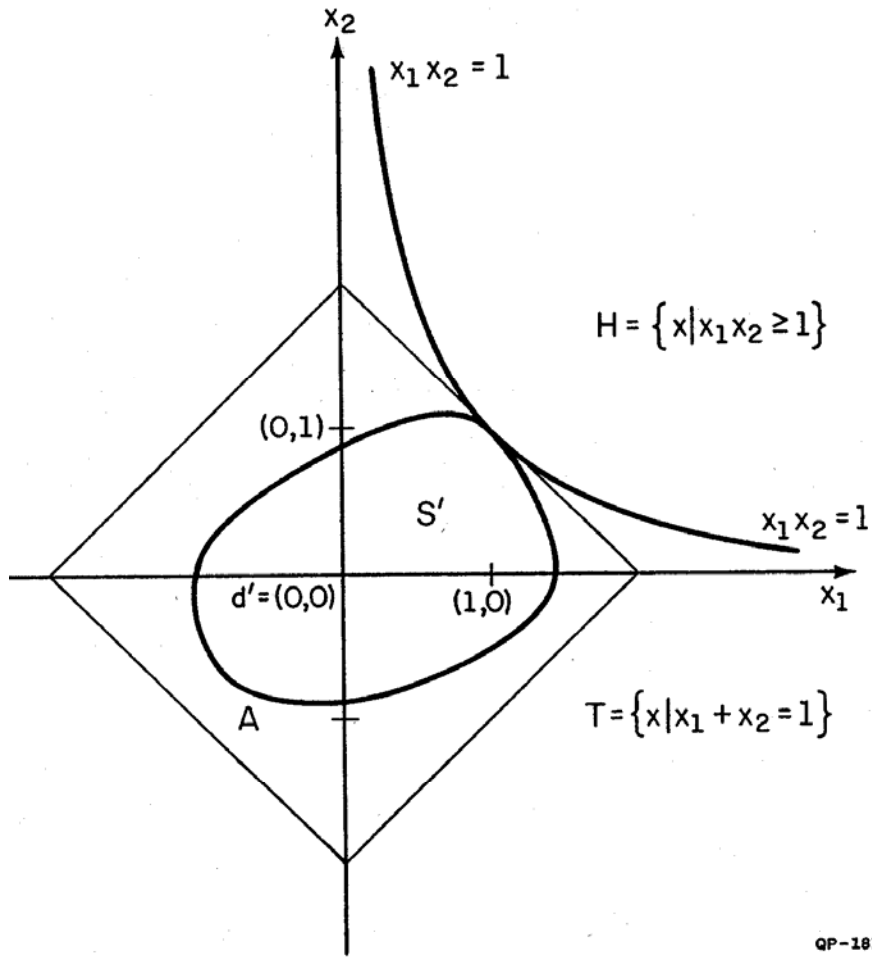
The function F clearly obeys Property 3, since the maximum which a function achieves on a set is always at least as large as the maximum which it achieves on any subset. Equally clearly, F possesses Property 4, since if a feasible payoff vector is not Pareto optimal, then it is majorized by another feasible payoff vector, which yields a higher average gain. Thus the Nash solution possesses each of the four properties. It remains to be shown that any solution f which possesses these four properties must be identical to F .

To see this, we first observe that if f is symmetric and Pareto optimal (i.e., if it possesses Properties 2 and 4) then it coincides with F for any symmetric bargaining game (S, d) and is equal to the unique Pareto optimal point x such that $x_1 = x_2 = \dots = x_n$. (By the definition of Pareto optimality, no bargaining game can contain more than one such point, and such a point is contained in S since S is symmetric, convex, and compact.) It will be convenient to note that, if (A, d) is a symmetric bargaining game whose Pareto optimal set is contained in the hyperplane $\{x \mid \sum x_i = n\}$ then $f(A, d) = F(A, d) = (1, \dots, 1)$.

Let f be a solution which possesses Properties 1-4, and let (S, d) be an arbitrary bargaining game with $z = F(S, d)$; then we need to show that $f(S, d) = z$. Let (S', d') be derived from (S, d) by changing the utility representations so that d is transformed into $d' = (0, \dots, 0)$, and z is transformed into $z' = (1, \dots, 1)$.¹⁰ By Property 1, $f(S, d)$ will be equal to z if and only if $f(S', d') = z'$. Now, $F(S', d') = z' = (1, \dots, 1)$ since F is independent of equivalent utility representations. So z' is the unique point in the intersection of the sets S and H , where H is the set $H \equiv \{x \mid \sum x_i \geq 1\}$. The hyperplane $T = \{x \mid \sum x_i = n\}$ is the unique tangent to H through the point z' ,¹¹ and since both H and S' are convex sets,

¹⁰ That is, each x in S is transformed into x' such that $x'_i = a_i x_i + b_i$ for $i = 1, \dots, n$, where $a_i = 1/(x_i - d_i)$ and $b_i = -d_i/(x_i - d_i)$.

¹¹ The tangent is unique, since the surface $\Pi(x_i - d_i) = k$ is differentiable.



QP-181

Figure 1

The case $n = 2$

the separating hyperplane theorem¹² implies that S' is a subset of $\{x \mid \sum x_i \leq n\}$.

(See Figure 1, p. 11).

Since S' is a compact set, it is contained in a (sufficiently large) symmetric set A , such that $P(A)$ is a subset of the hyperplane T (see Figure 1). As previously observed, $f(A, d) = z'$ since f possesses Properties 2 and 4. Thus Property 3 implies that $f(S', d') = z'$, which completes the proof.

To emphasize the distinct roles which were played by the different conditions, we state the following corollary of the proof.

Corollary 1.1: Suppose that for any symmetric bargaining game (A, d) such that $d = (0, \dots, 0)$ and $P(A)$ is a subset of the hyperplane $T = \{x \mid \sum x_i = n\}$, that f is a solution such that $f(A, d) = (1, \dots, 1)$. Then if f possesses Properties 1 and 3, f is identical to the Nash solution F .

Individual Rationality

In order to model a bargaining problem by the bargaining game (S, d) , we have assumed that each of the players is a rational individual. In particular, we have assumed that each player makes choices (i.e., exhibits preferences) which are sufficiently well behaved so that they can be represented by a utility function, and that this utility function completely summarizes the player's choice behavior over the set of outcomes. To say that an individual's choice behavior is summarized by a utility function is to say that, when he is faced with a choice between two alternatives which have different utilities, he chooses the one with the higher utility.

Of course, in a bargaining game, an individual player isn't free to simply choose any outcome, and the problem of bargaining over outcomes among players with different preferences is quite different from a problem of individual choice. But the rules of the bargaining game are such that any individual can always choose the disagreement outcome and insure that it will be the outcome of the game.

¹²The separating hyperplane theorem states that two compact convex sets which intersect at a unique point have a common tangent through that point.

Consequently, any individual will choose the disagreement outcome instead of agreeing to any outcome which gives him a lower utility. That is, no payoff vector x will be agreed to unanimously if there is even one player i such that $x_i < d_i$. Thus if the game is played by rational players, the only payoff vectors which can possibly arise are those vectors x such that $x \geq d$.

That is, in order to be consistent with the model, in which payoffs in the game are defined in terms of each individual's utility function, any solution must possess the following property.

Property 5. Individual Rationality: For every bargaining game (S, d) , $f(S, d) \geq d$.

This property is implicit in Nash's original treatment of the bargaining problem, and of course Nash's solution is individually rational. Many modern treatments of the subject explicitly include Property 5, along with the other four properties (cf. Owen [1968], p. 141; Harsanyi [1977], p. 196).

However in Nash's formal development of the model, the only reflection of the individual rationality of the players as captured in their utility functions is in Property 1. This property has nothing to do, of course, with the individual choice behavior implied by a utility function; it merely requires that the solution be independent of equivalent utility representations. The requirement of Pareto optimality, on the other hand, is a requirement concerning collective rather than individual choice. In some sense Pareto optimality is the strongest of the four properties, since it requires that the solution select a "good" outcome in every bargaining game. This is a particularly unsatisfying assumption to have to make in constructing a model of the bargaining process, since it eliminates a priori most of the potential outcomes,¹³ including the possibility that a disagreement will occur.

¹³That is, the assumption of Pareto optimality removes from consideration all of the feasible outcomes except for a set of measure zero. Even if we consider bargaining over a finite (rather than a convex) feasible set, it can be shown that the assumption of Pareto optimality generally removes most of the feasible outcomes from consideration (cf. O'Neill [1978]).

Surprisingly, however, the implications of individual rationality are sufficiently far-reaching that it is essentially unnecessary to impose the requirement of Pareto optimality in order to derive Nash's solution. The following theorem makes this precise.¹⁴

Theorem 2: There are precisely two solutions f possessing Properties 1, 2, 3, and 5. One is Nash's solution, $f = F$, and the other is the disagreement solution $f = D$ defined by $D(S,d) = d$ for all bargaining games (S,d) .

Proof: Both F and D possess Properties 1, 2, 3, and 5; we need to show that they are the only functions which do so. Suppose, therefore, that f is a solution which possesses these properties.

Consider the symmetric bargaining game $(\bar{A}, \bar{0})$, where $\bar{A} = \{x \in \mathbb{R}^n \mid x \geq 0, \sum x_i \leq n\}$, and the disagreement point is the origin (i.e., $\bar{0} \equiv (0, \dots, 0)$). The proof will proceed by establishing the following statements.

- i) if $f(\bar{A}, \bar{0}) = (0, \dots, 0)$, then $f(S,d) = D(S,d) = d$ for every (S,d) in B .
- ii) if $f(\bar{A}, \bar{0}) \neq (0, \dots, 0)$, then $f(\bar{A}, \bar{0}) = F(\bar{A}, \bar{0}) = (1, \dots, 1)$
- iii) if $f(\bar{A}, \bar{0}) = F(\bar{A}, \bar{0})$, then $f(S,d) = F(S,d)$ for every (S,d) in B .

To establish statement (i), consider an arbitrary bargaining game (S,d) . Suppose $f(\bar{A}, \bar{0}) = (0, \dots, 0)$, and let A' be the set $A' = \{kx + d \mid x \in \bar{A}\}$, where k is a positive number. Then, since S is compact, we can find k sufficiently large so that A' contains the set S^+ defined by $S^+ = \{x \in S \mid x \geq d\}$. Now Property 1 implies $f(A', d) = kf(\bar{A}, \bar{0}) + d = d$, and Property 3 implies $f(S^+, d) = f(A', d)$, since $d \in S^+$. But Properties 3 and 5 together imply $f(S^+, d) = f(S, d)$, so $f(S, d) = D(S, d) = d$.

To establish statement (ii), suppose that $f(\bar{A}, \bar{0}) = x \neq (0, \dots, 0)$. Since $(\bar{A}, \bar{0})$ is a symmetric bargaining game, Property 2 implies that $x_1 = x_2 = \dots = x_n \equiv k$. Property 5 and the fact that $x \neq (0, \dots, 0)$ imply that $k > 0$, and since the definition of a solution requires that $x \in \bar{A}$, we know that $k \leq 1$. Consider the set $A' = k\bar{A}$, and observe that x is an element of A' , which is a subset of \bar{A} . Property

¹⁴This result first appeared in Roth [1977a]. The presentation here follows the discussion in Roth [1979b].

3 therefore implies that $f(A', \bar{0}) = x$, while Property 1 requires that $f(A', \bar{0}) = kx$. Consequently $k = 1$, and $f(\bar{A}, \bar{0}) = (1, \dots, 1) = F(\bar{A}, \bar{0})$. Statement (iii) follows from Corollary 1.1 of Theorem 1, since Properties 2, 3, and 5 imply that, if $f(\bar{A}, \bar{0}) = (1, \dots, 1)$, then $f(A, \bar{0}) = (1, \dots, 1)$ for any symmetric game $(A, \bar{0})$ whose Pareto optimal subset $P(A)$ is contained in the hyperplane T . Specifically, if A is contained in \bar{A} , then Property 3 implies $f(A, \bar{0}) = f(\bar{A}, \bar{0})$. If \bar{A} is contained in A , then Properties 2 and 5 imply that $f(A, \bar{0})$ is an element of \bar{A} , and so Property 3 implies $f(A, \bar{0}) = f(\bar{A}, \bar{0})$. Thus if $f(\bar{A}, \bar{0}) = F(\bar{A}, \bar{0})$, the assumptions of Corollary 1.1 are met, and so statement (iii) follows. This completes the proof of the theorem.¹⁵

Thus Nash's first three properties admit only two modes of bargaining behavior consistent with individual rationality: one is the behavior which yields Nash's solution, the other is the behavior which yields disagreement in every game.

Symmetry and Asymmetry

In the proof of Nash's theorem we saw that the properties of Pareto optimality and symmetry serve simply to determine the solution on the class of symmetric games (i.e., see Corollary 1.1). In the presence of individual rationality, it becomes unnecessary to require Pareto optimality, and the symmetry property is only used to determine the solution f on the particular game $(\bar{A}, \bar{0})$ (i.e., see statements (i) and (ii) in the proof of Theorem 2). The two independence properties are then sufficient to determine the solution for arbitrary games.

It is easily verified from the proof of part (ii) of Theorem 2 that, if f is a solution different than the disagreement solution D and possessing Properties 1, 3, and 5, then $f(\bar{A}, \bar{0})$ must be Pareto optimal in \bar{A} . We might therefore conjecture that such a solution f can be uniquely determined by specifying $f(\bar{A}, \bar{0})$ to be an arbitrary Pareto optimal point p in \bar{A} . Then, if $f(\bar{A}, \bar{0}) = p = (1, \dots, 1)$, f would be the (symmetric) Nash solution, and for $p \neq (1, \dots, 1)$, we would hope to find f restricted to a unique non-symmetric Nash solution.

¹⁵The game $(\bar{A}, \bar{0})$ used in the proof will come up again and again in our study of the bargaining problem, and we will sometimes refer to it as the "canonical" bargaining game.

This conjecture is almost correct, but not quite. To see how it fails, observe that there are several solutions f obeying Properties 1, 3, and 5 such that $f(\bar{A}, \bar{0}) = (n, 0, \dots, 0)$, for instance. One such solution is $f(S, d) = x$ such that $x = (x_1, d_2, \dots, d_n)$, with x_1 maximal; another such solution is $f(S, d) = x$ such that x is lexicographically maximal in S^+ , where $S^+ = \{x \in S \mid x \geq d\}$. The second of these solutions is Pareto optimal for all games while the first is not; but it is easily verified that both solutions exhibit Properties 1, 3, and 5. For both of these solutions, $f(\bar{A}, \bar{0})$ has some non-positive components, and this, it turns out, is what causes the difficulty.

In particular, for each strictly positive vector $p \in P(\bar{A})$, there is a unique solution f with Properties 1, 3, and 5 such that $f(\bar{A}, \bar{0}) = p$. The function f picks the individually rational point $f(S, d) = x$ which maximizes the weighted geometric average¹⁶ of the gains, using the components of p as weights. That is, we have the following theorem. (For related results, see Harsanyi and Selten [1972], p. 101; Kalai [1977a]).

Theorem 3: For each strictly positive vector p with $\sum p_i = n$, there is a unique solution f possessing Properties 1, 3, and 5, such that $f(\bar{A}, \bar{0}) = p$. For any bargaining game (S, d) , $f(S, d) = x$, such that $x \geq d$ and $\prod (x_i - d_i)^{p_i} > \prod (y_i - d_i)^{p_i}$ for all y in S such that $y \geq d$ and $y \neq x$.

It is easily verified that f exhibits Properties 1, 3, and 5; and the following lemma shows that the description of f given in the theorem is consistent.

Lemma 3.1: The product $\prod x_i^{p_i}$ is maximized on \bar{A} when $x = p$.

Proof: The problem of maximizing $\prod x_i^{p_i}$ on \bar{A} is equivalent to maximizing $\sum (p_i \ln x_i)$ subject to the constraint $\sum x_i = n$. The corresponding Lagrangian is $L(x, \lambda) = \sum p_i \ln x_i - \lambda (\sum x_i - n)$, and the first order conditions are $\partial L / \partial x_i = (p_i / x_i) - \lambda = 0$, so $x_i = p_i / \lambda$ for all $i = 1, \dots, n$. Since $\sum x_i = \sum p_i = n$, it follows that $x = p$. (Since the geometric average is concave, the first order conditions are sufficient to identify a maximum.)

¹⁶ Again, we write the n -th power of the weighted geometric average for notational convenience. (See Footnote 9.)

The proof of the theorem requires only a slight modification of the proof of Theorem 1. For an arbitrary bargaining game (S, d) , let z be the point which maximizes $\prod (x_i - d_i)^{p_i}$ over all x in S^+ . Let (S', d') be the game derived from (S, d) by transforming z into $z' = p$, and d into $d' = \bar{0}$. Then, as in the proofs of Theorems 1 and 2, Properties 1 and 3 establish $f(S', d') = z' = p$, so $f(S, d) = z$, as required.

Nash [1950] originally interpreted the symmetry property as expressing "equal bargaining ability," although he subsequently (Nash [1953]) adopted the view that the property simply describes the information content of the model, as we discussed earlier. These two interpretations are not inconsistent, and in fact the property of symmetry in Theorems 1 and 2 serves precisely the same function that the information $f(\bar{A}, \bar{0}) = p = (1, \dots, 1)$ serves in Theorem 3. If $p \neq (1, \dots, 1)$, then we have some information that the bargaining abilities of the players¹⁷ (or some other factors "outside" the model) are not all equal, and this information is sufficient to uniquely determine the solution.

We saw that, to obtain uniqueness, it was necessary to require that $p > 0$, and the resulting non-symmetric Nash solutions have the property that $f(S, d) > d$. This is a property that is often called strong individual rationality, and is sometimes required as a property in treatments of the Nash solution.

Property 6. Strong individual rationality: For every bargaining game (S, d) , $f(S, d) > d$.

A better name for the above property might be "weak collective rationality," since it actually requires a solution to give each player strictly more than he can obtain on his own. But the name strong individual rationality is often used in the literature, and conveys the similarity in form between this property and Property 5, so we will use it here. Keep in mind, however, that unlike Property

¹⁷ So far as I know, the earliest explicit mention in the literature of the idea that different bargaining abilities might be modelled by non-symmetric Nash solutions of this sort occurs in a remark by Shubik [1959] (p. 350), who attributes the idea to Max Woodbury.

5, this property is not implied by the assumption that the payoffs are given in terms of each individual's utility function.

We have seen by example that ordinary individual rationality (Property 5), together with Properties 1 and 3, is not sufficient to imply Pareto optimality. A consequence of Theorem 3, however, is that strong individual rationality, together with these other two properties, implies the strong form of Pareto optimality stated below. (The proof of the preceding statement requires an argument precisely like that of part (ii) of the proof of Theorem 2, together with Theorem 3.)

Property 7. Strong Pareto Optimality: For any bargaining game (S,d) , if x and y are distinct elements of S such that $y \geq x$, then $f(S,d) \neq x$.

That is, a payoff vector x is strongly Pareto optimal in a set S if no player can obtain a higher payoff than he gets at x unless some other player receives a lower payoff than he gets at x . A payoff vector is (weakly) Pareto optimal, on the other hand, if it is not possible to simultaneously increase the payoff to every player. Any strongly Pareto optimal outcome is of course also weakly Pareto optimal. It is easy to see that the (symmetric or non-symmetric) Nash solutions are in fact all strongly Pareto optimal.

Kalai [1977a] has pointed out that the aggregation of players with identical utility functions leads to non-symmetric solutions in otherwise symmetric situations. For instance, consider an n -person bargaining game in which the players can be partitioned into two groups, having p and q members, respectively, such that within each group the players have identical utility functions. That is, at each outcome x in S , $x_1 = x_2 = \dots = x_p$ and $x_{p+1} = x_{p+2} = \dots = x_n$, where players $1, \dots, p$ are in one group and players $p+1, \dots, n$ are in the other. Then the symmetric Nash solution of the n -player bargaining game yields the same outcomes as the non-symmetric solution, with weights p and q , of the two-player game obtained by modelling each group as a single player.

Note that the symmetric Nash solution applied to the two-player problem in which each group is represented as a single player yields different outcomes, in general, than the symmetric solution of the n -player problem. The question of how to appropriately model a given bargaining problem (e.g., an industry-wide labor-

management dispute involving several unions and firms) is therefore a substantive one, (cf. Footnote 1, p. 4).

When there are more than two bargainers, we can observe some effects of asymmetry which do not occur in the special case $n = 2$. In particular, if f is a solution possessing Properties 1, 3, and 5, then the vector $p = f(\bar{A}, \bar{0})$ conveys some information about the bargaining success of the players, relative to one another, in the game $(\bar{A}, \bar{0})$. The success of a given player i in other bargaining games depends not only on p_i , but on the other components of p as well.¹⁸ (In the case $n = 2$, p_1 determines p .)

An Example

To see this, consider the three-player game $(S, \bar{0})$, where S is the convex hull of the points $(1,1,0)$, $(0,0,1)$ and $(0,0,0)$. Then the set of Pareto optimal outcomes in S is the set $\{(p,p,1-p) \mid 0 \leq p \leq 1\}$. Let f and g be non-symmetric Nash solutions such that $f(\bar{A}, \bar{0}) = (1, 3/2, 1/2)$ and $g(\bar{A}, \bar{0}) = (1, 1/2, 3/2)$.

Then $f(S, \bar{0})$ is the point x in S which maximizes the quantity $x_1 x_2^{3/2} x_3^{1/2}$. On the set S this is the point which maximizes $p^{5/2} (1-p)^{1/2}$, and it is readily verified that this is the point $f(S, \bar{0}) = (5/6, 5/6, 1/6)$. Similarly, g chooses the outcome x maximizing $x_1 x_2^{1/2} x_3^{3/2}$, which is the point $g(S, \bar{0}) = (1/2, 1/2, 1/2)$. The (symmetric) Nash solution F yields the outcome $F(S, \bar{0}) = (2/3, 2/3, 1/3)$.

So although the three solutions f , g , and F all give player 1 the same reward in the game $(\bar{A}, \bar{0})$, they each give him a different reward in the game $(S, \bar{0})$. The difference is due to the relative bargaining success of players 2 and 3 in the game $(\bar{A}, \bar{0})$, and the way this relative success is generalized by solutions which possess Properties 1, 3, and 5.

In particular, players 1 and 2 have identical interests in the game $(S, \bar{0})$, and so it is reasonable that player 1 should benefit when player 2 is an effective bargainer, and suffer when he is not. In this example, each of the solutions considered determines the (common) reward to players 1 and 2 by averaging their "bargaining ability,"³¹ as measured in the outcome of the game $(\bar{A}, \bar{0})$.

¹⁸I included an incorrect assertion on this point in Roth [1978], and corrected it in a subsequent errata.

C. Probabilistic Models

In the previous section we derived Nash's solution to the bargaining problem axiomatically, by specifying some of its properties. In this section, we consider some models which yield Nash's solution from other considerations.

What the models in this section have in common with one another is that they all describe some situation in which the bargainers are faced with a choice under (probabilistic) uncertainty. The Nash solution is then derived as resulting from the maximization of expected utility by each of the bargainers. Each of these models also describes a specific set of rules by which agreements may be reached, unlike the general model of bargaining with which we have been working, which describes only the set of possible agreements.

Only the last of the three models presented in this section has had a significant impact on the development of the theory of bargaining, but each of the three illuminates some aspect of the Nash solution. One of the reasons we include them at this point is to illustrate that, in the theory of bargaining as in other areas of game theory, a variety of approaches can be used to derive a given solution concept.

The models in this section were all originally formulated for the special case of two-person bargaining games, and that is how they will be presented here.

Bargaining as a Non-Cooperative Game¹

In his 1953 paper, Nash presented a model of bargaining as a non-cooperative game, permitting a derivation of the Nash solution which he regarded as comple-

¹A non-cooperative game in strategic form consists of a set of players $N = \{1, \dots, n\}$, a collection of strategy sets, $\Sigma_1, \dots, \Sigma_n$, and a payoff function $G: \Sigma_1 \times \dots \times \Sigma_n \rightarrow R^n$. The interpretation is that each player chooses an element of his own strategy set, and the resulting n -tuple of strategy choices determines the outcome of the game. At the time that Nash wrote, such a game was called non-cooperative if the players had to make their strategy choices without being able to communicate with one another. In the current literature, a game is considered to be non-cooperative if the rules are such that the players must make their strategy choices independently, without being able to conclude a prior binding agreement, while a game is cooperative if the players can conclude a binding agreement as to what outcome should be chosen. (Consequently cooperative games such as the bargaining game are usually not described in strategic form, which emphasizes the individual choices of the players, but are instead described by the set of outcomes which each coalition of players may potentially agree on.)

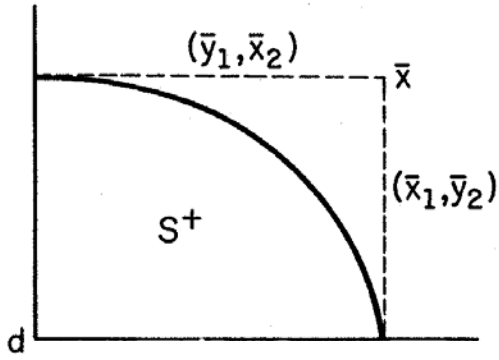
mentary to his earlier axiomatic derivation. In this model, a two-person game (S, d) is played by having each player i simultaneously announce a "demand" x_i . If the demand vector $x = (x_1, x_2)$ is compatible (in a way which we will describe) with the set of feasible outcomes, then each player gets the payoff he demanded. If not, then each player gets his disagreement payoff, d_i .

Specifically, the demand vector x is defined to be compatible with the set of feasible payoffs if there is some feasible outcome s in S such that $s > x$. Essentially the model assumes that utility is disposable, and enlarges the set of outcomes from the set S to the set $S_D \equiv \{x | x \leq s \text{ for some } s \text{ in } S\}$. Nash justified this manner of determining the payoff of the non-cooperative game by saying "It cannot be accused of contributing a bias to the final solution, and it gives the players a strong incentive to increase their demands as much as possible without losing compatibility." This enlargement of the feasible set has the effect of insuring each player that he will not suffer if the other player makes "too low" a demand (see, e.g., Figure 2b, p. 22).² Note that the set S_D , as we have defined it here, is not compact. However the individually rational subset $S_D^+ = \{x \in S_D | x \geq d\}$ is compact.³

For simplicity, we will require each player i to choose a demand x_i which is individually rational and at least potentially feasible: i.e., $d_i \leq x_i \leq \max_{x \in S} x_i$. Formally, player i 's strategy space is the compact interval $E_i = [d_i, \max_{x \in S} x_i]$, and for a strategy pair (x_1, x_2) the payoff function $G_i(x_1, x_2) = x_i \chi(x_1, x_2) + d_i(1 - \chi(x_1, x_2))$, where χ is the indicator function defined by

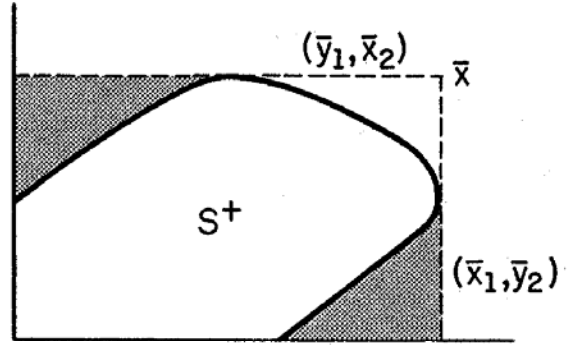
²That is, a pair of demands $x = (x_1, x_2)$ is compatible if neither player would have to lower his demand in order to achieve an outcome in S , i.e., if there is some $y \in S$ with $y \geq x$.

³As long as we are concerned with solutions which possess Properties 3 and 5, S^+ is the only set which concerns us. In any event, the sets S and S_D share the same strong Pareto optimal sets, so any strongly Pareto optimal solution is unaffected by the transition from S to S_D .



(a)

\bar{x} is an equilibrium



(b)

\bar{x} is not an equilibrium
 (the shaded area is contained in S_D
 but not in S)

Figure 2
9

$\chi(x_1, x_2) = 1$ if $(x_1, x_2) \in S_D$, and $\chi(x_1, x_2) = 0$ otherwise. An equilibrium⁴ pair of strategies is a pair (x_1, x_2) such that x_1 is the best response which player 1 could make to x_2 , and x_2 is the best response which player 2 could make to x_1 . That is (x_1, x_2) are an equilibrium if $G_1(x_1, x_2) \geq G_1(y_1, x_2)$ for all $y_1 \in \Sigma_1$, and $G_2(x_1, x_2) \geq G_2(x_1, y_2)$ for all $y_2 \in \Sigma_2$.

Then an equilibrium demand vector is any pair $x = (x_1, x_2)$ which gives neither player an incentive to raise or lower his demand, given the (fixed) demand of the other player. It is not too difficult to show that, since the demands of each player are restricted to be potentially feasible, any equilibrium demand vector x (with one possible exception) must be contained in S_D ; and since the game (S, d) is non-degenerate (i.e., since S contains points which are strictly greater than d), x must be distinct from the disagreement outcome d .

Starting with the exception, consider the "maximal demand point" $\bar{x} = (\bar{x}_1, \bar{x}_2)$ such that $\bar{x}_1 = \max_{x \in S^+} x_1$. Let (\bar{x}_1, \bar{y}_2) and (\bar{y}_1, \bar{x}_2) be the strongly Pareto optimal points in S^+ which achieve the maxima \bar{x}_1 and \bar{x}_2 , respectively. Then, if S^+ is a set such that $(\bar{y}_1, \bar{y}_2) = (d_1, d_2)$, the maximal demand vector (\bar{x}_1, \bar{x}_2) is an equilibrium. This is because $G(\bar{x}_1, \bar{x}_2) = (d_1, d_2)$, while $G_1(y_1, \bar{x}_2) \leq \bar{y}_1$ for all $y_1 \in \Sigma_1$ and $G_2(\bar{x}_1, y_2) \leq \bar{y}_2$ for all $y_2 \in \Sigma_2$. Of course, if $(\bar{y}_1, \bar{y}_2) \neq (d_1, d_2)$, then at least one of the players has the incentive to lower his demand, and secure a strictly larger payoff. (See Figures 2a, b, p. 22).

In a similar way, we see that if x is distinct from \bar{x} , then x cannot be an equilibrium unless $x \in S_D$. But for any demand vector $x = (x_1, x_2)$ contained in S_D , the payoff to the players is equal to their demands, so each player has an incentive to make his demand as high as possible, consistent with keeping x feasible. For x in S_D to be an equilibrium, it must be strongly Pareto optimal; and every strongly Pareto optimal outcome is an equilibrium demand vector. So the set of

⁴The notion of equilibrium for non-cooperative games was introduced by Nash [1951], and is sometimes referred to as "Nash equilibrium." We shall avoid this designation, so as not to invite confusion with the Nash solution which is our main subject. This equilibrium concept remains to this day the principal tool used in the analysis of non-cooperative games, although the question of how to interpret the set of equilibria in a game still presents unresolved difficulties.

equilibria is equal to the set of strongly Pareto optimal demand vectors, plus possibly the additional point \bar{x} .

In order to select a single equilibrium from this infinite set,⁵ Nash proposed to "smooth" the payoff function G , by replacing the discontinuous indicator function χ with a continuous approximation h ; such that h equals χ on S_D , but then drops off to zero in a continuous way. Nash proposed that the function h be interpreted as the probability that a given pair of demands would be in S_D , so that it can be thought of as introducing some uncertainty into the structure of the game. The smoothed payoff function is the expected value, $G'_i(x_1, x_2) = x_i h(x_1, x_2) + d_i(1 - h(x_1, x_2))$. For simplicity, we will normalize the disagreement payoff to $d_i = 0$, so that $G'_i(x_1, x_2) = x_i h(x_1, x_2)$ for $i = 1, 2$.

Thus in the smoothed game, a pair of demands (x_1, x_2) are in equilibrium if and only if x_1 maximizes $x_1 h(x_1, x_2)$ for fixed x_2 , and x_2 maximizes $x_2 h(x_1, x_2)$ for fixed x_1 . A given function h may yield more than one equilibrium, but Nash observed that there is a unique point in S which can be obtained as an equilibrium in the limit, as h approaches χ , regardless of the form of h .⁶ (Note that the maximal demand vector \bar{x} need not be an equilibrium in the smoothed game, unless it is feasible.)

In particular, let $x = (x_1, x_2)$ be a demand pair which maximizes the product $x_1 x_2 h(x_1, x_2)$ over all demand pairs. Then x is an equilibrium since $x_1 h(x_1, x_2)$ and $x_2 h(x_1, x_2)$ are maximal for fixed x_2 and x_1 , respectively. Let z be the point which maximizes $z_1 z_2$ on the set S_D ; i.e., $z = F(S_D, \bar{0}) = F(S, \bar{0})$. Then $x_1 x_2 \geq z_1 z_2$, since $h(x_1, x_2) \leq h(z_1, z_2) = 1$. But z is the only point common to both the set S_D and the set $H = \{x | x_1 x_2 \geq z_1 z_2\}$, so as h approaches χ , z is the limit point of any equilibrium x of the kind described here.

⁵If the point \bar{x} is feasible, then it is the unique equilibrium. Otherwise, the set of Pareto optimal outcomes is an infinite set.

⁶That is, depending on the nature of the limiting process, a given smoothing function h could yield more than one point in S as the limit of equilibria. But there is a unique point in S which is always obtained as such a limit.

The chief criticism of this argument (cf. Luce and Raiffa [1957], pp. 141-2) is that, although it identifies the Nash solution as an equilibrium with characteristics which uniquely distinguish it from other equilibria, it is not clear why these distinguishing characteristics should be relevant to the bargainers.

Bargaining as a Single-Player Decision Problem

Now we will consider a simple model of the two-player bargaining problem in which the Nash solution results if each player maximizes his expected utility under the assumption that the other player's behavior can be described by a probability distribution. The first part of the model as we present it here was proposed independently by Anbar and Kalai [1978], by Butterworth [1977], and briefly by Stone [1958]; our style of exposition follows that of Anbar and Kalai.

The rules of this model are the same as those in the last—a game (S, d) is played by having each player $i = 1, 2$ submit a demand x_i such that $d_i \leq x_i \leq \bar{x}_i$, where $\bar{x}_i = \max_{x \in S} x_i$, as before. If the demand vector $x = (x_1, x_2)$ is contained in the set S_D , then each player receives his demand x_i , otherwise the outcome is $d = (d_1, d_2)$.

Before proceeding further, it will be convenient to note that, in a two-player game S_D with disposable utility, the set of individually rational and (strongly) Pareto optimal points can be described as a function of one player's utility: e.g., let $\phi(x_1) = \max_{(x_1, x_2) \in S_D} x_2$ for any x_1 in the interval $[d_1, \bar{x}_1]$. Then all points of the form $(x_1, \phi(x_1))$ are Pareto optimal in S_D^+ , and all of the strongly Pareto optimal points are of this form. (Note that it makes essentially no difference which player's utility is taken to be the independent variable: the Pareto optimal set could also be defined as the set of points $(\psi(x_2), x_2)$, where ψ is defined analogously for $x_2 \in [d_2, \bar{x}_2]$.) The outcome of the game selected by the Nash solution can now be described as the point in S^+ which maximizes the product $(x_1 - d_1)(\phi(x_1) - d_2)$, or equivalently, as the point which maximizes $(\psi(x_2) - d_1)(x_2 - d_2)$.

Suppose now that player 1, say, believes that player 2 will make a random demand, chosen from a uniform distribution on $[d_2, \bar{x}_2]$. Then player 1, who is by

assumption a rational player (i.e., one who makes choices so as to maximize his expected utility) is faced with the problem of choosing his own demand.

For any demand x_1 , his expected reward is x_1 times the probability that (x_1, x_2) is in S_D , plus d_1 times the probability that it isn't: we'll denote this by $x_1 \Pr \{(x_1, x_2) \in S_D\} + d_1 \Pr \{(x_1, x_2) \notin S_D\}$. Since S_D is a set with disposable utility, the expected reward is $x_1 \Pr \{x_2 \leq \phi(x_1)\} + d_1 \Pr \{x_2 > \phi(x_1)\}$, and since x_2 is uniformly distributed on $[d_2, \bar{x}_2]$, this equals $x_1 [(\phi(x_1) - d_2)/(\bar{x}_2 - d_2)] + d_1(1 - [(\phi(x_1) - d_2)/(\bar{x}_2 - d_2)])$. So player 1 wants to choose his demand x_1 so as to maximize his expected reward, which we can rewrite as

$$[x_1 - d_1][(\phi(x_1) - d_2)/(\bar{x}_2 - d_2)] + d_1.$$

Since $(\bar{x}_2 - d_2)$ is a positive constant, player 1 maximizes his reward by choosing x_1 to maximize $(x_1 - d_1)[\phi(x_1) - d_2]$. That is to say, he maximizes his expected utility by demanding his share of the outcome picked by the Nash solution. If player 2 follows the same line of reasoning, then the demand vector is equal to the (feasible) point $z = F(S, d)$, which thus becomes the final outcome of the game.

As presented so far, this model is not in the game-theoretic tradition, since each player is assumed to behave as if the other were a random process rather than a utility-maximizer. That is, the Nash solution is the result only when each player makes an incorrect assumption about the behavior of the other. The model can, however, be restated⁷ in a way which does not require each player to act as if he were in possession of analytical abilities of a sort unavailable to the other player.

Consider the problem faced by player 1 in the game (S, d) , under the assumption that he knows player 2's preferences. Suppose that player 2 has the same sort of information about player 1, and that each player is aware of all the information in the possession of the other.

Unless he possesses a "theory" about rational behavior in game situations, it is still not clear that all of this information helps player 1 to decide what

⁷This restatement is intended merely to be suggestive and should not be viewed as an attempt to place this model on a rigorous game theoretic foundation. Some of the difficulties to be expected in this latter task will be discussed.

to demand. Since he wants his demand to be compatible with player 2's demand, he needs to have some idea of how player 2 will behave. But player 2's behavior depends on how he thinks player 1 will behave, and so the circle is complete.

We may suppose then, that being able to reach no conclusion about player 2's behavior, player 1 decides (by the "principle of insufficient reason")⁸ to make his own decision as if player 2 chose his demand from a uniform distribution on $[d_2, \bar{x}_2]$. We have already seen that player 1's optimal demand under this assumption is to demand his share of the outcome $(z_1, z_2) = F(S, d)$, and we can now carry this analysis a step further. Player 1 knows that player 2 is rational, and can also carry out this line of reasoning. Consequently, he can now consider a "second approximation" of player 2's behavior, based on the conclusion which he expects player 2 to reach. This second approximation is that player 2 will demand z_2 . Based on this second approximation, z_1 is still the optimal demand for player 1, and so z_2 will still be optimal for player 2, and so player 1's "third approximation" of player 2's behavior will be the same as the second. That is, we have described a process by which player 1, starting from a model of player 2's behavior, revises this model by attributing to player 2 the same line of reasoning which he (player 1) has pursued at each stage. We saw that the initial expectation of a uniform distribution on the interval $[d_1, \bar{x}_1]$ leads to convergence in three iterations.

The critical part of any such theory of "convergent expectations" is obviously the initial expectations of one player about the behavior of the other. Not only do some initial expectations fail to lead to convergence at all, but, more critically for the construction of a satisfactory theory, different initial expectations which do lead to convergence generally converge to different final

⁸ Loosely stated, the principle of insufficient reason states that, when no information is available about a set of possibilities, each possibility should be considered equally likely (there being "insufficient reason" to distinguish one from the other). Of course, any conclusions reached in this way depend in a critical way on how the set of possibilities is specified (cf. Footnote 9).

expectations.⁹ Since the process of convergence is thought of as going on independently in the minds of each of the players, different initial expectations would lead to incompatible actions. So any full-fledged theory of games based on this idea of convergent expectations must begin with an unambiguous specification of how initial expectations are to be arrived at.¹⁰

A Model of Negotiation

Undoubtedly the most famous of the models in this section is the negotiation model originally proposed by Zeuthen [1930] and subsequently extended and generalized by Harsanyi [1956], who was the first to appreciate the relationship between Nash's work and Zeuthen's. The key respect in which the rules of this model differ from those of the previous two, is that, instead of making just one demand, players can adjust their demands over a number of periods, by making concessions when their demands are incompatible. Specifically, the rules by which the game (S, d) is to be played are as follows. At each period, player 1 names an outcome $x = (x_1, x_2)$ in S , and player 2 names an outcome $y = (y_1, y_2)$. For simplicity, we will assume that both x and y are individually rational and Pareto optimal. If each player's demand can be met, i.e. if $(x_1, y_2) \in S_D$, then the game ends, with each player receiving his demand; i.e. the final payoff is (x_1, y_2) . Since both x and y are Pareto optimal, this occurs only when $x_1 \leq y_1$ and $y_2 \leq x_2$; i.e. when each player is offered at least as much as he demands. Otherwise, the game continues for at least one more period, and each player has an opportunity to propose a new outcome, by reducing his own demand, and increasing the payoff which he offers to the other player. The game ends with agreement at any period

⁹ For instance, instead of assuming a uniform distribution on $[d_1, \bar{x}_1]$, we could have incorporated Pareto optimality into our argument and argued for a uniform distribution on $[\bar{y}_1, \bar{x}_1]$, where \bar{y} is defined as in the discussion of the previous model. It is easy to verify that this leads to convergent final expectations different in general from the Nash solution.

¹⁰ A general theory of games based on the idea of convergent expectations has been proposed by John Harsanyi and Reinhard Selten. It is reported in Harsanyi [1975], [1978].

in which the demands of the two players are compatible, in which case they each receive their final demands; or else the game ends with disagreement at any period in which (prior to reaching agreement) neither player makes a concession from his previous demand, and in this case the players each receive their disagreement payoffs (d_1, d_2) . (In order to insure that the game does end in finitely many periods, we will require that if a player does wish to reduce his demand, he must reduce it by at least some (fixed) minimum amount.)

The principle object of Zeuthen's analysis is to determine, at any given period in which agreement has not yet been reached, which player should make the next concession. He argues that, at any such period, the player less willing to face the risk of conflict will be forced to concede. In particular, his analysis proceeds as follows.

Let $x = (x_1, x_2)$ and $y = (y_1, y_2)$ be the (incompatible) proposals of players 1 and 2 respectively. Suppose that player 1, say, considers only two alternatives: either he should capitulate completely, and accept the amount y_1 which player 2 has offered him, or else he should refuse to make any concession, and not lower his demand x_1 at any subsequent period. If he chooses the first alternative, then the game ends for certain at the next period, and his final payoff is y_1 . If player 1 chooses the second alternative, then the final outcome of the game depends on whether or not player 2 ultimately accepts the offer x_2 or not. If player 2 accepts x_2 , then player 1's utility will be x_1 , otherwise it will be d_1 . If p_1 is player 1's subjective probability that player 2 will not accept the offer x_2 , then his expected utility for choosing his second alternative is $p_1 d_1 + (1 - p_1)x_1$. Thus player 1 gets a higher expected utility from holding firm than from capitulating if and only if $p_1 d_1 + (1 - p_1)x_1 > y_1$, or equivalently, if and only if $p_1 < (x_1 - y_1)/(x_1 - d_1)$.

Define the quantity $(x_1 - y_1)/(x_1 - d_1) \equiv \bar{r}_1$ to be player 1's risk limit.¹¹ Only if player 1 believes the probability of a conflict (i.e., the probability p_1

¹¹In order to avoid cumbersome notation, we have suppressed the dependence of the risk limit \bar{r}_1 on the current offers x and y under consideration, but this dependence should not be overlooked, as it is critical to the subsequent argument.

that player 2 will refuse to capitulate) is less than \bar{r}_1 , does he prefer to hold firm to his demand x_1 rather than capitulate and accept y_1 . In precisely the same way, we define player 2's risk limit (with respect to the outcomes x and y) to be the quantity $\bar{r}_2 = (y_2 - x_2)/(y_2 - d_2)$. Thus the risk limit \bar{r}_1 represents the highest subjective probability of a conflict that player 1 would be willing to face and still insist on obtaining his own terms, rather than accept the terms offered to him by the other player.

The central assumption of the model is that player 1 will make some concession at the next period if and only if $\bar{r}_1 \leq \bar{r}_2$. Intuitively, think of the players sitting around the bargaining table after having made their most recent (incompatible) proposals. As time passes, the atmosphere grows more tense, and the probability of conflict rises, until one of the players (whose risk limit has been reached) breaks the deadlock by making a new proposal, with a more modest demand.

To see the implications of this assumption as it relates to the Nash solution, consider the case in which player 1's most recent proposal is $x = (x_1, x_2)$, and player 2's most recent proposal is $y = (y_1, y_2)$, with $x_1 > y_1$ and $y_2 > x_2$. Then Zeuthen's assumption is that player 1 will make a concession at the next period if and only if $\bar{r}_1 \leq \bar{r}_2$, i.e. if and only if $(x_1 - y_1)/(x_1 - d_1) \leq (y_2 - x_2)/(y_2 - d_2)$. But this inequality is equivalent to the inequality

$$\frac{(x_1 - d_1) + (d_1 - y_1)}{x_1 - d_1} \leq \frac{(y_2 - d_2) + (d_2 - x_2)}{y_2 - d_2},$$

which is equivalent to

$$(x_1 - d_1)(y_2 - d_2) + (d_1 - y_1)(y_2 - d_2) \leq (x_1 - d_1)(y_2 - d_2) + (x_1 - d_1)(d_2 - x_2).$$

Eliminating common terms and multiplying both sides by -1 , we conclude that $\bar{r}_1 \leq \bar{r}_2$ if and only if $(y_1 - d_1)(y_2 - d_2) \geq (x_1 - d_1)(x_2 - d_2)$. Thus according to Zeuthen's assumption, player 1 will not make a concession at the next period if and only if his most recent proposal yields a higher (geometric) average gain to the players than does player 2's most recent proposal. (In particular, a player who has proposed the outcome picked by the Nash solution will never make a concession.)

The second critical assumption of the model is that, when a player does decide to make a concession, he will make the smallest possible concession such that, in the following period, the other player will have to respond by also making a concession. Thus, if player 1, say, decides to make a concession from his last proposal x , he does not necessarily accept his opponent's last proposal y , but instead proposes an outcome z between x and y such that

$(z_1 - y_1)/(z_1 - d_1) > (y_2 - z_2)/(y_2 - d_2)$. As we have seen, this is equivalent to the inequality $(z_1 - d_1)(z_2 - d_2) > (y_1 - d_1)(y_2 - d_2)$.

Thus Zeuthen's analysis¹² leads to the conclusion that negotiations will result in a sequence of proposals, each yielding higher (geometric) average gains and terminating in agreement only when no feasible proposal yields still higher average gains. The final agreement can be made to approximate Nash's solution as closely as desired, by making the size of the minimal acceptable concession sufficiently small.¹³

A number of criticisms have been made of this line of analysis.¹⁴ One of the more frequently heard of these is that the players are assumed to make their decisions by evaluating only the two alternatives involving total capitulation by one of the players to the demands of the other (e.g., see the discussion leading to the definition of the risk limits \bar{r}_1 and \bar{r}_2). But when the time comes to act, the model predicts that in fact one of the players will make some intermediate concession, short of capitulation. Especially since we are asked to contemplate a sequence of such small concessions, it is difficult to argue that rational players should persist in reaching their decisions based on criteria concerned exclusively with large hypothetical concessions.

¹²As interpreted by Harsanyi [1956].

¹³Harsanyi ([1977], p. 152) assumes that no player will ever make a concession which overshoots the Nash solution, and that the minimal acceptable concession only applies when the demands are sufficiently far from the Nash solution, so that the outcome of this negotiation process is always identical to the Nash solution in his treatment.

¹⁴See e.g., Bishop [1963], [1964]; Pen [1952], [1959]; Saraydar [1965], [1971a], [1971b]; Wagner [1957], [1958], and Wagner [1979].

Harsanyi [1977] proposes a simple modification of the model which avoids this particular criticism, and which consequently casts the essential assumptions of the model in clearer terms. He defines the "compressed Zeuthen model" of bargaining to be a game similar to the one just described, but which ends after at most two periods. In the first period, each player simultaneously proposes an outcome in S , as in Zeuthen's model. If player 1's proposal x and player 2's proposal y are compatible (i.e., if $x_1 \leq y_1$ and $y_2 \leq x_2$), then the game ends with the utility payoff (x_1, y_2) . Otherwise, the game continues into the second (and last) period, in which each player's only option is to either repeat his own proposal, or accept the proposal of the other player. That is, in the second period, player 1 must make a proposal x' equal either to x or to y , and player 2 must likewise make a proposal y' equal to x or to y . If the proposals x' and y' are compatible (i.e., if at least one of the players capitulates to the other's proposal), then the game ends with the utility payoff (x'_1, y'_2) . Otherwise the game ends with the disagreement payoff (d_1, d_2) . In this game, unlike the previous game, no intermediate concessions are allowed. Therefore, in this game, the risk limits \bar{r}_1 and \bar{r}_2 are derived from the actual choice facing each player in the second period.

Suppose we again make the assumption that a player i will make a concession in the second period if and only if $\bar{r}_i \leq \bar{r}_j$. Given this assumption, the payoff to each player is determined by the proposals made in the first period. We can consider then, the one-period "constrained bargaining game," in which the players are constrained to act in accordance with this assumption. In the constrained game, players 1 and 2 select proposals x and y , and player 1's payoff is x_1 if $\bar{r}_1 > \bar{r}_2$, and y_1 otherwise; while player 2's payoff is y_2 if $\bar{r}_2 > \bar{r}_1$, and x_2 otherwise. Each player's strategy set is equal to the set of individually rational and Pareto optimal outcomes in S , and it is not difficult to see that the Nash solution is the unique equilibrium proposal. That is, we can state the following result.

Proposition: In the constrained bargaining game (S,d) , the unique equilibrium pair of proposals (x,y) is the pair such that $x = y = F(S,d)$.

Proof: Letting $z = F(S,d)$ we observe that if either player i proposes the outcome z , then his payoff in the constrained game will be z_i , regardless of the proposal of the other player. This is because, if the initial proposals of the two players are compatible, then each player receives his demand (so player i receives z_i), and if the initial proposals are incompatible, the other player will have to capitulate (since z maximizes the product $(z_1 - d_1)(z_2 - d_2)$ on S). Furthermore, if player i proposes z , then player j cannot receive a higher payoff than z_j , for the same reason. Thus the proposals $x = y = z$ are in equilibrium. (In fact, z is a maximin proposal for either player.)

To see that no other pair of proposals is in equilibrium, observe that if player 1, say, proposed a Pareto optimal outcome x with $x_1 < z_1$ then player 2's best response would be to also propose the outcome x . But this pair of proposals is not in equilibrium, since we have already seen that player 1 can assure himself of z_1 , regardless of player 2's proposal. Similarly, if player 1 proposes x with $x_1 > z_1$, then any counterproposal y on the open set on which $(y_1 - d_1)(y_2 - d_2) > (x_1 - d_1)(x_2 - d_2)$ will yield player 2 a payoff greater than x_2 . Since this set is open, there is no equilibrium of the form (x,y) , since player 2 has no best response to x . In any event, player 2 has a choice of responses to such a proposal x which leave player 1 worse off than if he had proposed z .

Thus the compressed Zeuthen model leads to the outcome $z = F(S,d)$ as the unique equilibrium outcome if we accept Zeuthen's assumption that a player will make a concession in the second period if and only if his risk limit is not higher than the other player's. But the assumption that the players are expected utility maximizers implies that a player will make a concession whenever his subjective probability of a conflict is higher than his risk limit. So Zeuthen's

assumption is really an assumption about the subjective probabilities which the players assign to a conflict.¹⁵

For our purposes, the most interesting feature about this model is that it leads to the Nash solution without explicit mention of either Properties 1 or 3; i.e., without specifically requiring independence of equivalent utility representations or independence of irrelevant alternatives. Since these are both properties of the solution, they must have been implicitly introduced into the model in some form.

Independence of equivalent utility representations results from the fact that the model only makes use of expected utility calculations, which are invariant under equivalent utility representations. In particular, the risk limit of each player is defined to be the ratio of two utility differences--the ratio of the loss from making a concession to the loss from ending in disagreement. These ratios are preserved under positive linear transformations.¹⁶ Similarly, independence from irrelevant alternatives arises because, given the fixed disagreement outcome, the model only considers binary comparisons between alternative proposals x and y . (Cf. Footnote 7 on page 7.)

¹⁵ Harsanyi [1977] argues that this assumption can be justified through some additional axioms.

¹⁶ Scales of measurement which are unique up to positive linear transformations are called interval scales because such transformations preserve the ratios of intervals.

D. Risk Posture

So far we have described bargaining problems in terms of the set of feasible utility payoffs available to the players, without talking about the particular bargains which might yield those utilities. This permitted us to summarize in a concise way each player's preferences over all possible bargains, including those which might involve chance events (i.e., lotteries), without explicitly considering those lotteries.

In order to consider an individual's attitude towards risk,¹ it will be necessary to explicitly consider the underlying alternatives over which the bargaining is conducted, in order to be able to compare his attitudes towards risky alternatives (i.e., lotteries) and towards alternatives which are free of risk. For instance, suppose that the bargaining is to determine the distribution of a certain sum of money, and consider the expected utility function u_i expressing individual i 's preferences for money. We will assume that an individual's utility function for money is increasing (i.e., more money is preferred to less), and that it has both first and second derivatives. An individual i is said to be risk averse if he prefers to retain his current wealth w for certain rather than to participate in a fair gamble in which he would either win or lose an amount c . That is, he prefers the (riskless) alternative which gives him wealth w for certain to the (risky) lottery $[1/2(w + c); 1/2(w - c)]$. Since these two alternatives have the same expected monetary value, the individual's preference reflects his aversion to the risk involved in the second alternative.

In terms of individual i 's expected utility, this preference can be written $u_i(w) > 1/2u_i(w + c) + 1/2u_i(w - c)$. This is equivalent to $u_i(w) - u_i(w - c) > u_i(w + c) - u_i(w)$, so that, for a risk averse individual, the utility increments corresponding to equal increments of wealth are decreasing as the wealth increases. Dividing both sides of the above inequality by c , and taking the limit as c goes to zero, we conclude that if u_i is a risk averse utility function then its first

¹The notion of risk posture considered in this section is motivated differently than the risk limit considered in the Harsanyi-Zeuthen negotiation model, but they are not entirely unrelated (cf. Proposition 5.1, p. 46).

derivative u_1' is a strictly decreasing function of wealth, or equivalently that its second derivative u_1'' is negative. Formally, a utility function is strictly risk averse if it is strictly concave, i.e., if $u_1''(w) < 0$ for all w .² It is (weakly) risk averse if $u_1''(w) \leq 0$ for all w . Thus an individual who is averse to fair gambles regardless of his current wealth w is also an individual who would prefer to have the expected value of any lottery for certain, rather than to participate in the lottery. That is, a risk averse utility function is concave, so $u(pa + (1-p)b) \geq pu(a) + (1-p)u(b) = u([pa; (1-p)b])$ for any prizes a and b and probability p .

Henceforth, unless otherwise indicated, we will assume that any utility function for money is not only increasing, but also at least weakly risk averse. So a utility function for money u_1 will have $u_1'(w) > 0$ and $u_1''(w) \leq 0$ for all w (see Figure 3, p. 37).

We can then consider the bargaining game which results when two individuals, whose wealth is w_1 and w_2 , bargain over how to divide a positive quantity Q of money. The rules of the game are that they may split the money in any way they like if they both agree, but that neither player will get any money if they don't reach an agreement. The resulting bargaining game is defined in terms of the utility payoffs (rather than the money payoffs) available to the players: it is the game (S^*, d^*) defined by

$$d^* = (u_1(w_1), u_2(w_2)) \text{ and}$$

$$S^* = \{(x_1, x_2) \mid x_i = u_i(w_i + c_i) \text{ for some } c = (c_1, c_2) \text{ such that} \\ c_1 + c_2 \leq Q \text{ and } c_i \geq 0 \text{ for } i = 1, 2\} .$$

The pair (S^*, d^*) is a bargaining game of the kind we have been considering, since the concavity of the utility functions u_i insures the convexity of the set S ,³

²That is, we reserve the term "risk averse" for an individual who is averse to fair gambles regardless of his current wealth.

³This follows from the fact that, if u is a concave function, then $u(pc + (1-p)c') \geq pu(c) + (1-p)u(c')$ for all $p \in [0,1]$. (This is often taken to be the definition of concavity.)

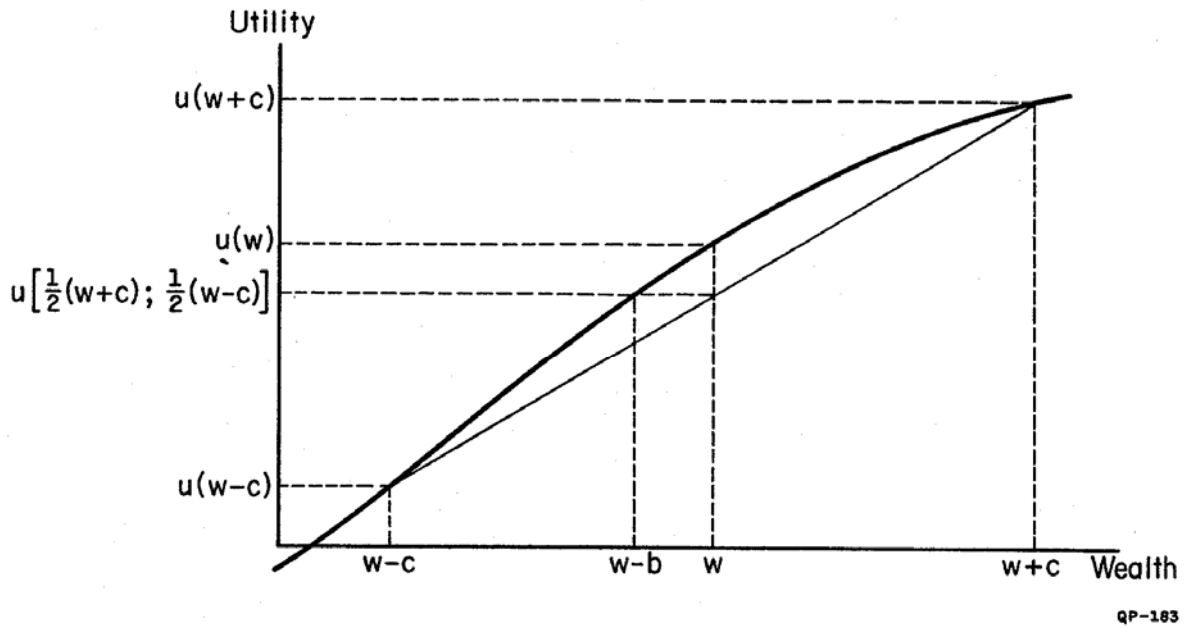


Figure 3

while the monotonicity of the u_i insures that there is a point x in S^* such that $x > d^*$.

Comparative Risk Aversion

We would like to have some way of comparing the risk aversion expressed by different utility functions. As we have seen, the risk aversion of a utility function for money is related to its concavity, and so we might hope, for instance, that the second derivative at a given wealth w might serve as a measure of the risk aversion at that wealth. However, if $v_i = au_i$ for some positive constant $a \neq 1$, then v_i and u_i are equivalent utility functions, but $u_i'' \neq v_i''$. So the second derivative is an inappropriate measure of individual i 's risk aversion, since it isn't constant for equivalent utility representations. We need a way to compare the risk aversion of two utility functions in a manner which is independent of the arbitrary features of those functions. The following definition, which depends only on the preferences which the utility functions represent, is due to Yaari [1969], who generalized some of the pioneering work of Arrow [1965], [1971] and Pratt [1964] in this area.

Let C be a convex subset of R^n representing a set of riskless outcomes (i.e., outcomes which do not involve lotteries), and let u and v be risk averse (i.e., concave) von Neumann-Morgenstern utility functions defined on C and on the mixture set M of lotteries generated⁴ by C . Note that C is a subset of M , since the lottery $[1c; 0b]$ is equal to c and is an element of M . For any outcome c in C , the acceptance set of an individual with utility function u ⁵ is defined to be the set

$$A_u(c) \equiv \{m \in M \mid u(m) \geq u(c)\}.$$

Thus $A_u(c)$ is the set of all lotteries which individual u would accept as alternatives to c ; i.e., the set of all alternatives which he likes at least as well as c .

⁴The mixture set generated by C is the smallest mixture set which contains C . It can be thought of as the set of lotteries whose outcomes are all in C . (For the definition of a mixture set, see p. 2).

⁵An individual with utility function u will sometimes be referred to simply as "individual u ."

A utility function v is defined to be (weakly) more risk averse than u if, for every outcome c in C , $A_u(c)$ contains $A_v(c)$. That is, v is more risk averse than u if every lottery which would be accepted by individual u in place of the riskless outcome c would also be accepted by individual v . If $A_u(c)$ strictly contains $A_v(c)$ for all c in C , then v is strictly more risk averse than u . Thus a strictly more risk averse individual would decline to accept some lotteries which a less risk averse individual would accept in place of any riskless outcome c .

If v is more risk averse than u , it follows that u and v must order the (riskless) elements of C in almost the same way: i.e., for all elements b and c of C , $v(b) \geq v(c)$ implies $u(b) \geq u(c)$, and $u(b) > u(c)$ implies $v(b) > v(c)$. This is because if $v(b) \geq v(c)$, then b is an element of $A_v(c)$ and consequently it is also an element of $A_u(c)$, which is to say $u(b) \geq u(c)$. If $u(b) > u(c)$, then c is not an element of the acceptance set $A_u(b)$, and consequently it is not an element of $A_v(b)$ either, which is to say $v(b) > v(c)$.⁶ Any two utility functions which coincide on the set of riskless outcomes, and whose risk aversion can be compared in this way, differ only on the set of risky outcomes. So the difference between two such functions lies entirely in the attitude which they express towards risk.

Having thus characterized the relative risk aversion of utility functions in terms of their acceptance sets, we can prove the following theorem which permits the same relationship to be expressed directly in terms of the utility functions (Kihlstrom and Mirman [1974], Proposition 6). We will henceforth assume that v and u induce the same ordering on C whenever v is more risk averse than u .

Theorem 4: A utility function v is more risk averse than u if and only if, for all riskless outcomes c in C , $v(c) = k(u(c))$, where k is an increasing concave function.

Furthermore, v is strictly more risk averse than u if and only if the function k is strictly concave.

⁶ Note that u and v need not order the elements of C in precisely the same way, since we have not ruled out the possibility that there exist elements b and c such that $u(b) = u(c)$ but $v(c) > v(b)$. Indeed, let $u(c) = k$ for all c in C , i.e., let u be constant on C . Then any (non-constant) utility function v is more risk averse than u , but u and v do not induce the same ordering on C . (See Roth [1979e]).

Proof: First suppose that v is more risk averse than u . Since v and u induce the same ordering on C , it follows that for all c in C , $v(c) = k(u(c))$ where k is an increasing function (i.e., $k(x) \geq k(y)$ if and only if $x \geq y$). We need to show that k is concave.

Suppose that k is not concave over the range of values taken by u on the set C . Then there exist outcomes b and c in C and some real number q in the interval $[0,1]$ such that $qk(u(b)) + (1-q)k(u(c)) > k(qu(b) + (1-q)u(c))$. Let \hat{a} in C be a riskless outcome such that $v(\hat{a}) = v([q;1-q]c) = qv(b) + (1-q)v(c)$. That is, \hat{a} is a riskless outcome which is equivalent to the lottery $[q;1-q]c$ as evaluated by v . (Such an outcome \hat{a} , which is sometimes called the certain equivalent of the lottery, exists because v is concave on C , which is a convex set). Then the lottery $[q;1-q]c$ is contained in the acceptance set $A_v(\hat{a})$.

$$\begin{aligned} \text{However } k(u(\hat{a})) &= v(\hat{a}) = qv(b) + (1-q)v(c) \\ &= qk(u(b)) + (1-q)k(u(c)) \\ &> k(qu(b) + (1-q)u(c)). \end{aligned}$$

Since k is an increasing function, this implies that $u(\hat{a}) > qu(b) + (1-q)u(c) = u([q;1-q]c)$. Thus the lottery $[q;1-q]c$ is not contained in the acceptance set $A_u(\hat{a})$, which contradicts the assumption that v is more risk averse than u . Thus, if v is more risk averse than u , k must be concave.

To prove the theorem in the other direction, suppose that for all outcomes c in C , $v(c) = k(u(c))$, where k is an increasing concave function. For outcomes a, b, c in C , let $m = [q;1-q]b$ be a lottery contained in the acceptance set $A_v(c)$. Then $k(qu(a) + (1-q)u(b)) \geq qk(u(a)) + (1-q)k(u(b))$ by the concavity of k , and

$$qk(u(a)) + (1-q)k(u(b)) = qv(a) + (1-q)v(b) = v(m) \geq v(c) = k(u(c))$$

since m is in the acceptance set $A_v(c)$. Therefore $k(qu(a) + (1-q)u(b)) \geq k(u(c))$. Since k is an increasing function, it follows that $u(m) = qu(a) + (1-q)u(b) \geq u(c)$, and so m is also contained in the acceptance set $A_u(c)$. Thus $A_v(c)$ is a subset of $A_u(c)$, and so v is more risk averse than u .

In what follows, we will consider two-player bargaining games (S, d) as arising from bargaining over a set of riskless alternatives C , by players with utility functions u_1 and u_2 . That is, the feasible set of utility payoffs can be represented as the set $S = \{(x_1, x_2) \mid x_1 = u_1(c) \text{ and } x_2 = u_2(c) \text{ for some } c \text{ in } C\}$ and the disagreement point d can be represented as the point $d = (u_1(\hat{c}), u_2(\hat{c}))$ where \hat{c} is the disagreement outcome. Every game (S, d) in the class B can be represented in this way.⁷ As in Section C, the Pareto optimal subset of S , $P(S)$, can be represented as the set of points of the form $(x_1, \phi(x_1))$, where $x_1 = u_1(c)$ and $\phi(x_1) = u_2(c)$ for some Pareto optimal outcome c in C , where ϕ is a monotone decreasing concave function defined on an interval $[\underline{x}_1, \bar{x}_1]$.

To see how risk aversion enters into Nash's model of the bargaining problem, let w be the utility function of player 2 in the game (S, d) just considered; i.e., let $u_2 = w$. Now consider another game (\hat{S}, \hat{d}) , derived from the first game by replacing player 2 with a more risk averse player, whose utility function is \hat{w} . Then for all $c \in C$, $\hat{w}(c) = k(w(c))$, where k is an increasing concave function. So the new disagreement point \hat{d} is given by $\hat{d} = (u_1(\hat{c}), \hat{w}(\hat{c})) = (d_1, k(d_2))$, while the new feasible set \hat{S} is given by

$$\begin{aligned} \hat{S} &= \{(y_1, y_2) \mid y_1 = u_1(c), y_2 = \hat{w}(c) \text{ for some } c \in C\} \\ &= \{(y_1, y_2) \mid y_1 = x_1, y_2 = k(x_2) \text{ for some } (x_1, x_2) \in S\}. \end{aligned}$$

The set of Pareto optimal points in \hat{S} is the set $P(\hat{S}) = \{(x_1, k(\phi(x_1))) \mid (x_1, \phi(x_1)) \in P(S)\}$, since both $P(\hat{S})$ and $P(S)$ are determined by the same underlying outcomes c in C .⁸

For example, consider the game (S^*, d^*) , defined earlier, which results from bargaining over the division of Q dollars, between players with utility functions u_1 and u_2 , and initial wealth w_1 and w_2 . The Pareto optimal set of S^* is the set $P(S^*) = \{(x_1, x_2) \mid x_1 = u_1(w_1 + c_1), x_2 = u_2(w_2 + Q - c_1) \text{ for } 0 \leq c_1 \leq Q\}$. The set $P(S^*)$

⁷ Of course bargaining games in the class B can arise in other ways as well, e.g., from bargaining over risky as well as riskless alternatives. See, for instance, Kihlstrom and Roth [1979] for a discussion of bargaining over insurance contracts.

⁸ This would not necessarily be the case if the bargaining were over lotteries instead of riskless outcomes, since a given lottery might be Pareto optimal for one pair of players, but no longer be Pareto optimal for more risk averse players.

can be described as the set of points $(x_1, \phi(x_1))$, where ϕ is a monotone decreasing, concave function defined on the interval $u_1(w_1) = d_1^* \leq x_1 \leq u_1(w_1 + Q)$.

Let v_2 be another utility function, such that v_2 is more risk averse than u_2 , and let $(T^\#, d^\#)$ be the game which results when u_2 is replaced by v_2 in the game (S^*, d^*) . That is,

$$\begin{aligned} T^\# &= \{(y_1, y_2) \mid y_1 = u_1(w_1 + c_1), y_2 = v_2(w_2 + c_2) \text{ for some } c_1, c_2 \geq 0 \\ &\quad \text{such that } c_1 + c_2 \leq Q\} \\ &= \{(y_1, y_2) \mid y_1 = x_1 \text{ and } y_2 = k(x_2) \text{ for some } (x_1, x_2) \in S^*\} \\ &\quad \text{where } k \text{ is an increasing, concave function.} \end{aligned}$$

So the Pareto optimal set of outcomes in $T^\#$ is

$$P(T^\#) = \{(y_1, y_2) \mid y_1 = u_1(w_1 + c_1), y_2 = k(u_2(w_2 + Q - c_1)) \text{ for } 0 \leq c_1 \leq Q\},$$

so $P(T^\#)$ is the set of points of the form $(x_1, k(\phi(x_1)))$. (See Figure 4, p. 43). Making player 2 more risk averse has the effect of making the Pareto optimal set of utility payoffs more concave as a function of player 1's utility.

Note that we have drawn Figure 4 so that $d^* = d^\#$, and $k(\phi(x_1)) = \phi(x_1)$ for some point $x_1 = u_1(w_1 + c_1)$. This simply amounts to choosing a normalization of v_2 such that $v_2(w_2) = u_2(w_2)$, and $v_2(w_2 + Q - c_1) = u_2(w_2 + Q - c_1)$.

To see how the risk aversion of the players affects the Nash solution, consider a bargaining game (S, d) whose Pareto set is the set of points of the form $(x_1, \phi(x_1))$, and let $z = F(S, d)$. Then z_1 maximizes the geometric average $A(x_1) = (\phi(x_1) - d_2)(x_1 - d_1)$ on S , and so we have the following lemma. (For simplicity, we state the lemma for the case that ϕ is differentiable.)

Lemma 5.1: $(\phi(z_1) - d_2)/(z_1 - d_1) = -\phi'(z_1)$.

Proof: Since $A(z_1)$ is maximal, we know that $A'(z_1) = \phi'(z_1)(z_1 - d_1) + \phi(z_1) - d_2 = 0$. The lemma follows immediately.

The lemma says that the slope of the line joining the disagreement point to the point $z = F(S, d)$ is the negative of the slope of the line tangent to S at z . That is, in Figure 5 (p. 43), the angles α and β are equal.

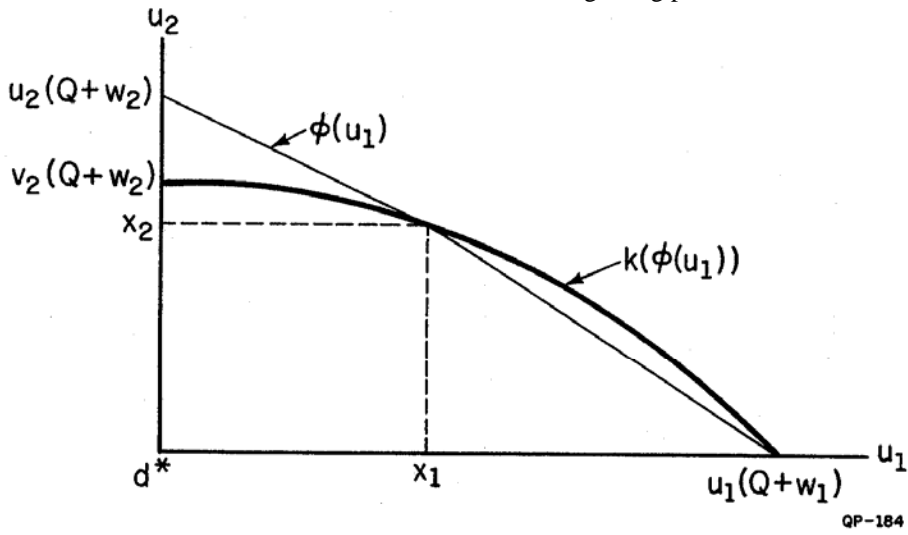


Figure 4

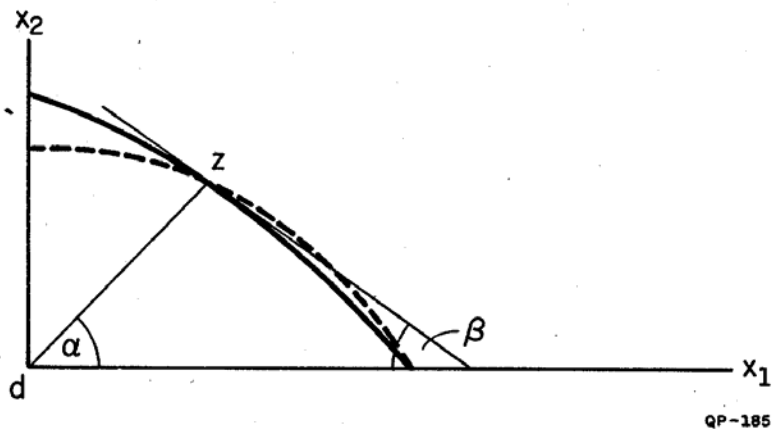


Figure 5

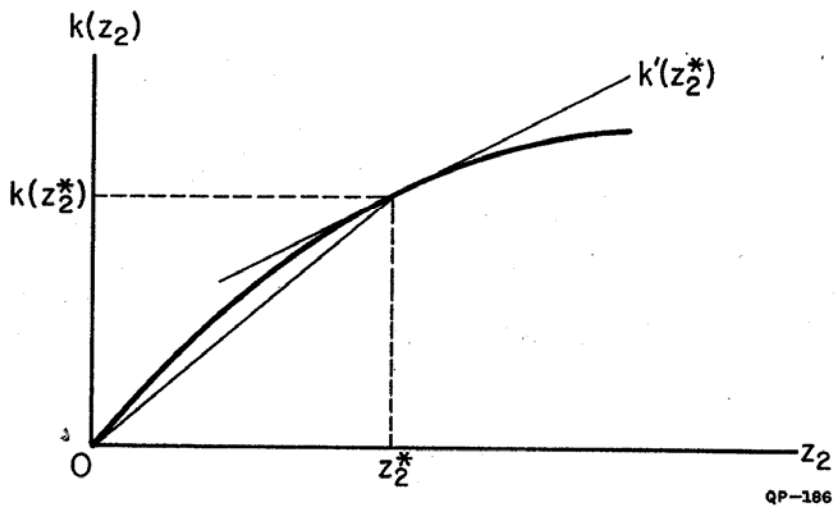


Figure 6

If ϕ is not differentiable, then several tangents to S may exist at any point $(z_1, \phi(z_1))$, but Nash's solution can still be characterized as the solution which selects the unique point z such that the line joining d to z has the negative slope of some tangent to S at z .

So we can see from the dotted line in Figure 5 that as player 2, say, becomes more risk averse, the Pareto optimal set of payoffs in S becomes more concave and "flatter" as a function of player 1's utility, so that the tangent at z^9 to the new set of feasible utility payoffs will make a more acute angle β than previously. Consequently, in the new game the Nash solution will occur at a point which makes the angle α more acute as well, which is to say player 1's utility will be greater than z_1 at the Nash solution to the new game. More formally, we can state the following result for two-player bargaining games. (Our presentation follows that of Kihlstrom, Roth, and Schmeidler [1979].)

Theorem 5: The utility which Nash's solution assigns to a player in a two-person game increases as his opponent becomes more risk averse. That is, $F_1(\hat{S}, \hat{d}) \geq F_1(S, d)$ where (\hat{S}, \hat{d}) is obtained from (S, d) by replacing player j , ($j \neq 1$), with a more risk averse player.

Proof: The game (\hat{S}, \hat{d}) is derived from (S, d) by replacing player 2, say, with a more risk averse player, so the Pareto optimal points in S are of the form $(x_1, \phi(x_1))$, while the Pareto optimal points in \hat{S} are of the form $(x_1, \hat{\phi}(x_1))$, where $\hat{\phi} = k(\phi)$, and k is increasing and concave. It will be sufficient for the proof of the theorem to consider only games (S, d) and (\hat{S}, \hat{d}) such that ϕ and $\hat{\phi}$ are differentiable. This is because Nash's solution F is continuous (in the Hausdorff topology) on the class B of bargaining games. Thus, for instance, if (S, d) is a game in B whose Pareto set is not differentiable, there exists a sequence of games (S^n, d) in B which converges to (S, d) , and whose Pareto sets are differentiable. The continuity of F insures that $F(S^n, d)$ converges to $F(S, d)$, and so if the theorem is true for differentiable games then it is true for all games.

⁹We can always choose player 2's utility functions so that the curves $\phi(x_1)$ and $k(\phi(x_1))$ which define the Pareto surface intersect at z . (See the comments on Figure 4 on p. 42).

Since the solution F is independent of equivalent utility representations, it will also be sufficient to prove the theorem for games normalized so that d and \hat{d} are equal to the origin. So let $(S, \bar{0})$ and $(\hat{S}, \bar{0})$ be bargaining games which differ only in the risk aversion of, say, player 2. Then the Pareto optimal points of S are the points $(x_1, \phi(x_1))$ for $0 \leq x_1 \leq \bar{x}_1$, and the Pareto optimal points of \hat{S} are the points $(x_1, k(\phi(x_1)))$ defined on the same interval, where k is an increasing, concave function. Let $z^* = F(S, \bar{0})$ and $z^\# = F(\hat{S}, \bar{0})$; we want to show that $z_1^\# > z_1^*$. It is sufficient to show that the geometric average $A(x_1) = k(\phi(x_1))x_1$ has a positive first derivative at z_1^* .

But

$$A'(z_1^*) = k'(\phi(z_1^*))\phi'(z_1^*)z_1^* + k(\phi(z_1^*)) .$$

By Lemma 5.1, $\phi'(z_1^*)z_1^* = -\phi(z_1^*)$, so

$$\begin{aligned} A'(z_1^*) &= -k'(\phi(z_1^*))\phi(z_1^*) + k(\phi(z_1^*)) \\ &= -k'(z_2^*)z_2^* + k(z_2^*) \\ &= z_2^* \left[-k'(z_2^*) + \frac{k(z_2^*)}{z_2^*} \right] . \end{aligned}$$

The concavity of the function k implies that $k(z_2^*)/z_2^* \geq k'(z_2^*)$ (see Figure 6, p. 43), while the individual rationality of the Nash solution implies $z_2^* > 0$, so $A'(z_1^*) \geq 0$, as required. (In particular, if player 2 is strictly more risk averse in $(\hat{S}, \bar{0})$ than in $(S, \bar{0})$, then k is strictly concave and $A'(z_1^*) > 0$.) This completes the proof.

(Note that although we proved the theorem in terms of the effect on player 1's payoff when player 2 is replaced by a more risk averse player, the symmetry of the Nash solution insures that the same conclusion holds for player 2.)

This theorem can be used to determine the relative monetary payoffs predicted by the Nash solution in games of the form (S^*, d^*) , in which players with utility functions u_1 and u_2 and initial wealths w_1 and w_2 bargain over the division of Q dollars. Let $v_1(c) = u_1(w_1 + c)$ and $v_2(c) = u_2(w_2 + c)$ for all values of c between 0 and Q . Then $v_1(c)$ is player 1's utility when he receives c dollars, and the game (S^*, d^*) can be represented by $d^* = (v_1(0), v_2(0))$, and

$$S^* = \{(x_1, x_2) \mid x_1 = v_1(c_1) \text{ and } x_2 = v_2(c_2) \text{ for } c_1 + c_2 \leq Q, \text{ and } c_1, c_2 \geq 0\}.$$

Let $z^* = F(S^*, d^*)$, and let c_1^* and c_2^* be the monetary payoffs which yield the utilities z_1 and z_2 (i.e., $z_1 = u_1(w_1 + c_1^*) = v_1(c_1^*)$). Then we can state the following corollary of Theorem 5.

Corollary 5.1: If v_i is strictly more risk averse than v_j on the interval $0 \leq c \leq Q$, then $c_i^* < c_j^*$.

That is, the Nash solution predicts that the player who is more risk averse on the range of feasible monetary payoffs will receive the smaller payoff.

Proof: Suppose $v_i(c) = k(v_j(c))$ for $0 \leq c \leq Q$, where k is an increasing concave function. If k were linear, then v_i and v_j would be equivalent utility functions, and the symmetry of Nash's solution would imply $c_i^* = c_j^* = Q/2$. However k is strictly concave, so v_j is strictly more risk averse than v_i , and the corollary now follows immediately from Theorem 5.

The Harsanyi-Zeuthen model of negotiation considered in the previous section can be used to provide an alternative, indirect proof of Theorem 5, by means of the following observation.

Proposition 5.1: If v_i and u_i are two utility functions such that v_i is more risk averse than u_i , then the risk limit \bar{r}_i for any pair of proposals is lower under v_i than under u_i .

Proof: Consider an individual whose utility function is u_i , evaluating his risk limit for proposals which give him a utility of y_i if his demand is met, x_i if he concedes to his opponent's demand, and d_i if conflict results, with $y_i > x_i > d_i$. His risk limit \bar{r}_i as defined by Harsanyi and Zeuthen is then $\bar{r}_i = (y_i - x_i)/(y_i - d_i)$. Now $v_i = k(u_i)$ where k is a concave function, and we want to show that the risk limit evaluated by the utility function v_i is always less than \bar{r}_i , i.e., we want to show that $(k(y_i) - k(x_i))/(k(y_i) - k(d_i)) \leq \bar{r}_i$. But observe that $x_i = \alpha d_i + (1 - \alpha)y_i$ where $\alpha = \bar{r}_i$, and

$k(x_1) = k(\alpha d_1 + (1 - \alpha)y_1) \geq \alpha k(d_1) + (1 - \alpha)k(y_1)$, since k is concave. Consequently

$$\frac{k(y_1) - k(x_1)}{k(y_1) - k(d_1)} \leq \frac{k(y_1) - (\alpha k(d_1) + (1 - \alpha)k(y_1))}{k(y_1) - k(d_1)} = \alpha = \bar{r}_1,$$

which completes the proof of the proposition.

Together with the results of the Harsanyi-Zeuthen negotiation model, this proposition provides an alternate proof of Theorem 5, since the lower an individual's risk limit, the more he must concede to his opponent at any stage of the negotiations.

So the Nash solution F is sensitive to risk aversion in any two-person bargaining game, which we can formalize as a property which we might find desirable in any solution f . In particular, let (S', d') be a two-person bargaining game derived from (S, d) by replacing one of the players, say player i , with a more risk averse player. That is, (S', d') is the game in which $d'_1 = k(d_1)$, $d'_j = d_j$, and $S' = \{(y_1, y_2) \mid y_1 = k(x_1) \text{ and } y_j = x_j \text{ for some } (x_1, x_2) \text{ in } S\}$, where k is an increasing concave function. (So if $\phi(x_j)$ describes player i 's Pareto optimal utility payoffs in S as a function of player j 's utility, then in S' these payoffs are described by $k(\phi(x_j))$.) Then the Nash solution possesses the following property.

Property 8: Risk sensitivity. If a two-person bargaining game (S, d) is transformed into a game (S', d') by replacing player i with a more risk averse player, then $f_j(S', d') \geq f_j(S, d)$.

Note that Property 8 makes a meaningful comparison, because player j 's utility function is unchanged in going from the game (S, d) to (S', d') .

Somewhat surprisingly, a solution which is both risk sensitive and Pareto optimal must also be independent of equivalent utility representations. That is, we have the following result.

Theorem 6: If f is a solution for two-player games which possesses Properties 4 and 8, then it also possesses Property 1.

Proof: The proof is simple, and consists of essentially two observations. The first is that the game (S, d) is derived from (S', d') by subjecting player i 's utility function in the latter game to a convex transformation, k^{-1} . So convex transformations¹⁰ of i 's utility lower f_j , just as concave transformations raise it, and so linear transformations of one player's utility leave the other player's payoff unchanged (since a linear transformation is both concave and convex). This, of course, is half of Property 1: the other half is that a linear transformation of a player's utility function should change his own payoff by the same linear transformation. But this follows from the Pareto optimality of f , which permits us to compute f_i as a function of f_j : since f_j is unchanged by a linear transformation of player i 's utility, the solution f picks the transformed point, as required.

An immediate corollary of Theorem 6 is, of course, that Property 8 can be used to characterize Nash's solution, as follows.

Corollary 6.1: The Nash solution is the unique solution for two-player games which possesses Properties 2, 3, 4, and 8.

Theorem 6 is somewhat counterintuitive, since it deduces linearity (Property 1) in part from the risk sensitivity property, which is specifically concerned with the non-linearity of risk averse utility functions. I think the best explanation is probably that the intuitive plausibility of the risk sensitivity property derives in part from the feeling that the outcome of bargaining may turn out not to be Pareto optimal. In particular, a disagreement may occur, and the fear of this eventuality may cause a highly risk-averse player to settle for an unfavorable agreement. Note that for games with more than two players it is not true in general that, as one player's utility function becomes more concave, the other players all receive a higher payoff under the Nash solution (e.g., consider the game described on p. 19).

¹⁰Of course not every convex transformation will leave the feasible set convex, i.e., not every convex transformation of a player's utility transforms a bargaining game into another bargaining game.

Boldness and Fear of Ruin

In considering the risk sensitivity of the Nash solution, we have so far been concerned only with fairly coarse comparisons of risk aversion. In the statement of Property 8, the game (S', d') is derived from the game (S, d) through a transformation of player 1's utility which affects every point in the set S . Since the Nash solution is independent of irrelevant alternatives, it should be clear that it will also be sensitive to more localized changes in player 1's utility function, affecting only the region near the payoff $z_1 = F_1(S, d)$. Lemma 5.1 makes it particularly clear that the Nash solution is determined by the utility functions of the players only in the neighborhood of the outcome z , and the disagreement point d .

A notion of risk aversion which captures the essential information needed to determine the Nash solution is the concept of "boldness" (or its inverse "fear of ruin") introduced by Aumann and Kurz [1977a], [1977b]. Consider the game (S^*, d^*) in which two players with initial wealth w_1 and w_2 bargain over how to split Q dollars. A feasible proposal is a proposed split (c_1, c_2) , such that $c_1 + c_2 \leq Q$ and $c_1, c_2 \geq 0$, which leaves each player i with total wealth $w_i + c_i$. Player i 's boldness with respect to the proposal (c_1, c_2) is defined to be the quantity $b_i(w_i, c_i) = u_i'(w_i + c_i) / (u_i(w_i + c_i) - u_i(w_i))$, and the inverse of this quantity is called his fear of ruin.¹¹

To see why this is a reasonable definition, suppose that player i has been offered the amount c_i , so that, if he accepts, his total wealth will be $w_i + c_i$. Suppose that he now considers a gamble which risks his entire gain c_i against the possibility of a small additional gain h . If h is very small in comparison with c_i , then the probability q that he will be "ruined" and lose the entire amount c_i would have to be quite small in order for him to prefer to accept the gamble, rather than to take c_i for certain. The maximum probability of ruin \bar{q} for which player i is willing to accept the gamble can be considered a measure of his boldness (or an inverse measure of his fear of ruin). That is, the bolder a player is (i.e., the

¹¹ Aumann and Kurz [1977a], [1977b] phrase their discussion primarily in terms of the fear of ruin, but in the present context we will find it more convenient to concentrate on boldness. In any event, our discussion closely follows Aumann and Kurz [1977a].

less fearful of ruin), the higher the probability of ruin must be to dissuade him from risking c_1 for a small additional gain h . Of course, \bar{q} tends to 0 as the additional gains h get smaller. It turns out that the boldness $b_1(w_1, c_1)$ is equal to the maximum probability of ruin per dollar of additional gains which player 1 is prepared to tolerate, for very small potential gains. That is,

$$b_1(w_1, c_1) = \lim_{h \rightarrow 0} \bar{q}/h.$$

To see this, observe that \bar{q} is the probability such that $u_1(w_1 + c_1) = \bar{q}u_1(w_1) + (1 - \bar{q})u_1(w_1 + c_1 + h)$. Consequently

$$\frac{\bar{q}}{h} = \frac{(u_1(w_1 + c_1 + h) - u_1(w_1 + c_1))/h}{u_1(w_1 + c_1 + h) - u_1(w_1)},$$

and so

$$\lim_{h \rightarrow 0} \frac{\bar{q}}{h} = \frac{u_1'(w_1 + c_1)}{u_1'(w_1 + c_1) - u_1'(w_1)} = b_1(w_1, c_1).$$

The Nash solution can now be characterized as selecting the point z^* in S^* at which the players are equally bold (or equally fearful of ruin). That is, we have the following result.

Theorem 7: $F(S^*, d^*) = z^* = (u_1(w_1 + c_1^*), u_2(w_2 + c_2^*))$ such that $c_1^* + c_2^* = Q$ and $b_1(w_1, c_1^*) = b_2(w_2, c_2^*)$.

Proof: The Nash solution picks the point which maximizes the product $A(c_1) = [u_1(w_1 + c_1) - u_1(w_1)][u_2(w_2 + Q - c_1) - u_2(w_2)]$. Setting dA/dc_1 equal to zero yields $-u_2'(w_2 + Q - c_1^*)[u_1(w_1 + c_1^*) - u_1(w_1)] + u_1'(w_1 + c_1^*)[u_2(w_2 + Q - c_1^*) - u_2(w_2)] = 0$, which gives us the required result.

Note that the assumption that utility functions for money are increasing and concave implies that a player's boldness is a decreasing function of his gains. That is

$$\frac{\partial b_1(w_1, c_1)}{\partial c_1} = \frac{[u_1(w_1 + c_1) - u_1(w_1)]u_1''(w_1 + c_1) - (u_1'(w_1 + c_1))^2}{[u_1(w_1 + c_1) - u_1(w_1)]^2} < 0,$$

since the denominator is always positive, and the numerator is always negative.

Consequently if, say, player 2's boldness were to decrease with respect to the gains c_2^* , then player 1's gains at the Nash solution of the resulting bargaining game would be greater than c_1^* . Another way of stating the relationship between boldness and the Nash solution in the game (S^*, d^*) is as follows.

Corollary 7.1: $c_1^* \geq c_2^*$ if and only if $b_1(w_1, \frac{Q}{2}) \geq b_2(w_2, \frac{Q}{2})$.

That is, the player who is bolder with respect to an equal division of the available money obtains the larger share according to the Nash solution. The proof is immediate from Theorem 7 and the fact that $b_1(w_1, c_1)$ is a decreasing function of c_1 . Note that Corollary 7.1 presents a more sensitive result than Corollary 5.1.

An example: Consider a game (S^*, d^*) in which two players bargain to divide Q dollars, and suppose that the utility functions of the two players for money are $u_1(c) = c$ and $u_2(c) = \sqrt{c}$. If the initial wealth of each player is 0, then $d^* = (0, 0)$, and $S^* = \{(x_1, x_2) \mid x_1 = c_1, x_2 = \sqrt{c_2}, \text{ where } c_1 + c_2 \leq Q, c_1, c_2 \geq 0\}$. Let c_1^* and c_2^* be the monetary payoffs which correspond to Nash's solution. Note that player 2 is strictly more risk averse than player 1, since u_2 is a concave function of u_1 , and so Corollary 5.1 (p. 46) implies that $c_2^* < c_1^*$. Furthermore, the boldness of the two players with respect to any payoffs c_1 and c_2 is $b_1(0, c_1) = \frac{1}{c_1}$ and $b_2(0, c_2) = 1/2c_2$. So the boldness of the two players with respect to an equal division of the money is $b_1(0, \frac{Q}{2}) = \frac{2}{Q}$ and $b_2(0, \frac{Q}{2}) = \frac{1}{Q}$. So player 1 is bolder with respect to an equal division, and so Corollary 7.1 also implies that $c_2^* < c_1^*$. Setting $b_1(0, c_1) = b_2(0, c_2)$ we find that in fact $c_1^* = \frac{2}{3}Q$ and $c_2^* = \frac{1}{3}Q$.

Naturally, Theorem 7 is equivalent to Lemma 5.1 for games of the form (S^*, d^*) , since they both characterize the Nash solution on these games. In fact, the two results describe essentially the same phenomenon. To see this, let $F(S^*, d^*) = z^* = (u_1(w_1 + c_1^*), u_2(w_2 + c_2^*))$, and let the Pareto surface of S^* be the set of points $(x_1, \phi(x_1))$ where $x_1 = u_1(w_1 + c_1)$ for $0 \leq c_1 \leq Q$, and ϕ is defined by $u_2(w_2 + c_2) = \phi(u_1(w_1 + Q - c_2))$. Then $u_2'(w_2 + c_2) = -\phi'(u_1(w_1 + Q - c_2))u_1'(w_1 + Q - c_2)$. For simplicity, we can consider the utility functions to be normalized so that $d^* = (u_1(w_1), u_2(w_2)) = (0, 0)$.

Then Theorem 7 asserts that

$$\frac{u_1'(w_1 + c_1^*)}{z_1^*} = \frac{u_2'(w_2 + c_2^*)}{z_2^*} = \frac{-\phi'(u_1(w_1 + Q - c_2^*))u_1'(w_1 + Q - c_2^*)}{z_2^*}$$

but $Q - c_2^* = c_1^*$, so this is equivalent to $z_2^*/z_1^* = -\phi'(z_1^*)$, which is precisely Lemma 5.1.

Strategic Risk Posture and the Utility of Bargaining

So far our study of bargaining has been divided conceptually into two distinct parts. First, the preferences of the players over a set of underlying alternatives ("prizes")¹² are codified in terms of expected utility functions. This then leads to the formal description of the bargaining game in terms of the available utility payoffs, which we have studied axiomatically by specifying the properties of a solution. Thus all of the information about a player's attitude towards risk is contained in his utility function (for the underlying prizes), while all the information about the bargaining process is contained in the solution.

This division is in some respects an artificial one. If a player believes that a given solution is an accurate model of the bargaining process, then when he considers two different bargaining situations, he prefers the one in which the solution predicts he will get a higher utility payoff. That is, a given solution induces a preference relation over bargaining situations. On the other hand, if an individual has sufficiently well behaved preferences over bargaining situations (cf. Section A) then his preferences can be modelled as a utility function for bargaining. This is the approach which we will now proceed to investigate.¹³

In particular, we will investigate what kind of preferences over bargaining situations lead to utility functions corresponding to the symmetric and non-symmetric Nash solutions. Just as formulating an individual's utility function for money forces us to explicitly take into account his attitudes towards the (probabilistic) risk associated with lotteries, we will find that, in order to

¹²Since we will be speaking here about alternative bargaining situations, we will designate the alternative bargains available in a given game as prizes.

¹³Our discussion closely follows that of Roth [1978].

formulate his utility function for bargaining games we must explicitly take into account his attitude towards the strategic risk associated with bargaining.

For simplicity, we will confine our attention to bargaining games with a common disagreement point $d = \bar{0}$, and denote a bargaining game $(S, \bar{0})$ by its set of outcomes S . Let B' be the set of all such (non-degenerate) bargaining games, together with the degenerate game $A_{\bar{0}} = \{\bar{0}\}$, in which the disagreement point is the only feasible outcome. For any positive vector x in R^n , it will also be convenient to denote by A_x the game $A_x = \{ax \mid 0 \leq a \leq 1\}$. That is, A_x is the game whose outcome set is the line segment joining the disagreement point $\bar{0}$ with the point x .

We will be interested in an individual's preference relation R over the (smallest) mixture set M' which contains $N \times B'$ (cf. p. 2). That is, we are interested in an individual's preferences over alternatives of the form (i, S) , where $i \in N = \{1, \dots, n\}$ and $S \in B'$, as well as over all lotteries which can be generated from alternatives of this form. For instance the statement $(i, S)P(j, T)$ means "it is preferable to play position i in the game S than to play position j in the game T ." (The strict preference relation P , and indifference relation I are defined as in Section A.)

We call the set $N = \{1, \dots, n\}$ the set of positions (rather than the set of players) to emphasize that we are now considering the preferences of a single individual over alternative bargaining situations. If, for instance, this individual chooses the alternative (i, S) it means that he will participate in the bargaining game S in position i , while the other positions will be filled by $n - 1$ other players (who we will regard as being drawn at random from some fixed set of potential players).

Of course the set of outcomes S in a bargaining game is defined in terms of the utility functions of each of the players for the underlying prizes (e.g., their utility functions for money). Thus if x is an outcome in the set S , then the alternative (i, S) represents a bargaining situation in which the prize corresponding to the payoff vector x gives the individual whose preferences we are studying the utility payoff \hat{x}_i (in terms of his own utility function over prizes) while

giving the players in other positions j the utility x_j (in terms of their own utility functions).¹⁴

We will assume that the preference relation R obeys the necessary conditions of substitutability and continuity (as on p. 3) to insure that an expected utility function θ on M' exists. That is, θ is a real-valued function on M' such that $\theta(i,S) > \theta(j,T)$ if and only if $(i,S)P(j,T)$, and $\theta[p(i,S); (1-p)(j,T)] = p\theta(i,S) + (1-p)\theta(j,T)$. We will sometimes write $\theta_i(S) = \theta(i,S)$, and $\theta(S) = (\theta_1(S), \dots, \theta_n(S))$. That is, $\theta(S)$ is a vector whose i -th component is the utility of playing position i in game S . To insure that the names of the positions have no influence on the desirability of a particular position in a game, we require that the preference relation obey the following conditions.

Condition 1:¹⁵ For all $i \in N$, $S \in B'$ and every permutation π of N , $(i,A)I(\pi i, \pi A)$.

Since we are considering preferences over bargaining situations, which are themselves defined in terms of the utility of prizes, we will also require that the preferences over bargaining games be consistent with the preferences for prizes which the utility payoffs of a given bargaining game convey. The following condition insures that the preferences over the "sure-thing" games A_x , which involve no conflict among the players, are compatible with the interpretation of x as a utility vector.

Condition 2: $(i,A_x)R(i,A_y)$ if and only if $x_i \geq y_i$.

¹⁴For instance, for the sake of concreteness we might imagine that before a game is played, each player's utility function for prizes (with an arbitrary origin and scale) is calibrated in terms of "play money". The set of outcomes S in a bargaining game then corresponds to a set of payoff vectors in play money. After the bargaining is concluded, each player exchanges his play money for prizes at the exchange rate established for his own utility function.

¹⁵This is essentially a symmetry condition, but we shall see that it plays a somewhat different role than the symmetry property of the Nash solution.

So $(i, A_x)P(i, A_y)$ if and only if $x_i > y_i$; i.e., it is preferable to play position i in the game A_x than in A_y if and only if the resulting utility payoff is higher.¹⁶

An immediate consequence of Conditions 1 and 2 in terms of the utility function θ is

Lemma 8.1: $\theta_i(A_x) \geq \theta_j(A_y)$ iff $x_i \geq y_j$.

In particular if $x_i = y_j > 0$ then $\theta_i(A_x) = \theta_j(A_y) > \theta_i(A_0)$, so we can choose the natural normalization for the utility function θ , and set $\theta(i, A_0) = 0$, and $\theta(i, A_{\bar{1}}) = 1$ where $\bar{1} = (1, \dots, 1)$.¹⁷

The conditions which insure that the utility function θ will share the Nash solution's two independence properties (Properties 1 and 3) are the following:

Condition 3: If $S' = \{(a_1 x_1, \dots, a_n x_n) | x \in S\}$ for $a_j > 0$, $j = 1, \dots, n$, and if $a_i \geq 1$, then $(i, S)I[(1/a_i)(i, S'); (1-1/a_i)(i, A_0)]$.

Condition 4: If A_x is a subset of S , and S is a subset of T , where x is Pareto optimal in T , and if $(i, A_x)I(i, T)$, then $(i, A_x)I(i, S)$.

Condition 4 might be called "indifference to irrelevant alternatives" while Condition 3 essentially imposes the requirement that θ be independent of scale changes. In particular, we have the following consequence of Condition 3.

Lemma 8.2: If $S' = \{(a_1 x_1, \dots, a_n x_n) | x \in S\}$ where $a_j > 0$ for $j = 1, \dots, n$ then $\theta_i(S') = a_i \theta_i(S)$ for all $i \in N$.

Proof: If $a_i \geq 1$, then Condition 3 implies that

$$\theta_i(S) = \theta\left[\frac{1}{a_i}(i, S'); \left(1 - \frac{1}{a_i}\right)(i, A_0)\right] = \frac{1}{a_i} \theta_i(S') \quad .$$

¹⁶The motivation for this condition is that, since all the players in the game A agree that x is the best feasible outcome, there should be no difficulty in reaching an agreement that x will be the final outcome of the game.

¹⁷For any real number c , we denote the vector (c, \dots, c) by \bar{c} .

while if $0 < a_1 < 1$ then $S = \{(b_1 x_1, \dots, b_n x_n) \mid x \in S^1\}$ where each $b_j = 1/a_j$, so $b_1 \geq 1$ and the first part of the proof applies.

Lemmas 8.1 and 8.2 together with the normalization $\theta(i, A_1) = 1$ imply that $\theta_1(A_x) = x_1$ for any game A_x , and any $i \in N$.

Thus the conditions imposed so far essentially embed the utility function for prizes into the utility function for games, but this is not yet sufficient to uniquely determine the preference relation R over the mixture set M^1 . To do this, it is necessary to specify the individual's attitude towards the canonical game $\bar{A} = \{x \geq \bar{0} \mid \sum x_i \leq n\}$. We will assume that $(i, A_n)R(i, \bar{A})R(i, A_0)$, and let c be the number between 0 and n such that

Condition 5: $(i, \bar{A})I(i, A_c)$.

Then c is the individual's certain equivalent for bargaining in the symmetric game \bar{A} ; that is, he is indifferent between bargaining in the game \bar{A} or receiving the utility c for certain (in the game A_c). Of course it follows immediately that $\theta(i, \bar{A}) = c$, and we will say that the quantity c represents the individual's strategic risk posture;¹⁸ i.e., it reflects his attitude towards the risk involved in bargaining. If $c = 1$ then he is neutral to strategic risk (since in the game \bar{A} , 1 is the (arithmetic) average award which the players receive if a Pareto optimal agreement is reached). If $c < 1$ then the individual is strategic risk averse, and if $c > 1$ then he is strategic risk preferring.

For instance, an individual might be neutral to strategic risk if he felt that he was an average bargainer, and that a Pareto optimal agreement was sure to be reached in the game \bar{A} . On the other hand, an individual who felt that there was some chance that bargaining in the game \bar{A} might result in a non-Pareto optimal outcome (e.g., who felt there was some chance that the disagreement outcome would result) might therefore be averse to strategic risk. That is, he would strictly prefer to get a utility payoff of 1 for certain, rather than to have to take his chances in the game \bar{A} .

¹⁸The notion of strategic risk posture was introduced in Roth [1977b], [1977c].

The strategic risk posture, as captured by the certain equivalent c , completely determines the utility function θ . That is, for a preference relation obeying Conditions 1 through 5, the (normalized) utility function θ is uniquely determined, as follows.

Theorem 8: For any $S \in B^1$, $k \in N$, $\theta_k(S) = x_k$ where x is the unique element of S^+ such that $\Pi x_i^{q_i} > \Pi y_i^{q_i}$ for all $y \in S^+$ such that $y \neq x$, where $q = (q_1, \dots, q_n)$ is the vector such that $q_k = c$, $\sum q_i = n$, and $q_i = q_j$ for $i, j \neq k$.

That is, for each $k \in N$, $\theta_k(S)$ is equal to the k -th component of the non-symmetric Nash solution f such that $f(\bar{A}, \bar{0}) = q$.¹⁹

Proof: What we already know about the non-symmetric Nash solutions tells us that the function θ_k defined in the theorem does behave in accordance with Conditions 2 through 5, as does the k -th component of any non-symmetric Nash solution g such that $g_k(\bar{A}, \bar{0}) = q_k = c$. However only if $g_i(\bar{A}, \bar{0}) = g_j(\bar{A}, \bar{0})$ for $i, j \neq k$ will the function g_k obey Condition 1. (To see this, consider the set S' equal to the convex hull of the points $(1,0,1)$, $(0,1,0)$ and $(0,0,0)$. Then Condition 1 requires that $\theta_1(S') = \theta_1(S)$, where S is the game described at the end of section B.) The proof that θ_k must in fact be equal to the k -th component of the non-symmetric Nash solution so described proceeds, with the help of Condition 4, precisely as in Theorem 3.

Note that (unless $c = 1$, in which case the vector q described in the theorem is the vector $\bar{1}$) each component of $\theta(S)$ is equal to the corresponding component of a different non-symmetric Nash solution. The vector $\theta(S)$ is not itself a Nash solution.²⁰ Note also that if S is a symmetric game, then $\theta_1(S) = \dots = \theta_n(S)$, regardless of the risk posture which it expresses. (For instance if $c = 3/4$, then $\theta(\bar{A}) = (3/4, \dots, 3/4)$.)

¹⁹The condition $q_i = q_j$ for $i, j \neq k$ was incorrectly omitted in Roth [1978], and I corrected it in a subsequent erratum.

²⁰In general, the vector $\theta(S)$ need not be a solution at all (i.e., it need not be an element of S). For instance a risk preferring individual with $c = 3/2$ has the utility vector $\theta(\bar{A}) = (3/2, \dots, 3/2)$, which is not an element of \bar{A} .

When $c = 1$, however, the vector $\theta(S)$ is equal to the Nash solution. Thus Theorem 8 provides the following characterization of the Nash solution.

Corollary 8.1: The Nash solution is equal to the utility vector θ reflecting strategic risk neutrality.

Since the Nash solution can be thought of as a utility function which is neutral to strategic risk, it is interesting to investigate its behavior with respect to ordinary risk--i.e., to lotteries between games. Of course, the function θ we have been studying is an expected utility function, so that the utility of a lottery between any games S and S' is its expected utility; e.g., $\theta[1/2(i,S); 1/2(i,S')] = 1/2\theta_i(S) + 1/2\theta_i(S')$. To see what kind of attitude towards risk this involves, we want to compare the utility of the lottery with the utility $\theta_i(T)$ of the "expected game" $T = (1/2)S + (1/2)S' = \{y | y = (1/2)x + (1/2)x' \text{ for some } x \text{ in } S, x' \text{ in } S'\}$.

If S and S' are games of the form A_x and $A_{x'}$, then θ would be linear: i.e., $(1/2)\theta(A_x) + (1/2)\theta(A_{x'}) = \theta(A_y)$ where $y = (1/2)x + (1/2)x'$ (i.e., $A_y = 1/2A_x + 1/2A_{x'}$). For arbitrary sets S and S' , however, the utility is not in general linear.

For example, in the case $n = 2$, let θ be risk neutral, so that it equals the Nash solution; let S be the convex hull of $(2,0)$, $(0,1)$ and $(0,0)$; and let S' be the convex hull of $(1,0)$, $(0,2)$ and $(0,0)$.

Then the set $T = (1/2)S + (1/2)S'$ is equal to the convex hull of the points $(1\frac{1}{2}, 0)$, $(0, 1\frac{1}{2})$, $(1,1)$, $(0,0)$. However $1/2\theta(S) + 1/2\theta(S') = (3/4, 3/4) < \theta(T) = (1,1)$. (See Figure 7, p. 59).

The subadditivity exhibited in this example²¹ has a simple and intuitive explanation in terms of the underlying bargaining process. The lottery $[1/2(i,S); 1/2(i,S')]$ can be thought of as one in which a coin is tossed to decide which of the two-player games, S or S' , should be played. After the coin is tossed, the individual whose preferences we are studying will play position i ($i = 1$ or 2) in the game selected, against some opponent. Since $F(S,0) = (1, 1/2)$ and $F(S',0) = (1/2, 1)$, this lottery yields an expected utility of $3/4$.

²¹Cf. Harsanyi [1968].

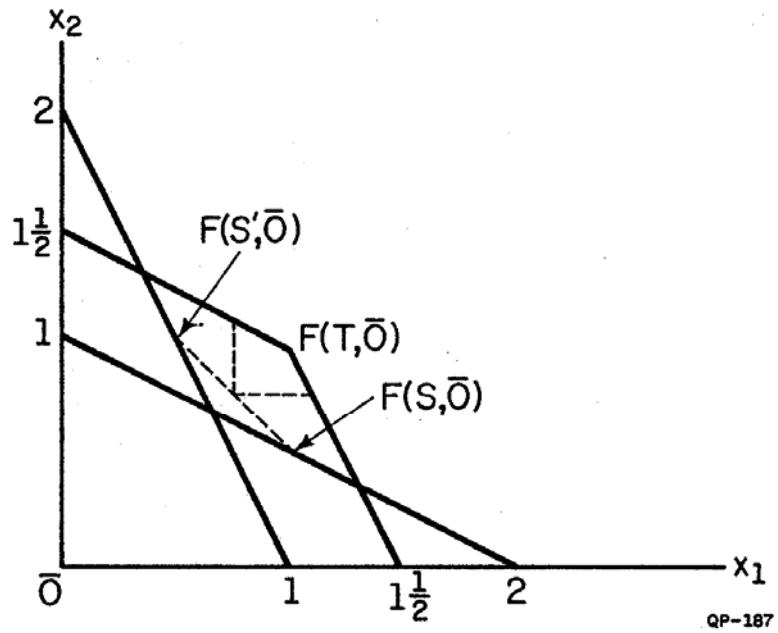


Figure 7

The game $T = \{(1/2)x + (1/2)x' \mid x \in S, x' \in S'\}$, on the other hand, can be interpreted as the game in which the players reach an agreement before the coin is tossed. Specifically, every utility vector y in T is of the form $y = (1/2)x + (1/2)x'$ where x and x' are utility vectors in S and S' . Since x and x' are expected utility vectors, y can be interpreted as the expected utility resulting from the contingent agreement of the form "choose x if the game S is selected, and choose x' if the game S' is selected." The set T of feasible utility payoffs is symmetric, and the payoff $F(T, \bar{0}) = (1,1)$ results from the contingent agreement that either the point $x = (2,0)$ or $x' = (0,2)$ will be chosen. Thus in this example, the risk neutral utility function equal to the Nash solution reflects a preference for bargaining before the coin is tossed rather than after.

The intuitive explanation is that, before the coin is tossed, the players can agree to choose one of the (extreme) payoffs $x = (2,0)$ and $x' = (0,2)$ since both players have an equal chance of receiving the higher payoff. If bargaining takes place after the coin toss, this form of compromise will be unavailable since only one of the points x and x' will be feasible.²²

²²Note, however, that it is not the case for arbitrary sets S and S' that $(1/2)F(S, \bar{0}) + (1/2)F(S', \bar{0})$ is always less than or equal to $F(T, \bar{0})$, for $T = (1/2)S + (1/2)S'$. A class of solutions which has this property on a subset of the class B' of games is studied by Perles and Maschler [1979].

PART II: OTHER MODELS OF BARGAINING

A. A Critical Evaluation of the Independence Properties

In Part I we saw that an individually rational solution f which is independent of equivalent utility representations and independent of irrelevant alternatives can be uniquely determined by specifying the outcome which it selects in the 'canonical' game (\bar{A}, \bar{O}) . Each of the two independence properties represents an assumption concerning the way in which different bargaining games are related to one another, and together they essentially imply that the process of bargaining in an arbitrary game is no different from the bargaining which occurs in the game (\bar{A}, \bar{O}) , or at least no different in any way which affects the final agreement.

Insofar as a solution is intended to be a descriptive model of the bargaining process, these assumptions can be tested empirically,¹ and in addition, both of the independence properties can also be challenged on conceptual grounds. In this section we will consider some of the objections which can be raised against the assumptions implicit in the two independence properties, and in the subsequent sections we will consider the consequences of replacing them with alternative assumptions.

Independence of Equivalent Utility Representations

The assumption that a solution (which models a given bargaining process) is independent of equivalent utility representations can be interpreted in several ways, depending on what information is assumed to be available to the bargainers. The traditional game-theoretic assumption is that the players are completely informed of one another's preferences, as captured by their utility functions.

¹There has been a fair amount of empirical work done which relates to bargaining, but much of it is in a form which makes it difficult to connect with the kinds of pure bargaining situations we are considering here. Even the various experimental investigations which have been explicitly designed to evaluate Nash's solution have been conducted under such a diversity of conditions that it is difficult to draw systematic conclusions from their results. For a sample of the empirical work, see Crott [1971], de Menil [1971], Fouraker and Siegel [1963], Nydegger and Owen [1975], O'Neill [1976], Rapoport, et. al [1976], [1977], Siegel and Fouraker [1960], and Stone [1958]. Some of this work is reviewed in Roth and Malouf [1979].

However in many cases of practical interest, the bargainers have much less complete information about the preferences of the other bargainers. In particular, we shall consider bargaining situations in which players know one another's preferences only over riskless events, but not over lotteries.

In such situations it might still happen that the bargaining process depends on the attitude of the players towards risk, even though each player knows only his own risk posture. That is, each player's negotiating behavior can still be influenced by his own risk posture, which in turn could be reflected in the final agreement. However it is also plausible that, in situations of this sort, the attitude of the players towards risk would influence the bargaining process only indirectly and incompletely, or even not at all.² In this case we would expect that bargaining would be independent of a wider class of transformations than those required by property 1. That is, for bargaining games in which a player's attitude towards risk is of no consequence, a larger class of utility functions are equivalent to one another. In Section B of Part II we will consider the consequences this has for a theory of bargaining.

We will also want to consider bargaining situations in which the players not only have complete information about one another's preferences and risk postures, but also possess other information which may effect the bargaining process. Note that the requirement that a solution be independent of equivalent utility representations and also symmetric constitutes a requirement that it be completely insensitive to any information other than that contained in the utility functions of the players. Even without the symmetry property, so long as a bargaining game is specified only by its utility payoffs, a solution which is independent of equivalent utility representations makes no distinction between different bargaining situations which may be perfectly distinguishable to the participants.

²Some of the experimental evidence can be interpreted as supporting this latter conclusion. Of course the details of how the negotiating process is implemented in an experimental bargaining situation determine how much information can be indirectly conveyed through the behavior of the bargainers, and so it is difficult to draw any very general conclusions on the basis of limited experimental evidence.

Consider, for instance, the following two-player bargaining problem. Players 1 and 2 bargain over a probability $p \in [0,1]$ to determine a lottery which, with probability p , gives player 1 a prize of c_1 dollars and player 2 zero dollars; and, with probability $(1-p)$, gives a prize of c_2 dollars to player 2 and zero dollars to player 1. The players are free to agree on any probability p between 0 and 1, and if they fail to reach an agreement, each player receives zero dollars. Then if players 1 and 2 start with initial wealth w_1 and w_2 respectively, and if their utility functions for money are u_1 and u_2 , then this bargaining game can be represented in the space of utility payoffs by the pair (S,d) where $d = (u_1(w_1), u_2(w_2))$ and $S = \{(x_1, x_2) \geq d \mid x_1 \leq pu_1(w_1 + c_1) + (1-p)u_1(w_1) \text{ and } x_2 \leq (1-p)u_2(w_2 + c_2) + pu_2(w_2) \text{ for } 0 \leq p \leq 1\}$. That is, S is the convex hull of the three points d , $(u_1(w_1 + c_1), u_2(w_2))$ and $(u_1(w_1), u_2(w_2 + c_2))$.

If we choose the origin and scale of the utility functions u_1 and u_2 so that $u_1(w_1 + c_1) = u_2(w_2 + c_2) = 2$ and $u_1(w_1) = u_2(w_2) = 0$, then S is the convex hull of the points $(0,0)$, $(2,0)$ and $(0,2)$, and so $(S,d) = (\overline{A}, \overline{0})$. Consequently $F(S,d) = (1,1) = ((1/2)u_1(w_1 + c_1), (1/2)u_2(w_2 + c_2))$. That is the Nash solution $F(S,d)$ selects the utility vector corresponding to agreement on the probability $p = 1/2$.

Of course the same bargaining situation is equally well represented by the utility payoffs available in terms of the utility functions u_1 and v_2 , where u_1 is player 1's utility function normalized as above, while v_2 is a utility function for player 2 normalized so that $v_2(w_2) = 0$ and $v_2(w_2 + c_2) = 100$. That is $v_2 = 50u_2$ is simply an alternative representation of player 2's preferences, since the choice of scale is arbitrary. Then the set of feasible utility payoffs in terms of the utility functions u_1 and v_2 is simply the set S' equal to the convex hull of the points $(0,0)$, $(2,0)$, and $(0,100)$, where $(0,0)$ is still the disagreement payoff. So, since the Nash solution F is independent of equivalent utility representations, it follows that $F(S',d) = (1,50)$, which in the game (S',d) corresponds as before to the agreement on the probability $p = 1/2$. That is, the games (S,d) and (S',d) as described here are equivalent representations of the same bargaining situation in which the potential cash awards are c_1 and c_2 , and so the Nash solution makes no distinction between them. The underlying bargaining situation can be represented

as a symmetric game in utility space, and the Nash solution selects the agreement at which each player has an equal chance of receiving his prize, and so player i 's expected monetary reward is equal to $c_1/2$ for $i=1,2$.

Now suppose the underlying bargaining situation is changed by increasing the amount of money which player 2 may receive, from c_2 dollars to some larger amount c_2' such that $u_2(w_2 + c_2') = 100$. That is, c_2' is so much larger than c_2 that player 2 would be indifferent between receiving c_2 for certain, or participating in a lottery which gave him a $1/50$ probability of winning c_2' and $49/50$ probability of winning nothing.

Let (T, d') be the resulting bargaining game when the utility payoffs are expressed in terms of the (original) utility functions u_1 and u_2 . Then $(T, d') = (S', d)$, and so $F(T, d') = (1, 50)$, corresponding to agreement on the probability $p = 1/2$ (see figure 8, p. 65). So the same property of the Nash solution which makes it insensitive to the purely formal change in representation from u_2 to v_2 also makes it insensitive to the actual change from c_2 to c_2' . That is, the model of the game in utility space, where the origins and scales of the payoffs are arbitrary, is a model which takes no account of the magnitude of c_1 and c_2 . Indeed, the game is completely defined from the point of view of this model even before c_1 and c_2 have been specified.

However the available empirical evidence suggests that the size of the potential monetary gains of the players may in fact affect the outcome of bargaining when both players have complete information about the size of the prizes.³ For instance, an experimental study using games of the form just discussed is reported in Roth and Malouf [1979]. In that study, one of the experimental manipulations involved the information made available to each player about the monetary prizes c_1 and c_2 . In the "high information" condition, each player knew the monetary value of both prizes, while in the "low information" condition, each player was informed only of the value of his own prize. For the games studied in this experiment, it was found that the agreements reached in the low information condition,

³ See e.g., Nydegger and Owen [1975], Rapoport et al. [1976], [1977].

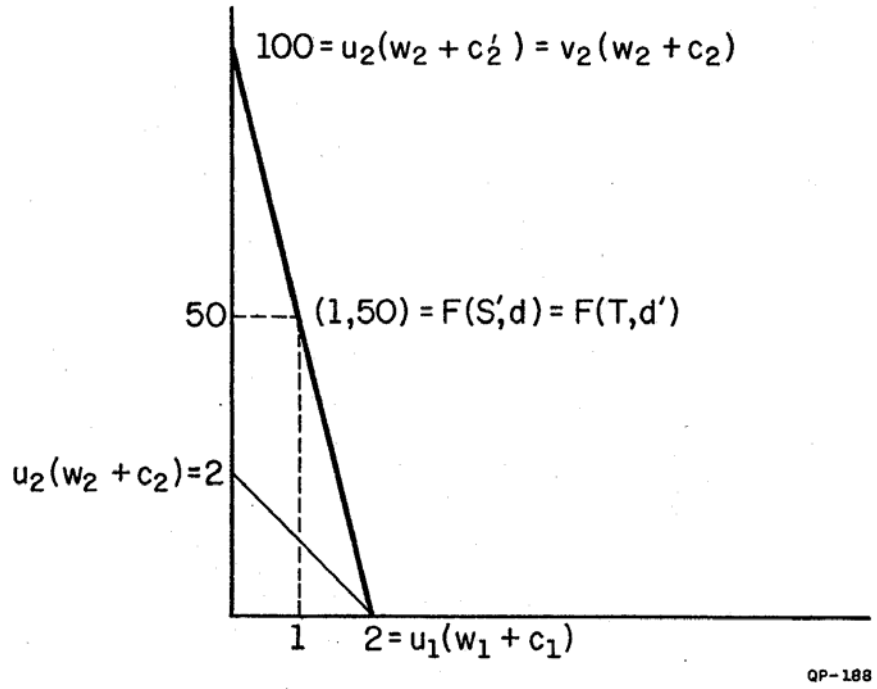


Figure 8

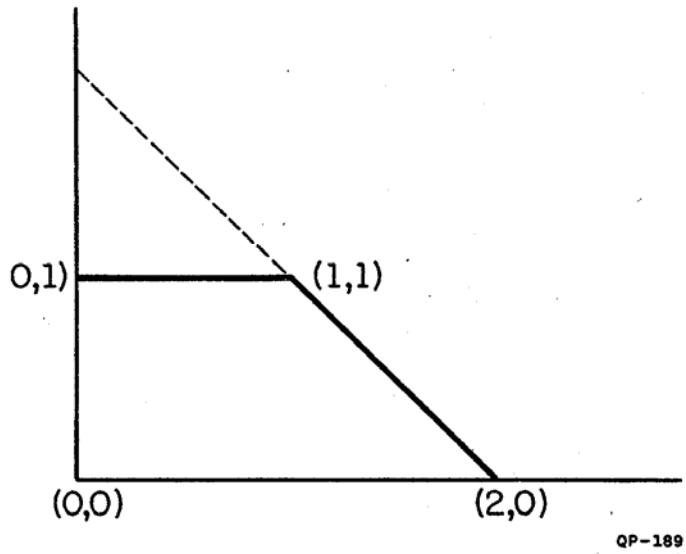


Figure 9

regardless of the values of c_1 and c_2 , were not significantly different from the agreement predicted by Nash's solution, at which each player 1 received a probability of 1/2 of receiving his potential prize c_1 . For games in which the prizes c_1 and c_2 were equal, it was found that Nash's solution was also a good predictor in the high information condition; i.e., the agreements reached continued to give each player an equal probability of winning his own prize. However for games in which c_1 and c_2 were unequal, the agreements reached in the high information condition were significantly different from the agreement predicted by Nash's solution. In particular, the agreements reached tended to give a higher probability of winning his own prize to the player with the smaller monetary prize, although not so much higher as to equalize the expected monetary value of the lottery which each player received. Thus in the high information condition, the outcomes reached in the games studied seem to have been influenced by comparisons of the monetary payoffs available to the players.

It is not hard to see why this might be the case. When the players know the magnitude of the potential monetary gains, an element is introduced into the bargaining process which is absent from Nash's model. For instance, the agreement at which players get equal gains is well defined in a monetary context, but has no meaning at all in terms of utility functions each of which is scaled arbitrarily. That is, if neither player knows the size of the other player's monetary prize then the only property of agreements which the players can negotiate about is the probability p . When the players each know the size of both prizes, then they can also describe agreements by the expected monetary gain which results. Thus the additional information enriches the language available in the negotiations towards reaching an agreement,⁴ and we will investigate some ways in which this might affect the agreement reached.

⁴If we wanted to model this bargaining situation as a noncooperative game, the difference between the two information conditions would be that the strategy sets of the players are larger in the high information condition. That is, certain bargaining strategies are available in the high information condition which are unavailable in the low information condition.

In Section C we will explore some models of bargaining which permit at least some well-defined comparisons to be made of the gains which an agreement offers to the different players.

Independence of Irrelevant Alternatives

Much of the criticism directed at the Nash solution has been concerned with its independence of irrelevant alternatives, which has also been the property most difficult to motivate.⁵ This property requires that a solution select an outcome using criteria which depend only on the disagreement point and the selected outcome itself, and not on any other outcomes in the feasible set.⁶ By calling other outcomes "irrelevant," the property's name reflects the assumption which it makes regarding the bargaining process.

There is some empirical evidence on each side of the question of how realistically this assumption models various forms of bargaining, but it is not difficult to argue that the entire range of feasible outcomes available in a game might reasonably affect the negotiations which take place, if only by affecting the initial aspirations of the bargainers. Consider for instance the game $(S, \bar{0})$ shown in Figure 9 (p. 65), whose feasible set is contained in the set \bar{A} . The Nash solution predicts that, since $F(\bar{A}, \bar{0}) = (1, 1)$ is contained in S , this will be the point selected in the game $(S, \bar{0})$ as well, even though the symmetry which might encourage the players to agree on this point in the game $(\bar{A}, \bar{0})$ is no longer present in the game $(S, \bar{0})$.

In Sections B and D we will consider some models which do not incorporate the assumption of independence of irrelevant alternatives, including some models which explicitly involve the "aspiration levels" which the players in a game might form.

⁵See e.g., Luce and Raiffa [1958].

⁶If we view a solution as a function which sets up some binary relation between alternative outcomes, and then selects the maximal element of this relation in the feasible set, then the property of independence of irrelevant alternatives specifies that, for a fixed disagreement point, the binary relation is independent of the feasible set. That is, the criteria in terms of which such a solution compares alternative outcomes are the same for all possible feasible sets.

B. Ordinal Models of Bargaining

In this section we will investigate the possibility of developing a theory of bargaining which depends only on the ordinal information contained in the players' utility functions.¹ By ordinal information, we mean information about each player's preference ordering over riskless alternatives. If A is a set of riskless alternatives over which player i has a preference ordering, then any real-valued function u_i on A such that $u_i(a) > u_i(b)$ if and only if player i prefers a to b is an ordinal utility function for player i . Thus if u_i is a utility function representing player i 's preferences on A , then v_i is also such a utility function if and only if $v_i = m_i(u_i)$ ² where m_i is a monotone increasing (i.e., order preserving) function from the set of real numbers to itself. That is, for any real numbers x and y , $m_i(x) > m_i(y)$ if and only if $x > y$. We will confine our attention to continuous (invertible) order preserving transformations, and for any x in R^n denote $(m_1(x_1), \dots, m_n(x_n))$ by $m(x)$.

As in Part I, an n -player bargaining game can be represented in the space of utility payoffs by the pair (S, d) , where S is a compact subset of R^n containing d and at least one point x such that $x > d$. However, since we will no longer be considering alternatives which involve lotteries, the requirement that S be a convex set is no longer a natural one. We will impose a weaker restriction on bargaining games (S, d) and require only that the set of strongly Pareto optimal outcomes in S be contractible (i.e., that it contains no "holes"). For two player games, this is equivalent to requiring that each player's utility payoff in the set $P(S)$ can be

¹Our attention only to ordinal properties can be interpreted in two different ways. On the one hand, we can assume that the players' choice behavior can be fully modelled by a cardinal (expected) utility function, but that only the ordinal properties of this utility function are relevant to the bargaining process in the situations we are considering. Alternatively, we can consider the case where each player's choice behavior is insufficiently regular to be accurately represented by a cardinal utility function (cf. p. 2), but obeys the weaker conditions needed for representation by a purely ordinal utility function.

²That is for any a in A , $v_i(a) = m_i(u_i(a))$.

described as a decreasing continuous function of the other player's utility.³

Formally, let B^* denote the set of all non-degenerate bargaining games (S, d) of this kind.⁴

Then a solution is a function $f: B^* \rightarrow R^n$ such that $f(S, d) \in S$, and it is independent of ordinal transformations if it possesses the following property.

Property 9. Independence of Ordinal Transformations: For any bargaining game (S, d) in B^* and any continuous, order preserving functions m_i , $i = 1, \dots, n$, let the bargaining game (S', d') be defined by $S' = m(S) \equiv \{y \in R^n \mid y = m(x) \text{ for some } x \text{ in } S\}$ and $d' = m(d)$. Then $f_i(S', d') = m_i(f_i(S, d))$ for $i = 1, \dots, n$.

Note that Property 9 is a stronger independence condition than Property 1, in the sense that the positive linear transformations described in Property 1 are a strict subset of the monotone transformations described in Property 9. That is, any solution which possesses Property 9 also possesses Property 1.

Since the (non-symmetric) Nash solutions possess Property 1 but not Property 9, and since the class B of games with convex payoff sets is contained in the class B^* , we can draw the following conclusion.

Theorem 9: No ordinally independent solution exists which also possesses the properties of strong individual rationality and independence of irrelevant alternatives.

That is, no solution which possesses Property 9 can also possess Properties 3 and 6.

³This would be the case, for instance, for games whose outcomes correspond to the division of divisible commodities. See, however, the last two paragraphs of this section.

⁴Enlarging the class of allowable games reduces the set of solutions which are defined on that class. Since some of our results in this section shall be impossibility results (e.g., Theorem 9 and 10, and Corollary 10.1) we should note that these results would continue to hold even on the class B of bargaining games considered in Part I. The proofs of these results have been constructed so as to make this perfectly clear.

Proof: Suppose that, contrary to the theorem, a solution f existed which possessed the properties named in the theorem. Then since it is ordinally independent, it is also independent of equivalent utility representations, and so the argument of part (ii) in the proof of Theorem 2 can easily be extended to the non-symmetric case to show that f must be Pareto optimal. Then by Theorem 3 this solution must coincide with a non-symmetric Nash solution for games with convex payoff sets. Since these solutions are not independent of arbitrary monotone transformations (e.g., the concave transformations considered in Section D of Part I), this provides the contradiction needed to complete the proof.

In fact, Theorem 9 can be considerably strengthened: Shapley [1969] proved the following result for the case $n = 2$.

Theorem 10: For two-player games, no ordinally independent solution exists which also possesses the property of strong individual rationality.

Proof: Consider the 2-player game (S, d) , where S is the convex hull of the points $(0,0)$, $(1,0)$ and $(0,1)$, and $d = (0,0)$. Let (S', d') be the game derived from (S, d) as in the statement of Property 9 by the transformations $m_1(x) = 2x/(1+x)$, $m_2(x) = x/(2-x)$ defined on the interval $0 \leq x \leq 1$. It is easily verified that both m_1 and m_2 are order preserving on this interval, that $m_1(0) = m_2(0) = 0$ and $m_1(1) = m_2(1) = 1$, and that $m_1(x) > x$ and $m_2(x) < x$ for $0 < x < 1$. Also, if $x_1 + x_2 = 1$, then $m_1(x_1) + m_2(x_2) = 1$ as well, and so $S' = m(S) \equiv \{y \mid y = m(x) \text{ for some } x \in S\} = S$. That is, the transformation $m = (m_1, m_2)$ maps the set S into itself, leaving only the points $(0,0)$, $(1,0)$, and $(0,1)$ fixed (see Figure 10, p. 71).

So $(S, d) = (S', d')$; consequently $f(S, d) = f(S', d')$. If f obeys Property 9, then $f(S, d) = m(f(S, d))$ as well: i.e., $f(S, d)$ is a fixed-point of m . But none of the three fixed-points of m contained in S is strictly positive, and so we have shown that ordinal independence is incompatible with strong individual rationality, as stated in the theorem.

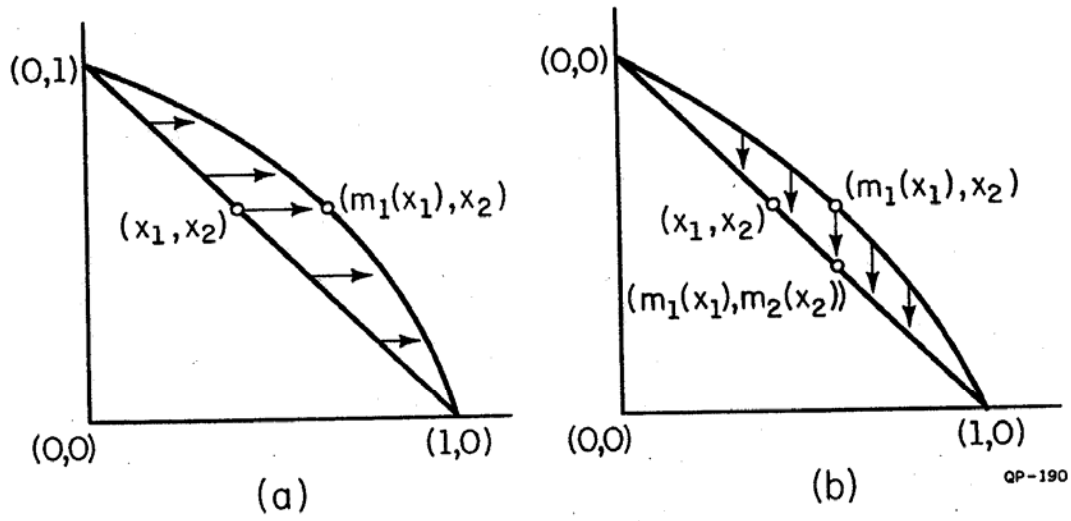


Figure 10

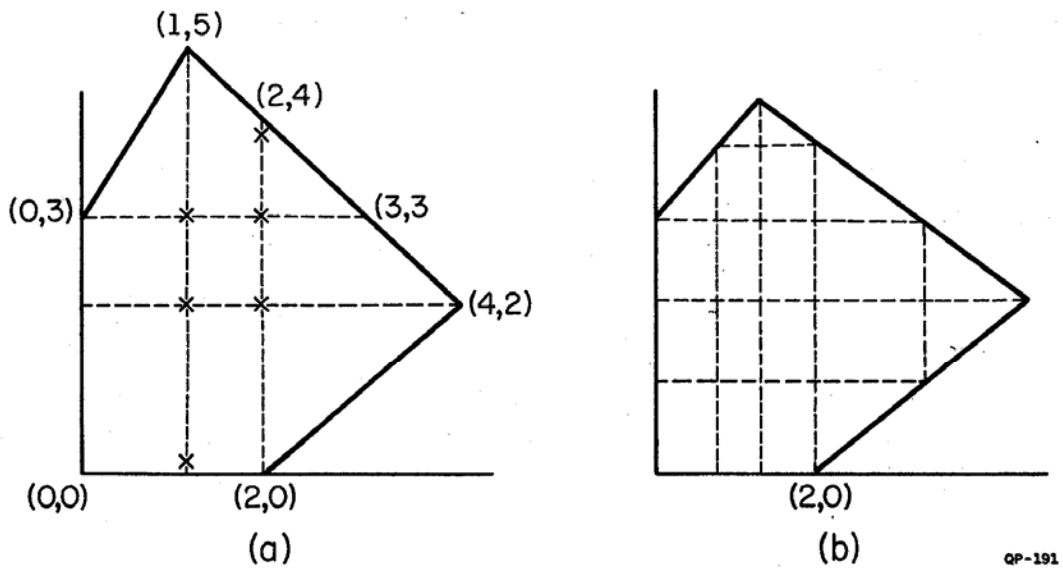


Figure 11

The theorem we have just proved reflects the fact that two-player games of the kind considered in the proof can be ordinally transformed into themselves without leaving "enough" fixed-points on which to base an ordinally-independent theory. It is not the case, however, that all two-player games yield as few fixed points.

Consider, for example, the game $(S, \bar{0})$ where S is the convex hull of the points $(0,0)$, $(2,0)$, $(4,2)$, $(1,5)$, and $(0,3)$. (See Figure 11a, p. 71.)

If $m = (m_1, m_2)$ is a pair of monotone transformations such that $(S, d) = (m(S), m(d))$, then each of the five extreme points of S must be a fixed point of S , since each can be described by purely ordinal properties.⁵ For instance, the point $(2,0)$ is player 1's most preferred outcome in the set of outcomes which player 2 finds indifferent to the disagreement point, while the point $(4,2)$ is player 1's most preferred outcome in the set of Pareto optimal outcomes. These are properties which must be preserved by any ordinal transformations of the players' utility functions.

Each of these fixed-points serves as a sort of benchmark from which other fixed points can be determined. For instance, the point $(2,4)$ can be described in purely ordinal terms by noting that it is the unique Pareto optimal point which gives player 1 the same utility as the fixed point $(2,0)$. Each of these "second-generation" fixed points is indicated by an x in Figure 11a, while Figure 11b shows the full set of third and fourth generation fixed points as well. That is, each of the 24 points shown in Figure 11b can be specified by purely ordinal properties, and so each of these points is fixed under the transformation m .

If we could specify a subclass of the games contained in B^* which were "sufficiently rich" in fixed points, then we might hope to be able to find reasonably well behaved, ordinally-independent solutions on this subclass. Shapley and Shubik [1974] have shown, in fact, that for games with at least

⁵That is, these points can be uniquely described using only ordinal comparisons ($>$, $=$, $<$), which are independent of ordinal transformations.

three players, the subclass of games whose outcome sets are "full-dimensional"⁶ admits ordinally independent solutions which are both symmetric and Pareto optimal, as well as strongly individually rational. In particular, they construct a solution with these properties for the case $n = 3$.

To get an idea of how the two-person case is different from the n -person case for $n \geq 3$ with respect to constructing an ordinally independent solution, recall that in the two-person case it is possible to transform the simplex into itself using monotone transformations on each player's utility, in such a way that no interior points remain fixed. It is also possible, using monotone transformations of each player's utility, to transform any Pareto surface so that all of its points lie on the simplex: i.e., so that every point (x_1, x_2) is transformed into a point $(m_1(x_1), m_2(x_2))$ such that $m_1(x_1) + m_2(x_2) = 1$. (See Figure 10(b), p. 71).

Neither of these observations carries over into the case $n \geq 3$. In three or more dimensions the simplex cannot be mapped non-trivially into itself by means of monotone transformations of the players' utility functions, and arbitrary Pareto surfaces cannot be mapped into the simplex.⁷

To demonstrate this latter observation, consider a Pareto surface containing the points $(0, 2, 4)$, $(2, 4, 0)$ and $(4, 0, 2)$ as well as the points $(1, 5, 3)$, $(3, 1, 5)$ and $(5, 3, 1)$. Suppose some monotone transformation $m = (m_1, m_2, m_3)$ did map these points into the simplex, then it would satisfy the following equations

$$m_1(0) + m_2(2) + m_3(4) = 1 = m_1(1) + m_2(5) + m_3(3)$$

$$m_1(2) + m_2(4) + m_3(0) = 1 = m_1(3) + m_2(1) + m_3(5)$$

$$m_1(4) + m_2(0) + m_3(2) = 1 = m_1(5) + m_2(3) + m_3(1)$$

However these equations cannot be simultaneously satisfied, since the monotonicity of each m_i implies that the sum of the nine terms on the left is strictly less

⁶ Specifically, they consider only Pareto optimal sets which are the intersection with the nonnegative orthant of an open set containing the origin.

⁷ That Pareto surfaces cannot, in general, be mapped into the simplex is an observation due to Bradley and Shubik [1974]. Our exposition of this point closely follows that of Shapley and Shubik [1974].

than the sum of the nine terms on the right, since each term $m_1(k)$ on the left can be paired with a larger term $m_1(k+1)$ on the right (i.e., $m_1(0) < m_1(1)$, $m_1(2) < m_1(3)$, $m_1(4) < m_1(5)$, $m_2(2) < m_2(3)$, etc.) So there is no vector m of monotone transformations of each player's utility function which maps these six points into the simplex.⁸

So when there are more than two players, the simplex and other full dimensional bargaining surfaces possess sufficient "rigidity" with respect to ordinal transformations so as to admit symmetric ordinal solutions. However, regardless of the number of bargainers, the following corollary of Theorem 10 shows that no satisfactory ordinally invariant solution can be defined on the entire class B^* of bargaining games.

Corollary 10.1: For any n , there is no solution defined on the class B^* which is both ordinally independent and strongly individually rational.

Proof: In Theorem 10 we have already considered the case $n = 2$, so let $n = k + 1$ for some $k \geq 2$. Consider the game $(S, \bar{0})$ where S is equal to the convex hull of the set $\{(p/k, \dots, p/k, (1-p)) \mid 0 \leq p \leq 1\}$ and the origin. Let $m = (m_1, \dots, m_k, m_{k+1})$ be the vector of monotone transformations such that for $i = 1, \dots, k$, $m_i(x) = (1/k) - (1 - kx)/(k - kx)$ on the interval $0 \leq x \leq 1/k$, and $m_{k+1}(x) = kx/(k - 1 + x)$, on the interval $0 \leq x \leq 1$. Then the vector m of transformations takes the set S into itself, leaving only the origin and the points $(0, \dots, 0, 1)$ and $(1/k, \dots, 1/k, 0)$ fixed. As in the proof of Theorem 2, any ordinally independent solution must therefore select one of these points. Since none of these points is strictly positive, the corollary is proved.

Note that although the game considered in the proof of the corollary is very much less than full dimensional, it is not at all degenerate in the game theoretic sense. In particular, no player is indifferent between any two Pareto-optimal

⁸ Bradley and Shubik show that, in three dimensions, at least six points are needed to observe this phenomenon. If $n \geq 4$, it is possible to exhibit four points which cannot all be transformed into the simplex.

outcomes of the game, so that it represents a situation in which each player has an active interest in the bargaining. Thus, regardless of the number of players, any ordinally independent solution will only behave in a satisfactory manner on a subclass of games which excludes at least some games which arise in a natural way.

It can nevertheless be illuminating to consider the consequences of ordinal invariance even over a restricted class of games. For instance, consider the class of games which can arise when n individuals ($n > 2$) with initial wealth w_i ($i = 1, \dots, n$) must bargain over how to split a positive quantity Q of money. When we considered such games before (Part I, Section D) we assumed that a utility function for money was an increasing concave function. Since concavity is not preserved under arbitrary monotone transformations, a utility function for money in our present context can be any increasing function. That is, from the point of view of an ordinally independent solution, any increasing function (of money) is equivalent to any other.

Specifically, for an individual with initial wealth w_i , any utility function for money is equivalent to the function $u_i(w_i + c) = nc/Q$, and so the set of Pareto optimal outcomes for a bargaining game over money is ordinally equivalent to the simplex. That is, if (S, d) is a bargaining game for money, then there exists a vector $m = (m_1, \dots, m_n)$ of monotone transformations of each player's utility function such that $m(S, d) = (\bar{A}, \bar{0})$.⁹ If f is an ordinally independent solution which is

⁹Note that it is immaterial whether we consider the unit simplex $S = \{x \geq 0 \mid \sum x_i = 1\}$ or the set $P(A) = \{x \geq 0 \mid \sum x_i = n\} = nS$.

symmetric and Pareto optimal¹⁰ (on at least the class of monetary games), then $f(\bar{A}, \bar{0}) = (1, \dots, 1)$, corresponding to the monetary split $c_i^* = Q/n$ for $i = 1, \dots, n$. Consequently $f_i(S, d) = m_i^{-1} f_i(\bar{A}, \bar{0}) = m_i^{-1}(1) = m_i^{-1}(u_i(w_i + c_i^*))$ for $i = 1, \dots, n$, where m_i^{-1} denotes the inverse of m_i , and u_i is the utility function defined at the beginning of the paragraph. So for any game (S, d) representing a situation in which the players can split a given sum of money, the solution f selects the utility vector corresponding to an equal division of money among all n bargainers.

In some respects this is a curious phenomenon. In Part I we observed that, when we take into account the attitudes of the players towards risk, bargaining games for money can become very asymmetric, and the (symmetric) Nash solution can lead to an unequal division of money. In the previous section we observed that only when additional information is added to the models considered in Part I does it become meaningful to compare the gains which a given agreement gives to different players. In this section we considered a model which contains strictly less information than the model of Part I, since it makes no use of the players' risk posture. But, even though the model is concerned only with the set of utility payoffs, it now becomes possible in the class of monetary games to identify the outcome corresponding to an equal division of the available money among the players,¹¹ and this is the unique outcome selected by an ordinally indepen-

¹⁰To formally demonstrate that such a solution exists, we need at least to show that the point $(1, \dots, 1)$ is fixed under monotone transformations which take \bar{A} into itself. For the case $n = 3$, suppose that $m_1, m_2,$ and m_3 are order-preserving transformations which take \bar{A} into itself, and let v be the image under m of the point $\bar{1} = (1, 1, 1)$; i.e., $v = (m_1(1), m_2(1), m_3(1))$. Note that since $m(\bar{A}) = \bar{A}$, it follows from the fact that each m_i is order-preserving that $m_i(0) = 0$ and $m_i(3) = 3$ for $i = 1, 2, 3$, and that for any x in $P(\bar{A})$, $\sum m_i(x_i) = \sum x_i = 3$.

Suppose that $v \neq \bar{1}$; e.g., suppose that $m_1(1) > 1$. Let w be the point $w = (1, 0, 2)$, then since $m_1(1) > 1$ it follows that $m_3(2) < 2$ since $\sum m_i(w_i) = 3$ and $m_2(0) = 0$. By considering the point $(0, 1, 2)$ in the same way, the fact that $m_3(2) < 2$ implies that $m_2(1) > 1$. So the assumption that $m_1(1) > 1$ implies that $m_2(1) > 1$ also, and the parallel argument implies that $m_3(1) > 1$ as well. This contradicts the requirement that $\sum m_i(1) = 3$, and so it follows that $\bar{1}$ is a fixed point of m .

¹¹Of course nothing in our treatment specifies that the quantity Q to be divided isn't some form of "play money" which may have different values to different players. That is, the exchange rates of play money into dollars can be different order-preserving functions for the different players.

dent solution which is symmetric and Pareto optimal on the class of monetary games.

It is worth noting that the results in this section depend to a large extent on the fact that we have been considering games with a continuum of feasible outcomes. If we considered games in which the Pareto optimal set were finite, then even the two person case would yield a variety of ordinally independent solutions. For instance the "second individually rational Pareto optimal point from the left" is an ordinally independent description in such a game, and so we could define an ordinally independent solution which selects such a point, (although it obviously would not be independent of irrelevant alternatives).

It seems likely that, in some situations with a finite number of alternatives, ordinally independent solutions of this kind may actually be descriptive. For instance, suppose that you and a friend need to decide to what restaurant you should go for dinner. If each of you has a different first choice, you might well agree on a common second choice. This is an ordinally independent proposal, since it takes no account of the intensity of each person's preferences. In disagreements of greater importance, however, it might be that considerations of intensity would be more likely to at least enter into the negotiations.

C. Interpersonal Comparisons

In the previous section we considered solutions which depend only on the preferences of the players for riskless alternatives, and not on the other information contained in their expected utility functions. Such solutions are independent of a wider range of transformations than are solutions which depend on the cardinal properties of the players' utilities. But whether we have been considering ordinal or cardinal solutions, we have thus far been concerned with transformations of any given bargaining problem obtained by transforming independently the utility functions of each player. Both Properties 1 and 9 specify that a game $G' = (S', d')$ is essentially equivalent to a game $G = (S, d)$ if and only if G' can be obtained from G by allowable transformations of each player's utility. Since the class of allowable transformations is defined in terms of functions of a single variable--each player's utility payoff at a given outcome--no information about the relative payoffs of different players at a given outcome need be preserved in the passage from G to G' . Solutions which are independent of such transformations are, therefore, insensitive to information concerning the relative payoffs of different players.

We have already argued in Section A that information about the relative payoffs (e.g., monetary payoffs) to the players may have some influence on the bargaining process. In this section we will consider some solutions which are not required to be insensitive to all information concerning relative payoffs. The first of the solutions we will consider is sensitive to a very wide range of changes in the relative payoffs to the players while the second is sensitive to relative payoffs only in certain limited respects.

Proportional Solutions¹

In Section A of Part II we considered some games with monetary prizes, and saw that a solution which is independent of equivalent utility representations

¹Solutions related to those considered here have been considered by Raiffa [1953], Isbell [1960], Kalai [1977b], Myerson [1977], and Roth [1979a]. Our presentation here follows primarily from Kalai [1977b] and Roth [1979a].

(Property 1) is incapable of distinguishing a purely formal change in the scale of a player's utility function from an actual change in the monetary awards for which he is bargaining. Now we will consider a family of solutions defined on games in which the payoffs to the players are taken to be defined in terms of a common scale.

Specifically, we will consider solutions f defined on the class B of games which we considered in Part I, which possess the following independence property.

Property 10: Independence of Common Scale Changes: For any bargaining game (S, d) and real numbers $a > 0$ and $b_i, i = 1, \dots, n$, let the game (S', d') be defined by $S' = \{y \in R^n \mid \text{for some } x \text{ in } S, y_i = ax_i + b_i \text{ for } i = 1, \dots, n\}$ and $d'_i = ad_i + b_i$ for $i = 1, \dots, n$. Then $f_i(S', d') = af_i(S, d) + b_i$ for $i = 1, \dots, n$.

Note that Property 10 is strictly weaker than Property 1; any solution which possesses Property 1 must also possess Property 10, but not conversely. Property 10 requires a solution to treat games (S, d) and (S', d') as equivalent only if the proportional gains made by each of the players at corresponding outcomes in the two games are equal. (If $y = ax + b$, where $b = (b_1, \dots, b_n)$ then for any $i, j \in N, (x_i - d_i)/(x_j - d_j) = (y_i - d'_i)/(y_j - d'_j)$ where $d' = ad + b$, as in the statement of Property 10.) Of course a Nash solution treats such games as equivalent since it possesses Property 1; i.e., it treats a strictly larger class of games as equivalent. Consideration of the equivalence class of games implied by Property 10 leads naturally to the definition of the family of proportional solutions, as follows.

A proportional solution $f: B \rightarrow R^n$ is a solution such that, for some strictly positive n -tuple $p = (p_1, \dots, p_n)$ and any bargaining game (S, d) in B ,

$$f(S, d) = \lambda(S, d)p + d,$$

where $\lambda(S, d)$ is the real number defined by

$$\lambda(S, d) \equiv \max\{t \mid (tp + d) \in S\}.$$

So if f is the proportional solution corresponding to a given n -tuple p , then for any game (S, d) and any players i and j , if $f(S, d) = x$ then $(x_i - d_i)/(x_j - d_j) = p_i/p_j$. The solution f selects the maximal feasible outcome at which the players'

gains over their disagreement payoffs are in proportion to the components of p . A given proportional solution f can always be specified by an n -tuple p normalized so that $\sum p_i = n$, in which case $f(\bar{A}, \bar{0}) = p$. A proportional solution can therefore be viewed as generalizing the results of the game $(\bar{A}, \bar{0})$ to arbitrary games, in such a way that the gains which the players receive in any game are in the same proportion as in the game $(\bar{A}, \bar{0})$.

It is straightforward to verify that a proportional solution as defined above is always individually rational and independent of irrelevant alternatives, and also that it is independent of common scale changes; i.e., it possesses Property 10. (This latter observation follows quickly from the fact that if (S', d') and (S, d) are related as in the statement of Property 10, then $\lambda(S', d') = \alpha\lambda(S, d)$.) However a proportional solution is neither Pareto optimal nor strongly individually rational over the entire class B of bargaining games. Consider, for example, the symmetric proportional solution f in the case $n = 2$, so that $f(\bar{A}, \bar{0}) = (1, 1)$. Let S be the convex hull of the points $(0, .25)$, $(1.5, .5)$, $(2, 0)$ and $(0, 0)$, and let T be the convex hull of the last three of these four points (see Figure 12, p. 81). Then $f(S, \bar{0}) = (.3, .3)$, and $f(T, \bar{0}) = (0, 0)$ neither of which is Pareto optimal (since $(1.5, .5)$ is a feasible payoff in both games).

Thus a proportional solution describes bargaining among players who are willing to sacrifice some potential gain for themselves in order to prevent any of the other players from receiving disproportionately large gains.² This kind of behavior can perhaps best be understood if we consider bargaining situations in which the final outcome will serve as the disagreement point for some future bargaining situation. Then a proportional solution describes bargaining among players who are unwilling to compromise their future bargaining positions.

Formally, we will show that the class of proportional solutions f can be characterized by the following property.

²Note that the notion of "disproportionate gains" can only be formulated in a context in which the ratios of the players' gains at each outcome convey some meaning. This is not the case if the payoffs are expressed in terms of independently scaled expected utility functions.

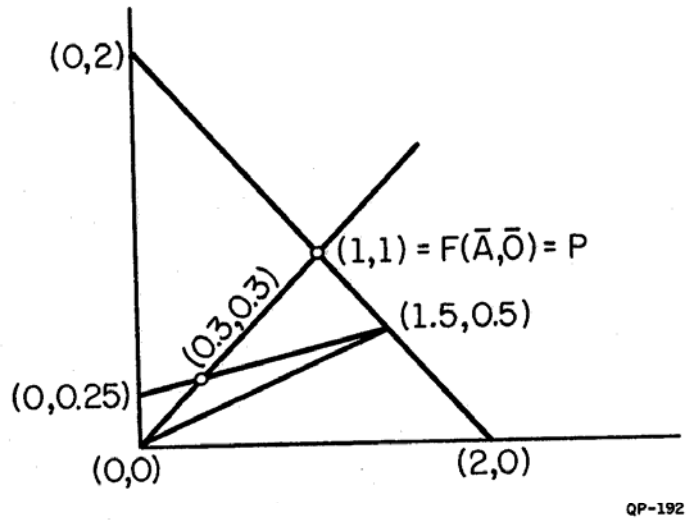


Figure 12

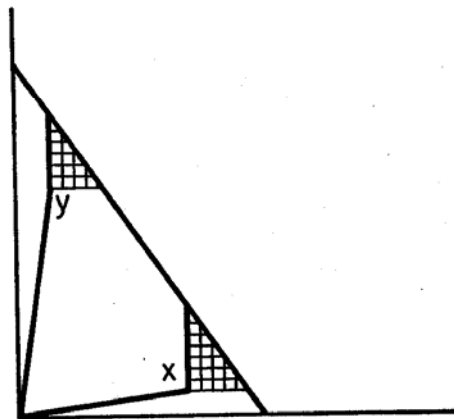


Figure 13

Property 11: Decomposibility. If (S,d) and (T,d) are bargaining games such that T contains S , then $f(T,d) = f(T,f(S,d))$, whenever $(T,f(S,d)) \in B$.

A solution which possesses Property 11 represents a bargaining process in which the set of feasible outcomes T can be decomposed into subsets without changing the final outcome. That is, if the players first negotiate over a subset S of outcomes (such that S contains d), under the usual rules of bargaining, and subsequently negotiate over the entire set T of outcomes using the agreement reached in the first stage as a disagreement point, the outcome will be the same as if the entire set T of outcomes were subject to negotiations in one stage.³

Decomposibility is closely related to the following property.

Property 12: Monotonicity. If (S,d) and (T,d) are bargaining games such that T contains S , then $f(T,d) \geq f(S,d)$.

That is, a monotonic solution models a bargaining process in which all of the bargainers benefit from any enlargement of the set of feasible outcomes. In the context of our present discussion, monotonicity is related to decomposibility through the following lemma.

Lemma 11.1: If f is a solution which is independent of common scale changes, decomposable, and individually rational, then f is monotonic.

Proof: Since f possesses Property 10, we can without loss of generality confine our attention to games whose disagreement point is equal to the origin. So let $(T,\bar{0})$ and $(S,\bar{0})$ be games such that T contains S , and for any number a let $aS = \{ax \mid x \in S\}$. Then for any a such that $0 < a < 1$, $(T,f(aS,\bar{0}))$ is a non-degenerate bargaining game; i.e., $(T,f(aS,\bar{0})) \in B$, since T contains points which are strictly greater than $f(aS,\bar{0})$.

³Kalai [1977b] suggests Kissinger's "step-by-step" disengagement accords in the Middle East as an example of a negotiating process which might be expected to obey Property 11. That is at each stage the participants evaluate any agreement in terms of the bargaining position which it will leave them in for subsequent negotiations. In Roth [1979a] Property 11 is called "path independence."

So the decomposibility of f implies that $f(T, \bar{0}) = f(T, f(aS, \bar{0}))$, while independence of common scale changes and individual rationality imply that $f(T, \bar{0}) \geq f(aS, \bar{0}) = af(S, \bar{0})$. Since the number a can be taken arbitrarily close to 1, it follows that $f(T, \bar{0}) \geq f(S, \bar{0})$, as required.

The lemma shows that decomposibility is a very powerful assumption, since it implies monotonicity, which is itself a powerful assumption. For instance, the assumption that a solution is monotonic is incompatible with the assumption that it is even weakly Pareto optimal.⁴ To see this, consider the points $x = (1.5, 0.1)$ and $y = (0.1, 1.5)$ and let A_x and A_y be the line segments joining the origin to x and to y , respectively. Then if f is any Pareto optimal solution, $f(A_x, \bar{0}) = x$ and $f(A_y, \bar{0}) = y$, while if f is monotonic then $f(\bar{A}, \bar{0}) = z$ such that $z \geq x$ and $z \geq y$. But $z \in \bar{A}$ implies $z_1 + z_2 \leq 2$, so no feasible point in \bar{A} is greater than both x and y (see Figure 13, p. 81). Consequently there is no solution which is both monotonic and Pareto optimal for all games.

However if we confine our attention to games with disposable utility, monotonicity is no longer incompatible with (weak) Pareto optimality.⁵ Specifically, define the class D of games with disposable utility to be the class of games $(S, d) \in B$ such that the set S^+ defined by $S^+ = \{x \in S \mid x \geq d\}$ is equal to the set S_D^+ defined by $S_D^+ = \{y \mid d \leq y \leq x \text{ for some } x \text{ in } S\}$. Then we can state the following result.

Lemma 11.2: If f is a decomposable solution which is strongly individually rational on the class D , then f is (weakly) Pareto optimal on the class D .

⁴Cf. Luce and Raiffa [1957], pp. 133-4.

⁵Isbell [1959] considers monotonic solutions and argues that, in practice, all games can be viewed as having disposable utility.

Proof: Let $(S, d) \in D$ and suppose the point $f(S, d) = x$ is not Pareto optimal in S . Then there is a $y \in S$ such that $y > x$. Then $(S, x) \in D$, so (letting $S = T$), decomposibility implies that $f(S, d) = f(S, x) = x$. But if f is strongly individually rational for games with disposable utility, then $f(S, x) > x$, which gives the contradiction needed to prove that x must be Pareto optimal.

Note that a proportional solution meets the conditions of the lemma, and is Pareto optimal on the class D . In fact proportional solutions can be characterized on the class D as follows.⁶

Lemma 11.3: If f is an individually rational solution which is independent of common scale changes and decomposable, and if f is strongly individually rational on the class D , then f is proportional on the class D .

That is, there is a positive n -tuple p such that if (S, d) is a game with disposable utility, then $f(S, d) = \lambda(S, d)p + d$.

Proof: Let f be a solution which obeys the conditions stated in the lemma, and let $p = f(\bar{A}, \bar{0})$. Then Lemma 11.2 implies that $\sum p_i = n$. Let ϵ be a positive number such that $\epsilon < n - p_i$ for $i = 1, \dots, n$, and let $q^i = (q_1^i, \dots, q_n^i)$ be the vector such that $q_j^i = 0$ for $j \neq i$ and $q_i^i = p_i + \epsilon$ for $i = 1, \dots, n$. For any such ϵ , let A_ϵ be the convex hull of the points p, q^1, \dots, q^n , and $\bar{0}$. Then A_ϵ is a subset of A , so the monotonicity of f (Lemma 11.1) implies $f(A_\epsilon, \bar{0}) \leq f(\bar{A}, \bar{0}) = p$. But Lemma 11.2 implies that $f(A_\epsilon, \bar{0})$ is Pareto optimal in A_ϵ , and so $f(A_\epsilon, \bar{0}) = p$. (See Figure 14a, p. 85.)

We are now in a position to show that $f(S, \bar{0}) \geq \lambda(S, \bar{0})p$ for any game $(S, \bar{0}) \in D$. Observe that, because $(S, \bar{0})$ is a game with disposable utility, it contains all points y such that $\bar{0} \leq y \leq \lambda(S, \bar{0})p$. Consequently, for every positive number $\delta < 1$, there exists a set A_ϵ of the kind described in the previous paragraph such that $\delta\lambda(S, \bar{0})A_\epsilon$ is contained in S (see Figure 14b, p. 85). So $f(S, \bar{0}) \geq f(\delta\lambda(S, \bar{0})A_\epsilon) = \delta\lambda(S, \bar{0})p$ (where the inequality is a consequence of

⁶Here we follow Kalai [1977b].

where

$$m(A,D) = \inf \{ (Ax, x) : \|x\| = 1, x \in H(A,D) \} ,$$

$$M(A,D) = \sup \{ (Ax, x) : \|x\| = 1, x \in H(A,D) \} .$$

Before recording further properties of the phase shift it will be instructive to consider an example.

Let A be a self-adjoint $n \times n$ matrix considered as an operator on \mathbb{C}^n . Let $Df = (f, y)y$ be a one dimensional operator on \mathbb{C}^n with range spanned by the vector $y \in \mathbb{C}^n$. After diagonalization of A the phase shift of the perturbation $A \rightarrow A+D$ has a fairly explicit description. In fact, assume A is diagonalized as the matrix

$$\tilde{A} = \begin{pmatrix} \alpha_{11} & & & \\ & \cdot & & 0 \\ & & \cdot & \\ 0 & & & \cdot \\ & & & & \alpha_{nn} \end{pmatrix}$$

where $\alpha_{11} \leq \alpha_{22} \leq \dots \leq \alpha_{nn}$. In this diagonalization of A the operator $D = (, y)y$ is unitarily equivalent to $\tilde{D} = (, x)x$, for some $x = (x_1, \dots, x_n) \in \mathbb{C}^n$.

For z not equal to any of the eigenvalues $(\alpha_{11}, \dots, \alpha_{nn})$, the perturbation determinant has the form

$$\Delta_{B/A}(z) = 1 + (\alpha_{11}-z)^{-1}|x_1|^2 + \dots + (\alpha_{nn}-z)^{-1}|x_n|^2 . \quad (2.8)$$

From (2.7) it is clear that the phase shift for the perturbation $A \rightarrow A+D$ equals the characteristic function of the set in \mathbb{R} where the function (2.8) is negative. The graph of (2.8) on \mathbb{R} is the following:

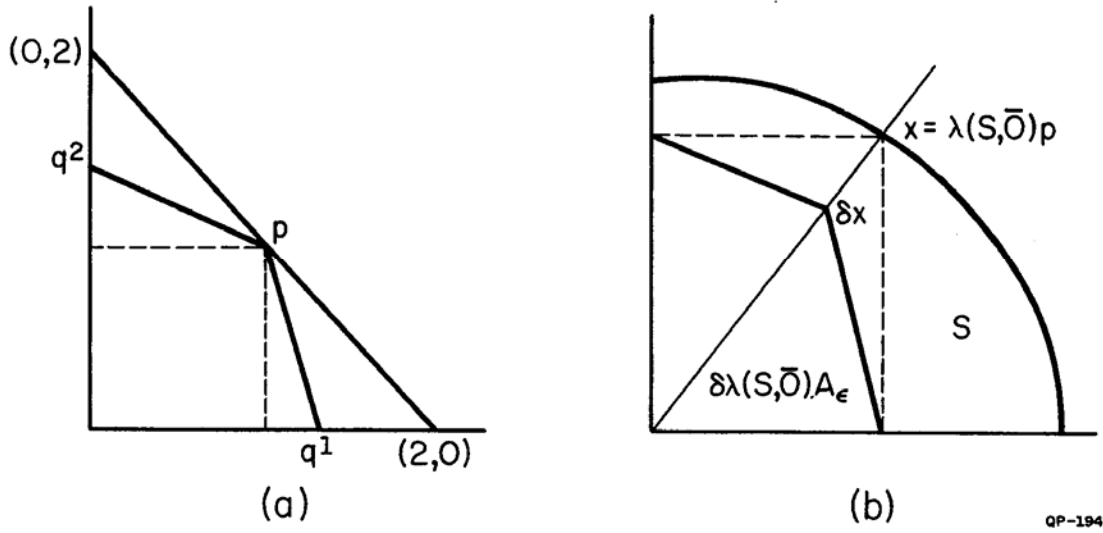


Figure 14

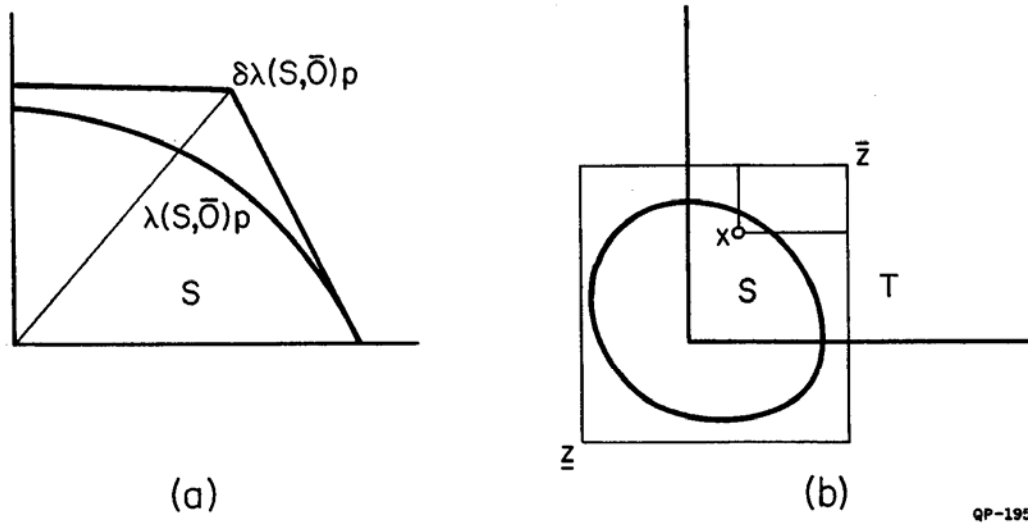


Figure 15

monotonicity, and the equality a consequence of Property 10). Since δ can be arbitrarily close to 1, it follows that $f(S, \bar{0}) \geq \lambda(S, \bar{0})p$.

To complete the proof we need to show that $f(S, \bar{0}) \leq \lambda(S, \bar{0})p$. For any $\delta > 1$ define the set

$$S_\delta = \{x \geq \bar{0} \mid x \leq y \text{ for some } y \text{ in the convex hull of } S \text{ and the point } \delta\lambda(S, \bar{0})p\}.$$

Then $(S_\delta, \bar{0})$ is the (smallest) game with disposable utility which contains the set S and the point $\delta\lambda(S, \bar{0})p$ (see Figure 15a, p. 85). Note that $\lambda(S_\delta, \bar{0}) = \delta\lambda(S, \bar{0})$, and since $\delta\lambda(S, \bar{0})p$ is strongly Pareto optimal in S_δ , $f(S_\delta, \bar{0}) = \delta\lambda(S, \bar{0})p$ follows immediately from the conclusion drawn in the previous paragraph that $f(S_\delta, \bar{0}) \geq \delta\lambda(S, \bar{0})p$. But S_δ contains S , so by monotonicity $\delta\lambda(S, \bar{0})p = f(S_\delta, \bar{0}) \geq f(S, \bar{0})$ for any $\delta > 1$, and so $f(S, \bar{0}) \leq \lambda(S, \bar{0})p$.

We have shown that $f(S, \bar{0}) = \lambda(S, \bar{0})p$; and since by assumption the solution f possesses Property 10, $f(S, d) = \lambda(S, d)p + d$ for any $(S, d) \in D$.

We can reach almost the same conclusions on the entire class B of games, as follows.⁷

Theorem 11: Let f be an individually rational solution which is strongly individually rational on the class D , and which is decomposable and independent of common scale changes. Then for any game $(S, d) \in B$, $f(S, d) = k(S, d)p + d$ where $0 \leq k(S, d) \leq \lambda(S, d)$.

Proof: For any $(S, d) \in B$, let \bar{z} and \underline{z} be n -tuples such that, for any $y \in S$, $\underline{z} < y < \bar{z}$, and let $T = \{x \mid \underline{z} \leq x \leq \bar{z}\}$. Then T contains S and $(T, x) \in D$ for any $x \in S$, so Lemma 11.3 implies $f(T, x) = \lambda(T, x)p + x$. So, by the decomposibility of f , $\lambda(T, d)p + d = f(T, d) = f(T, f(S, d)) = \lambda(T, f(S, d))p + f(S, d)$. So $f(S, d) = [\lambda(T, d) - \lambda(T, f(S, d))]p + d = k(S, d)p + d$, as required (see Figure 15b, p. 85).

⁷Here we follow Roth [1979a].

So proportional solutions are maximal solutions obeying the conditions of the theorem. That is, if g is a solution obeying the conditions of Theorem 11, then there exists a proportional solution f such that, for every $(S,d) \in B$, $f(S,d) \geq g(S,d)$, while for every $(S,d) \in D$, $f(S,d) = g(S,d)$.⁸

Proportional solutions also possess the following properties, which we introduce now in order to provide an alternative characterization of proportional solutions, and which will also be of interest in the next section.

Property 13: Individual Monotonicity. If (T,d) and (S,d) are games such that T contains S and, for some $i \in N$, T_D^+ intersected with $\{x \in R^n | x_i = d_i\}$ is equal to the intersection of S_D^+ with $\{x \in R^n | x_i = d_i\}$, then $f_i(T,d) \geq f_i(S,d)$.

This property can best be understood by observing that for every outcome x in S^+ there is a vector y in the intersection of S_D^+ with $\{x \in R^n | x_i = d_i\}$ such that $x_k = y_k$ for every $k \neq i$. So the property states that if the set S of feasible outcomes is enlarged to a set T in such a way that, for every $(n-1)$ -tuple of payoffs to players other than i the range of feasible payoffs to player i is increased, then player i 's final payoff in the enlarged game should be at least as large as his final payoff in the original game. Note that individual monotonicity is a weaker property than monotonicity; i.e., a solution which possesses Property 12 also possesses Property 13, but not conversely.⁹

⁸To see that not every solution which obeys the conditions of Theorem 11 need be proportional (i.e., to see that $k(S,d)$ need not be identical to $\lambda(S,d)$) consider the case $n = 2$. For any $(S,d) \in B$ define

$$s_1 = \max \{x_1 | (x_1, d_2) \in S\} \quad \text{and} \quad s_2 = \max \{x_2 | (d_1, x_2) \in S\},$$

and let $p = (p_1, p_2) > \bar{0}$. Let f be the solution defined by $f(S,d) = d$ if $(s_1, s_2) = (d_1, d_2)$, and $f(S,d) = \lambda(S,d)p + d$ otherwise. Then f is not a proportional solution, but it is straightforward to verify that f obeys the conditions of the theorem.

⁹In section D we will consider a solution which is individually monotonic but not monotonic. The notion of individual monotonicity was introduced in Kalai and Smorodinsky [1975] and Kalai [1977b].

Property 14: Continuity on the class D. Let $(S_j, d) \in D$ be bargaining games defined for a sequence of sets S_j such that, in the limit as j goes to infinity, S_j converges to S (in the Hausdorff topology), where (S, d) is also in the class D. Then $\lim_{j \rightarrow \infty} f(S_j, d) = f(S, d)$.

It is easy to verify that a proportional solution is individually monotonic, and continuous on the class D. (However, unlike the Nash solutions, proportional solutions are not continuous over the entire class B of games.)¹⁰ In fact, proportional solutions can be characterized by properties 3, 10, 13, and 14, as follows.

Theorem 12: If f is an individually rational solution which is strongly individually rational on the class D, and if f is independent of irrelevant alternatives, independent of common scale changes, individually monotonic, and continuous on the class D, then f is a proportional solution on the class B.

The proof will proceed in two parts. In Part 1, we will show that a solution f which meets the conditions of the theorem must be proportional on the class D, and in Part 2 we will show that it must be proportional on the entire class B.¹¹ Specifically, we will first demonstrate that for any $(S, \bar{0}) \in D$, $f(S, \bar{0})$ is of the form $\lambda(S, \bar{0})p$, and in Part 2 we will relax the assumption that $(S, \bar{0}) \in D$. Since f possesses Property 10, it is independent of the origin of each player's payoff scale, and so establishing the theorem for any game $(S, \bar{0})$ is sufficient to establish it for any (S, d) as well.

¹⁰For instance in the two-person case let f be the symmetric proportional solution, and let S_j be the line joining the origin to the point $x^j = (1, 1 + (1/j))$. Then as j goes to infinity, S_j converges to the line segment S joining the origin to $(1, 1)$, but $f(S_j, \bar{0}) = 0$ for all j , while $f(S, \bar{0}) = (1, 1)$.

¹¹Part 1 follows Kalai [1977b] while Part 2 follows Roth [1979a].

Proof of Part 1: Let f be a solution obeying the conditions of the theorem, and let $p = f(\bar{A}, \bar{0})$. Let $(S, \bar{0}) \in D$ be a game chosen so that the point $z = \lambda(S, \bar{0})p$ is strongly Pareto optimal in S , and let

$$\bar{C} = \{x \geq \bar{0} \mid \sum_{j=1}^n x_j - x_i \leq n, \text{ for } i = 1, \dots, n\}.$$

(See Figure 16a, (p. 90), and note that \bar{C} contains \bar{A} , and that the intersection of \bar{C} with $\{x \mid x_i = 0\}$ is equal to the intersection of \bar{A} with $\{x \mid x_i = 0\}$ for every i .)

Independence of common scale changes implies that $f(\lambda(S, \bar{0})\bar{A}, \bar{0}) = \lambda(S, \bar{0})p$, and so independence of irrelevant alternatives implies that $f(S', \bar{0}) = \lambda(S, \bar{0})p$, where S' is the intersection of S with $\lambda(S, \bar{0})\bar{A}$. But the set S^* equal to the intersection of S with $\lambda(S, \bar{0})\bar{C}$ contains S' , so individual monotonicity implies $f_i(S^*, \bar{0}) \geq f_i(S', \bar{0}) = \lambda(S, \bar{0})p_i$ for every $i = 1, \dots, n$. That is, $f(S^*, \bar{0}) \geq \lambda(S, \bar{0})p = z$, and so the assumption that z is strongly Pareto optimal in S (and hence in S^*) implies $f(S^*, \bar{0}) = z$.

To see that $f(S, \bar{0}) = z$, observe that if T is any subset of S which contains S^* (as S does) then independence of irrelevant alternatives implies that either $f(T, \bar{0}) = z$ or else $f(T, \bar{0}) \notin S^*$. That is, either $f(T, \bar{0}) = z$ or else $f(T, \bar{0}) \notin \lambda(S, \bar{0})\bar{C}$. But the point z is contained in the interior of the set $\lambda(S, \bar{0})\bar{C}$ (since $p > \bar{0}$ is contained in the interior of \bar{C}), and consequently the continuity of f insures that $f(S, \bar{0}) = z$.¹² (See Figure 16b, p. 90.)

This almost completes the proof of Part 1, except that we began with the assumption that the game $(S, \bar{0}) \in D$ was chosen so that z is strongly Pareto optimal in S . To see that we can relax this assumption, first note that for any game

¹²Let $\{T_a\}$ be a family of sets defined for $a \in [0, 1]$ such that $T_0 = S^*$, $T_1 = S$, and $a < b$ implies T_a is contained in T_b . In addition, let the family $\{T_a\}$ be continuous: i.e., if $\{a_i\}$ is a sequence of numbers converging to a , then T_{a_i} converges to T_a . Then continuity implies that $f(T_a, \bar{0})$ traces a continuous path from $f(T_0, \bar{0})$ to $f(T_1, \bar{0})$. But since, for every a , $f(T_a, \bar{0})$ is either equal to z and thus in the interior of $\lambda(S, \bar{0})\bar{C}$ or else $f(T_a, \bar{0}) \notin \lambda(S, \bar{0})\bar{C}$, it follows that $f(T_a, \bar{0}) = z$ for all $a \in [0, 1]$, since no continuous path can "jump" from the interior of a set to its exterior, without crossing its boundary.

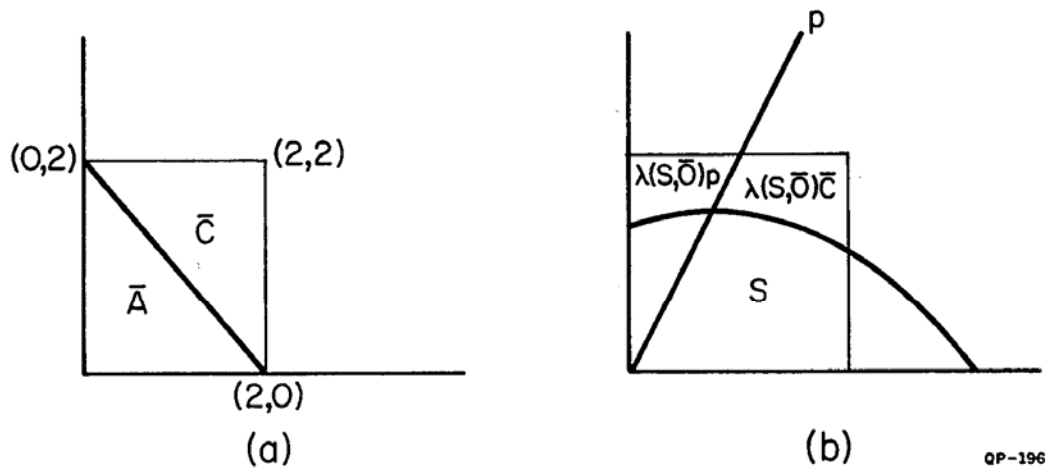


Figure 16

$(S, \bar{0}) \in D$ the point $z = \lambda(S, \bar{0})_p$ must be at least weakly Pareto optimal. (If not, there would exist some point y in S with $y > z + \epsilon p$ for some $\epsilon > 0$. Since $(S, \bar{0}) \in D$, $z + \epsilon p$ must be contained in S , which contradicts the definition of $\lambda(S, \bar{0})$.) Since z is weakly Pareto optimal in S , we can construct a sequence of games $(S_j, \bar{0}) \in D$ in which z is strongly Pareto optimal such that the sets S_j converge to S . The continuity of f then implies that $f(S, \bar{0}) = \lim_{j \rightarrow \infty} f(S_j, \bar{0})$, which our previous argument has shown is the point z , as required.

Proof of Part 2: For any game $(S, \bar{0}) \in B$, define $S_D^+ = \{y \geq \bar{0} \mid y \leq x \text{ for some } x \text{ in } S\}$. Then $(S_D^+, \bar{0}) \in D$, and so we have just shown that $f(S_D^+, \bar{0}) = \lambda(S_D^+, \bar{0})_p$. We need to show that $f(S, \bar{0}) = \lambda(S, \bar{0})_p$.

Let $z = \lambda(S, \bar{0})_p$, and observe that if z is Pareto optimal in S then it is also Pareto optimal in S_D^+ , so $\lambda(S_D^+, \bar{0}) = \lambda(S, \bar{0})$. Thus in this case $f(S_D^+, \bar{0}) = z$, and independence of irrelevant alternatives implies that $f(S, \bar{0}) = z$ as well.

Next observe that for any game $(S, \bar{0}) \in B$, it must be the case that $f(S, \bar{0}) \geq z$. If $z = \bar{0}$, this follows immediately from the individual rationality of f . Otherwise $z > 0$. For any i in N , define $S^i = \{x \in S \mid x_i \leq z_i\}$. Then $\lambda(S^i, \bar{0}) = \lambda(S, \bar{0})$, and z is Pareto optimal in S^i , so the argument of the previous paragraph insures that $f(S^i, \bar{0}) = z$. Independence of irrelevant alternatives now implies that either $f(S, \bar{0}) = z$ or else $f(S, \bar{0}) \notin S^i$. In either case, $f_i(S, \bar{0}) \geq z_i$, and since i was chosen arbitrarily, $f(S, \bar{0}) \geq z$.

To show that $f(S, \bar{0}) = z$, let $y = f(S, \bar{0})$ and let $T = \{x \in S \mid x \leq y\}$. Then if $(T, \bar{0}) \in B$ (i.e., if $y > \bar{0}$) then $f(T, \bar{0}) = y$ by independence of irrelevant alternatives. So $f(T_D, \bar{0}) \geq y$ by individual monotonicity, which implies $y = f(T_D, \bar{0}) = \lambda(T_D, \bar{0})_p$, since y is strongly Pareto optimal in T_D . But in the previous paragraph we showed $y \geq z$, so $z = \lambda(S, \bar{0})_p$ is an element of T , which implies $\lambda(S, \bar{0}) \leq \lambda(T_D, \bar{0})$. Similarly, $y = \lambda(T_D, \bar{0})_p$ is an element of S , so $\lambda(T_D, \bar{0}) \leq \lambda(S, \bar{0})$. Thus $\lambda(S, \bar{0}) = \lambda(T_D, \bar{0})$, so $f(S, \bar{0}) = y = z$.

To complete the proof, we need to show that the same conclusion holds when $(T, \bar{0}) \notin B$, i.e., when $y_k = 0$ for some k . Since we have shown that $y \geq z$, this can only occur when $\lambda(S, \bar{0}) = 0$, so $z = \bar{0}$. Let \hat{y} be the positive vector defined by $\hat{y}_i = y_i$ if $y_i > 0$, and $\hat{y}_i = \epsilon$ if $y_i = 0$, where ϵ is an arbitrarily small positive

number. Letting $\hat{T} = \{x \in S \mid x \leq \hat{y}\}$, independence of irrelevant alternatives implies $f(\hat{T}, \bar{0}) = y$, and individual monotonicity implies $\lambda(\hat{T}_D, \bar{0})_p = f(\hat{T}_D, \bar{0}) \geq y$. Since $\lambda(\hat{T}_D, \bar{0})$ can be made arbitrarily small by choosing $\varepsilon = \hat{y}_k$ small, it follows that $y = z = \bar{0}$, which completes the proof.

Ordinal Interpersonal Comparisons¹³

The proportional solutions which we have been considering involve cardinal comparisons of different players' payoffs, which are meaningful only in a context in which there is a great deal of information about the relative payoffs of the various players at every outcome. Now we will consider interpersonal comparisons in the context of solutions which are sensitive only to ordinal information. The relationship our treatment of this subject bears to our treatment of proportional solutions is approximately the same as the relationship between our treatments of ordinally and cardinally independent solutions involving no interpersonal comparisons. We will consider two-player games in the class B^* which we have already studied in an ordinal context without interpersonal comparisons, and we will show that allowing ordinal interpersonal comparisons will add sufficient structure to the problem to permit us to derive a non-trivial solution.

Formally, we will consider two-player games $(S, d) \in B^*$. Recall that S is a compact subset of the plane whose strong Pareto set $\bar{P}(S)$, which by definition is the set $\{x \in S \mid y \geq x \text{ and } y \in S \text{ implies } y = x\}$, can also be described as the set $\bar{P}(S) = \{(x_1, \phi(x_1)) \mid \underline{x}_1 \leq x_1 \leq \bar{x}_1\}$ where ϕ is a real valued, continuous, strictly decreasing function defined on the interval $[\underline{x}_1, \bar{x}_1]$.

If the set S is the set of feasible utility payoffs to the two players, then it conveys (at least) ordinal information about each player's preferences for outcomes. We will be interested in solutions which may be sensitive to ordinal interpersonal comparisons as well, i.e., we will be looking at solutions which obey the following independence property.

¹³The material presented under this heading follows Roth [1979c], and Roth and Malouf [1979].

Property 15: Independence of Ordinal Transformations Preserving Interpersonal Comparisons. Let (S, d) and (S', d') be two-player games in B^* such that $d'_i = t_i(d_1, d_2)$ for $i = 1, 2$ and $S' = \{(t_1(x_1, x_2), t_2(x_1, x_2)) \mid (x_1, x_2) \in S\}$ where $t = (t_1, t_2)$ is a transformation $t: S \rightarrow \mathbb{R}^2$ such that for all x, y in S and $i = 1, 2$,

- (i) $t_i(x_1, x_2) \geq t_i(y_1, y_2)$ if and only if $x_i \geq y_i$ and
- (ii) $t_1(x_1, x_2) - t_1(d_1, d_2) \geq t_2(x_1, x_2) - t_2(d_1, d_2)$ if and only if $x_1 - d_1 \geq x_2 - d_2$.

Then $f_1(S', d') = t_1(f(S, d))$.

So $G' = (S', d')$ and $G = (S, d)$ are equivalent from the point of view of this property if G' is derived from G via a transformation t (of feasible payoff vectors in G to feasible payoff vectors in G')¹⁴ which (i) preserves each player's ordinal preferences, and (ii) preserves information about which player makes larger gains at any given outcome. A solution which possesses this property can thus be sensitive not only to the players' preferences over outcomes, but also to their feelings about which one of them benefits more at any given outcome. Note that a transformation t which preserves interpersonal comparisons divides the plane into the three regions $\{x \mid x_1 - d_1 = x_2 - d_2\}$, $\{x \mid x_1 - d_1 > x_2 - d_2\}$, and $\{x \mid x_1 - d_1 < x_2 - d_2\}$, and it transforms a point from any of these regions into a point in the same region.

For two-player games in the class B^* we can state the following result.

Theorem 13: There is a unique solution f which is strongly individually rational, strongly Pareto optimal, independent of irrelevant alternatives, and independent of ordinal transformations preserving interpersonal comparisons. It is the solution $f = E$ defined by $E(S, d) = x$ such that $\min \{x_1 - d_1, x_2 - d_2\} > \min \{y_1 - d_1, y_2 - d_2\}$ for all $y \in \bar{P}(S)$ such that $y \neq x$.

¹⁴The transformation is defined on payoff vectors, and part (ii) of Property 15 requires that in fact the transformations of one player's payoffs must be related to the transformation of the other's. Note that Property 15 is strictly stronger than Property 10; i.e., any solution which is independent of ordinal transformations preserving interpersonal comparisons is also independent of common scale changes, but not conversely. On the other hand Property 15 is strictly weaker than Property 9, independence of ordinal transformations.

The solution E picks the strongly Pareto optimal point which maximizes the minimum gains available to the players.¹⁵ The letter E was chosen to reflect the fact that this solution selects the outcome which gives both players equal gains, whenever there exists a strongly Pareto optimal outcome with this property. That is, whenever there is a point x in $\bar{P}(S)$ such that $x_1 - d_1 = x_2 - d_2$, then $E(S, d) = x$. If no point in $\bar{P}(S)$ gives the players equal gains, then it must be that $\bar{P}(S)$ is contained either in the set $\{x | x_1 - d_1 > x_2 - d_2\}$ or in the set $\{x | x_1 - d_1 < x_2 - d_2\}$.¹⁶ In the first case $E(S, d) = x$ is the point at which x_2 is maximized, in the second case it is the point which maximizes x_1 . That is, the solution E always chooses the strongly Pareto optimal point which comes closest to giving the players equal gains.

Proof: It is straightforward to verify that the function E defined in the statement of the theorem is well-defined and possesses the properties specified by the theorem. We need to show that it is the unique solution with those properties; i.e., that if f is a solution possessing the specified properties, then $f(S, d) = E(S, d)$ for any (S, d) in B^* .

First observe that if f is individually rational, strongly Pareto optimal, and independent of irrelevant alternatives, then for any $(S, d) \in B^*$, $f(S, d) = f(T, d)$ where T is the union of $\bar{P}(S^+)$ and $\{d\}$. (This follows since individual rationality implies that $f(S, d) \in S^+ = \{x \in S | x \geq d\}$, strong Pareto optimality implies $f(S, d) \in \bar{P}(S^+)$, and so independence of irrelevant alternatives implies that $f(S, d) = f(T, d)$ for any subset T of S which contains $\bar{P}(S^+)$.) We will sometimes denote the game (T, d) by $(\bar{P}(S^+), d)$, which is a slight abuse of our

¹⁵ There is a unique such point which is strongly Pareto optimal. Of course there may be several points in a set S which maximize the minimum gains available, but all but the point $E(S, d)$ must be only weakly Pareto optimal. The function E can be described as a lexicographic maximin function, since it selects the point $(x_1, x_2) \in S$ such that $\max \{x_1 - d_1, x_2 - d_2\} \geq \max \{y_1 - d_1, y_2 - d_2\}$ where the maxima are taken over all points $z \in S$ such that $\min \{z_1 - d_1, z_2 - d_2\} \geq \min \{w_1 - d_1, w_2 - d_2\}$ for all $w \in S$.

¹⁶ This follows from the Mean Value Theorem, since $\bar{P}(S)$ is defined by a continuous, decreasing function ϕ .

notation, since $d \notin \bar{P}(S^+)$. However it will be understood that in this case the set of feasible payoff vectors is the union of $\bar{P}(S^+)$ and $\{d\}$. So it will be sufficient to show that f and E coincide on games of the form $(\bar{P}(S^+), d)$, and Property 15 further insures that it will be sufficient to show this when $d = \bar{0}$.

First we show that $f(\bar{P}(\bar{A}), \bar{0}) = E(\bar{P}(\bar{A}), \bar{0}) = (1,1)$, (where $\bar{P}(\bar{A})$ is of course the line segment joining the points $(2,0)$ and $(0,2)$). To show this it is sufficient to note that the union of $\bar{P}(\bar{A})$ and $\{\bar{0}\}$ can be mapped into itself by the transformation $t = (t_1, t_2)$ given by

$$t_1(x_1, x_2) = \begin{cases} 2x_1 / (1 + x_1) & \text{for } 0 \leq x_1 \leq x_2 \\ x_1 + x_2 - t_2(x_1, x_2) & \text{for } 0 \leq x_2 < x_1 \end{cases}$$

$$t_2(x_1, x_2) = \begin{cases} x_1 + x_2 - t_1(x_1, x_2) & \text{for } 0 \leq x_1 \leq x_2 \\ 2x_2 / (1 + x_2) & \text{for } 0 \leq x_2 < x_1 \end{cases}$$

This transformation t defined on the union of $\bar{P}(\bar{A})$ and $\{\bar{0}\}$ satisfies the conditions of Property 15 and leaves only the points $\bar{0}$, $(2,0)$, $(1,1)$, and $(0,2)$ fixed.¹⁷

But since t transforms the game $(\bar{P}(\bar{A}), \bar{0})$ into itself, Property 15 requires that $f(\bar{P}(\bar{A}), \bar{0}) = t(f(\bar{P}(\bar{A}), \bar{0}))$; i.e., $f(\bar{P}(\bar{A}), \bar{0})$ must be a fixed point of t . The unique fixed point of t in $\bar{P}(\bar{A})$ which is strongly individual rational is the point $(1,1)$, and so $f(\bar{P}(\bar{A}), \bar{0}) = (1,1)$, as required.

Next, we will demonstrate that it will be sufficient for our proof to show that f and E coincide on games $(\bar{P}(S^+), \bar{0})$ such that $\bar{P}(S)$ is a subset of $\bar{P}(\bar{A})$. To see this, consider an arbitrary game of the form $(\bar{P}(S^+), \bar{0})$, and let $(\bar{P}(T), \bar{0})$ be a game such that $\bar{P}(T)$ contains $\bar{P}(S)$, and $\bar{P}(T) = \{(x_1, \phi(x_1)) \mid 0 \leq x_1 \leq \bar{x}_1\}$ where ϕ is a continuous decreasing function such that $\phi(\bar{x}_1) = 0$. (Thus $\bar{P}(T)$ touches both axes; i.e., it contains points of the form $(0, \bar{x}_2)$ and $(\bar{x}_1, 0)$.) Then there is a (unique) point x^* in $\bar{P}(T)$ which gives the players equal gains;¹⁸ let $x^* = (c, c)$, where c is a positive real number. The transformation $t = (t_1, t_2)$ given by

¹⁷ The transformation t moves each non-fixed point of $\bar{P}(\bar{A})$ closer to the point $(1,1)$.

¹⁸ Again, by the Mean Value Theorem.

$$t_1(x_1, x_2) = \begin{cases} x_1 & \text{for } 0 \leq x_1 \leq c \\ 2c - x_2 & \text{otherwise} \end{cases}$$

$$t_2(x_1, x_2) = \begin{cases} 2c - x_1 & \text{for } 0 \leq x_1 \leq c \\ x_2 & \text{otherwise} \end{cases}$$

transforms the game $(\bar{P}(T), \bar{0})$ into the game $(c\bar{P}(\bar{A}), \bar{0})$ (see Figure 17, p. 97). So the transformation $t' = t/c$ transforms $(\bar{P}(T), \bar{0})$ into $(\bar{P}(\bar{A}), \bar{0})$, and thus transforms the arbitrary game $(\bar{P}(S^+), \bar{0})$ into a game whose strong Pareto set is a subset of $\bar{P}(\bar{A})$. Property 15 thus assures that if f and E coincide on subsets of $\bar{P}(\bar{A})$ then they coincide everywhere.

But if $\bar{P}(S)$ is a subset of $\bar{P}(\bar{A})$ which contains the point $f(\bar{P}(\bar{A}), \bar{0}) = (1,1)$, then independence of irrelevant alternatives implies $f(\bar{P}(S), \bar{0}) = (1,1)$ as well. If $(1,1)$ is not an element of $\bar{P}(S)$ then $\bar{P}(S)$ is contained either in a line segment joining $(0,2)$ to $x = (\bar{x}_1, \bar{x}_2)$ such that x maximizes player 1's payoff in $\bar{P}(S)$ and $\bar{x}_1 < \bar{x}_2$, or else in a line segment joining $(2,0)$ to the point $x = (\bar{x}_1, \bar{x}_2)$ which maximizes player 2's payoff, and for which $\bar{x}_2 < \bar{x}_1$.

But either line segment can be transformed into itself leaving only its endpoints fixed, so that Property 15 together with independence of irrelevant alternatives and strict individual rationality implies that $f(\bar{P}(S), \bar{0}) = x = E(\bar{P}(S), \bar{0})$, which completes the proof.

Note that the solution E can be experimentally distinguished from the symmetric proportional solution only on games which have no strongly Pareto optimal point which gives the players equal gains. Much of the experimentation in the psychological literature has concentrated on games in which the payoffs are in money, and there is some evidence¹⁹ that in such games some aspects of the player's behavior can be explained in terms of the solution E .

¹⁹ See for example the work of Rapoport and his colleagues, and Roth and Malouf [1979].

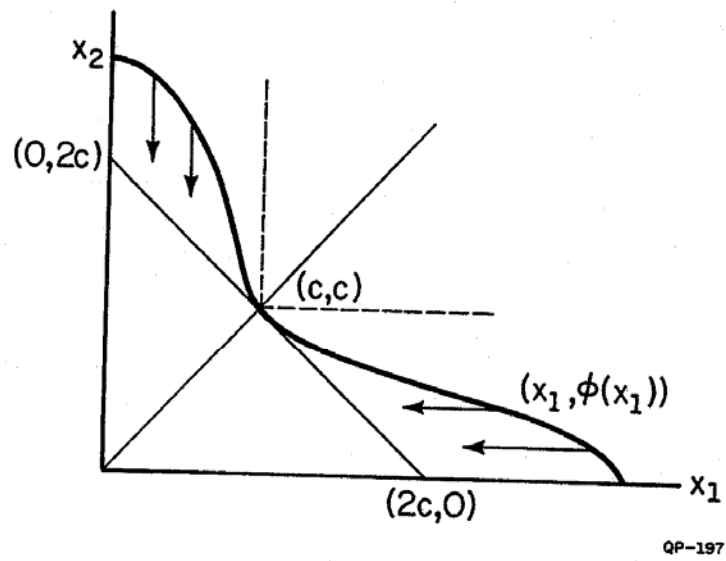


Figure 17

D. "Irrelevant" Alternatives

In the previous section we explored some solutions which, like the (symmetric and non-symmetric) Nash solutions, are independent of irrelevant alternatives but which, unlike the Nash solutions, are not independent of equivalent utility representations. In this section we will consider some solutions which are independent of equivalent utility representations, but which are not independent of "irrelevant" alternatives. That is, we will consider some solutions which are sensitive to changes in the set of feasible outcomes in a way which the Nash solutions are not.

When we first discussed the two independence properties, we suggested that independence of equivalent utility representations can be thought of as specifying the information on which bargaining might depend, while independence of irrelevant alternatives specifies the kind of bargaining process being modelled. So in this section we will be considering bargaining which may proceed in a different way than the bargaining modelled by the Nash solution, but which will have available the same information—i.e., will have available information about each player's preferences just sufficient to determine his expected utility function up to an arbitrary choice of origin and scale. We will again be considering bargaining games in the class B.

The word "irrelevant" in the name of Property 3 obviously prejudices the issue. A more descriptive name would be "independence of alternatives other than the disagreement point," since the property essentially specifies that a solution possessing it may not depend on any point in the feasible set except possibly the disagreement point.¹ In this section we will consider solutions which may depend on other points in the feasible set as well.

An Individually Monotonic Solution

The solution considered here has received considerable attention in the recent literature, and together with some closely related solutions, has been studied from several different points of view by a number of investigators interested in

¹Cf. Roth [1977d].

two-person bargaining games. (See e.g., Butrim [1976], Crott [1971], Kalai and Smorodinsky [1975], Nydegger and Owen [1975], Raiffa [1953], and Schmitz [1977], and cf. Footnote 3.) This solution is not independent of "irrelevant" alternatives, in the sense that it depends not only on the disagreement payoff but also on the maximum payoffs each player might hope to receive. Specifically, for any two-player game (S, d) let $\bar{x}(S) = (\bar{x}_1, \bar{x}_2)$ be the vector defined by $\bar{x}_i = \max \{x_i \mid (x_1, x_2) \in S^+\}$ for $i = 1, 2$.² Let G be the solution such that $G(S, d)$ is the maximal point on the line joining d to $\bar{x}(S)$ which is contained in S . That is, $G(S, d) = x$ such that $x \in S$, $(x_1 - d_1)/(x_2 - d_2) = (\bar{x}_1 - d_1)/(\bar{x}_2 - d_2)$, and $x \geq y$ for all $y \in S$ such that $(y_1 - d_1)/(y_2 - d_2) = (\bar{x}_1 - d_1)/(\bar{x}_2 - d_2)$. (See Figure 18, p. 100.)³

Kalai and Smorodinsky [1975] provide the following elegant characterization of the solution G in terms of Properties 1, 2, 7, and 13.

Theorem 14: The solution G is the unique solution defined on two player games in the class B which is strongly Pareto optimal, symmetric, independent of equivalent utility representations, and individually monotonic.

²When no confusion will result we will sometimes suppress the dependence on S and write $\bar{x}_i(S)$ as simply \bar{x}_i , just as we suppress the dependence on d when we write $S^+ = \{x \in S \mid x > d\}$.

³Crott [1971] independently proposed this solution for games whose disagreement point is the origin to model the results of an experimental investigation. Kalai and Smorodinsky [1975] observed that the solution is individually monotonic, and provided the axiomatic characterization which we will present as Theorem 14. Nydegger and Owen [1975] conducted some experiments motivated by the results of Kalai and Smorodinsky, and unlike Crott, obtained results for which the solution G did not provide an adequate model. (In fact their results were instead more consistent with either the symmetric proportional solution or the "equal gains" solution E considered in the previous section.) Butrim [1976] also considers the solution G from an axiomatic point of view.

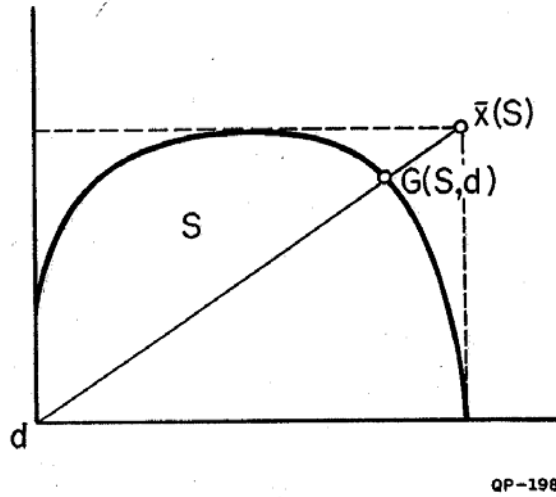


Figure 18

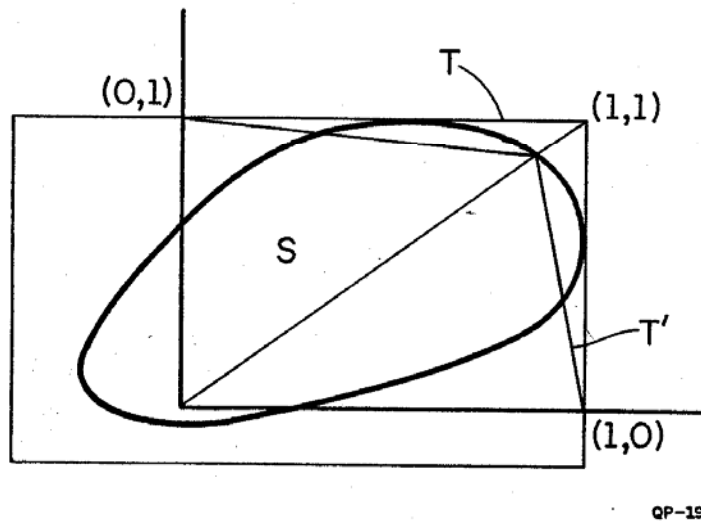


Figure 19

Proof: It is relatively straightforward to verify that the solution G for two-player games possesses the properties named in the theorem. To see that it is the only solution which does, let f be any solution possessing the properties named in the theorem: we will show that $f(S,d) = G(S,d)$ for any game (S,d) . Since f and G are both independent of equivalent utility representations, it will be sufficient to show that they coincide on all games of the form $(S,\bar{0})$, such that $\bar{x}(S) = (1,1)$.⁴

So let $(S,\bar{0})$ be a two-player game with $\bar{x}(S) = (1,1)$, and let $(T,\bar{0})$ be a game such that T contains S and the set $T^+ = \{x \in T \mid x \geq \bar{0}\}$ is equal to the set $S_D^+ = \{y \mid \bar{0} \leq y \leq x \text{ for some } x \text{ in } S\}$. (See Figure 19, p. 100.) Then $\bar{x}(T) = (1,1)$ and the property of individual monotonicity implies that $f_i(T,\bar{0}) \geq f_i(S,\bar{0})$ for $i = 1,2$, so that $f(T,\bar{0}) \geq f(S,\bar{0})$.⁵ But any point which is strongly Pareto optimal in S is also strongly Pareto optimal in T , so that $f(S,\bar{0})$ is strongly Pareto optimal in T . Consequently $f(T,\bar{0}) = f(S,\bar{0})$.

Let T' be the convex hull of the points $\bar{0}$, $(0,1)$, $(1,0)$ and $G(T,\bar{0})$, all of which are contained in T . Then T contains T' , which is a symmetric set, since $G_1(T,\bar{0}) = G_2(T,\bar{0})$. Since the solution f is symmetric as well as Pareto optimal, it follows that $f(T',\bar{0}) = G(T,\bar{0})$. But T' is contained in T and $\bar{x}(T') = \bar{x}(T) = (1,1)$, so individual monotonicity implies (cf. Footnote 5) that $f(T,\bar{0}) \geq f(T',\bar{0}) = G(T,\bar{0})$, and since $G(T,\bar{0})$ is strongly Pareto optimal in T , it follows that $f(T,\bar{0}) = G(T,\bar{0})$. But $G(T,\bar{0}) = G(S,\bar{0})$, so $f(S,\bar{0}) = f(T,\bar{0}) = G(S,\bar{0})$, as required.

⁴That is, we can choose the utility representations so that any game is equivalent to one which has been normalized as indicated.

⁵It is straightforward to verify that if f is a solution which possesses Property 13, individual monotonicity, then it also possesses the following weaker property.

Property 13': Restricted monotonicity. If (T,d) and (S,d) are games such that T contains S and $\bar{x}(T) = \bar{x}(S)$, then $f(T,d) \geq f(S,d)$.

This property is the only consequence of individual monotonicity needed in the proof. (Cf. Kalai and Smorodinsky [1975], Schmitz [1979], Roth [1979d].)

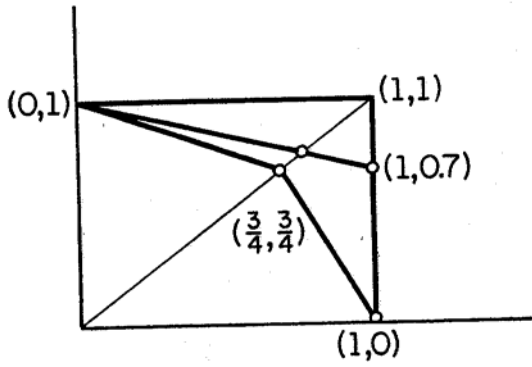
We can best compare the solution G with the Nash solution F by observing that in Theorems 1 and 14 we have characterized them with essentially the same list of properties, except that F is independent of irrelevant alternatives while G is individually monotonic. Since F and G are not equal, an immediate corollary of the two theorems is that G is not independent of irrelevant alternatives (a fact which is apparent from its definition) while F is not individually monotonic. Of course F and G coincide on all symmetric games, and on all games which can be converted into symmetric games by equivalent utility representations. Thus for instance $F(\bar{A}, \bar{0}) = G(\bar{A}, \bar{0}) = (1,1)$, but the two solutions can be thought of as generalizing this outcome to other games in different ways.⁶

For example, let S be the convex hull of the points $(2,0)$, $(1,1)$, $(0,1)$ and $\bar{0}$. Then S is contained in \bar{A} , and $F(\bar{A}, \bar{0}) = (1,1)$ is an element of S , so the fact that F is independent of irrelevant alternatives implies that $F(S, \bar{0}) = (1,1)$ also (cf. Figure 9, p. 65). However $\bar{x}(S) = (2,1)$, so $G(S, \bar{0}) = (4/3, 2/3)$. Thus G models a bargaining process in which a player's final payoff decreases as his maximum feasible payoff decreases.

To see that F is not individually monotonic, consider the sets S and T where S is the convex hull of the points $\bar{0}$, $(0,1)$, $(1,0)$, and $(3/4, 3/4)$, and T is the convex hull of the points $\bar{0}$, $(0,1)$, $(1,0)$, and $(1, .7)$.⁷ Then S is symmetric, so $F(S, \bar{0}) = G(S, \bar{0}) = (3/4, 3/4)$. Furthermore, T contains S and $\bar{x}(T) = \bar{x}(S)$, so individual monotonicity implies that $G(T, \bar{0}) \geq G(S, \bar{0})$, and in fact $G(T, \bar{0}) = (10/13, 10/13)$, (see Figure 20, p. 103). But $F(T, \bar{0}) = (1, .7)$, so that, unlike the solution G , the Nash solution models a bargaining process in which player 2 does less well in the game $(T, \bar{0})$ than in the game $(S, \bar{0})$, even though the set T offers him a wider range of payoffs for any given payoff to player 1 (cf. the discussion of Property 13).

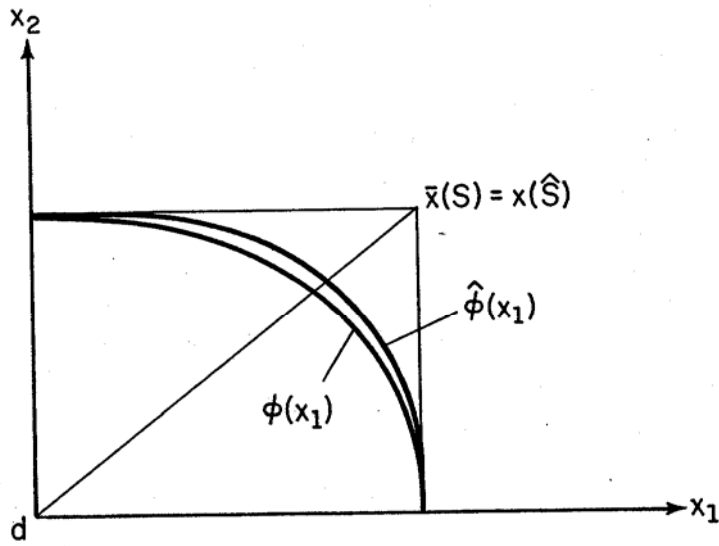
⁶Specifically, independence of irrelevant alternatives permits the solution of one game to be deduced from another game with a larger set of feasible outcomes, while individual monotonicity permits the solution of a game to be deduced from a game with a smaller set of feasible outcomes (cf. the proof of Theorem 14).

⁷This example was used by Kalai and Smorodinsky [1975] to motivate the solution G by contrast with the Nash solution.



QP-200

Figure 20



QP-201

Figure 21

We note in passing that the solution G shares with F the property of risk sensitivity which we studied in Section D of Part I. That is, in two-player bargaining games, a player's final payoff as modelled by G increases as his opponent becomes more risk averse. Specifically, we have the following result, which parallels Theorem 5 (cf. Kihlstrom, Roth, and Schmeidler [1979]).

Theorem 15: The utility which the solution G assigns to a player in a two-person game increases as his opponent becomes more risk averse. That is, for $i = 1$ or 2 , $G_i(\hat{S}, \hat{d}) \geq G_i(S, d)$, where (\hat{S}, \hat{d}) is obtained from (S, d) by replacing player j ($j \neq i$) with a more risk averse player.

Proof: Let (\hat{S}, \hat{d}) be derived from (S, d) by replacing the utility function w of player 2, say, with a more risk averse utility function \hat{w} , such that $\hat{w}(c) = k(w(c))$ for all c in the underlying set of alternatives C . (Recall from Theorem 4 that k is an increasing concave function.) Since the solution G is independent of equivalent utility representations, we can choose any normalization for \hat{w} , and hence for k . So let $k(d_2) = d_2$ and $k(\bar{x}_2) = \bar{x}_2$. This is equivalent to letting $\hat{w}(\bar{c}) = w(\bar{c}) = d_2$ and $\hat{w}(m_2) = w(m_2) = \bar{x}_2$, where \bar{c} is the disagreement alternative and m_2 is the alternative in C which yields player 2 his maximum utility in the Pareto set of the set $S^+ = \{x \in S \mid x \geq d\}$. Similarly, let m_1 be the alternative which yields player 1 his maximum payoff in $P(S^+)$, so $u_1(m_1) = \bar{x}_1$.

Note that $w(m_2) \geq w(m_1) \geq w(\bar{c})$, and so there exists some number α between 0 and 1 such that $w(m_1) = \alpha w(\bar{c}) + (1-\alpha)w(m_2)$. Consequently the concavity of k implies that $\hat{w}(m_1) \geq w(m_1)$, since

$$\hat{w}(m_1) = k(w(m_1)) \geq \alpha k(w(\bar{c})) + (1-\alpha)k(w(m_2)) = w(m_1),$$

where the last equality follows from the fact that $k(w(\bar{c})) = w(\bar{c})$ and $k(w(m_2)) = w(m_2)$.

So if we denote the Pareto sets $P(S)$ and $P(\hat{S})$ as sets consisting of points of the form $(x_1, \phi(x_1))$ and $(x_1, \hat{\phi}(x_1))$, respectively, then $\hat{\phi}(\bar{x}_1) \geq \phi(\bar{x}_1)$, and $\hat{\phi}(u_1(m_2)) = \phi(u_1(m_2))$. That is $\hat{\phi}$ is greater than ϕ at either end point of the Pareto set $P(S^+)$, and since $\hat{\phi} = k(\phi)$ and k is concave, this implies that

$\hat{\phi}(x_1) > \phi(x_1)$ for all $x_1 \in [u_1(m_2), \bar{x}_1]$. That is, every point in $P(S^+)$ is less than or equal to the corresponding point in $P(\hat{S}^+)$; (see Figure 21, p. 103). Since $\bar{x}(S) = \bar{x}(\hat{S})$ and $d = \hat{d}$, it therefore follows that $G(\hat{S}, \hat{d}) = G(\hat{S}^+, \hat{d}) \geq G(S^+, d) = G(S, d)$, and so $G_1(\hat{S}, \hat{d}) \geq G_1(S, d)$, as required.

Although the literature concerning the solution G has been concerned exclusively with two-player games, G is nevertheless well-defined for games with any number of players.⁸ However G need not be Pareto optimal for games involving more than two players. In fact, for games with more than two players, no solution exists which possesses the properties used to characterize G in Theorem 14. In fact, we have the following result (Roth [1979d]).

Theorem 16: For bargaining games with 3 or more players, no solution exists which is Pareto optimal, symmetric, and individually monotonic.

Specifically, we will prove that no solution for n person games ($n \geq 3$) possesses the properties of Pareto optimality, symmetry, and restricted monotonicity (cf. footnote 5).

Proof: We will assume that f is a solution which possesses the properties named in the theorem, and show that this leads to a contradiction. Consider the n -person game ($n \geq 3$) whose disagreement point is equal to the origin (which we will denote by $\bar{0}$), and whose feasible set S is equal to the convex hull of $\bar{0}$ and the points p and q such that $p_1 = 0$ and $p_i = 1$ for $i \neq 1$, and $q_2 = 0$ and $q_j = 1$ for $j \neq 2$ (see Figure 22, p. 106). Then the set of Pareto optimal points in S is the line segment joining p to q , so $f(S, \bar{0}) = x$ is a convex combination of p and q . In particular, $x_3 = 1$.

Now consider the game $(T, \bar{0})$, where $T = \{x \geq \bar{0} \mid \sum x_i \leq (n-1) \text{ and } x_i \leq 1 \text{ for } i = 1, \dots, n\}$. Then $(T, \bar{0})$ is a symmetric game, so $f_1(T, \bar{0}) = f_2(T, \bar{0}) = \dots = f_n(T, \bar{0})$, and the Pareto optimality of f implies that $f(T, \bar{0}) = z = (\frac{n-1}{n}, \frac{n-1}{n}, \dots, \frac{n-1}{n})$.

⁸For an n -player game (S, d) , $G(S, d) = x$ is the maximal point such that $(x_i - d_i)/(x_j - d_j) = (\bar{x}_i - d_i)/(\bar{x}_j - d_j)$ for all $i, j \in N$.

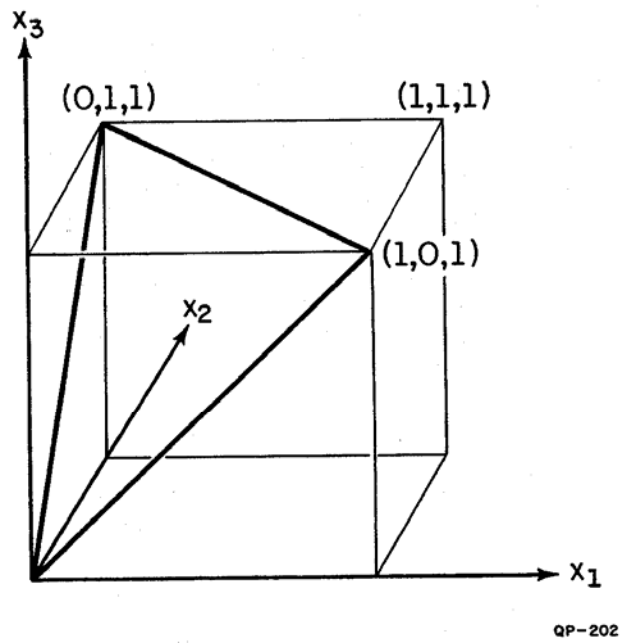


Figure 22

But the point $(1, 1, \dots, 1)$ is the ideal point both of the game $(S, \bar{0})$ and of the game $(T, \bar{0})$, so the fact that T contains S and the restricted monotonicity of f (cf. footnote 5) imply that $z \geq x$. Since we have shown that $z_3 = \frac{n-1}{n}$ and $x_3 = 1$, this gives the contradiction needed to complete the proof.

It is straightforward to verify that the solution G is symmetric and individually monotonic for any n . Theorem 16 therefore implies that G does not possess the property of Pareto optimality for $n \geq 3$; i.e., $G(S, d)$ need not be Pareto optimal in S . In fact, for the game $(S, \bar{0})$ considered in the proof, $G(S, \bar{0}) = \bar{0}$, which is not Pareto optimal.

This should serve to emphasize that G is very closely related to the proportional solutions studied in the previous section. For games (S, d) normalized so that $d_1 = d_2 = \dots = d_n$ and $\bar{x}_1(S) = \bar{x}_2(S) = \dots = \bar{x}_n(S)$, G coincides with the symmetric proportional solution. The difference between the two solutions lies in the way the outcomes of such games are generalized to other games: a proportional solution gives the players gains which are in fixed proportion to one another, while for any game (S, d) , G assigns the players gains in proportion to the quantities $\bar{x}_1(S) - d_1$.⁹

Dependence on the Ideal Point

The solution G is not independent of alternatives other than the disagreement point, since it is also dependent on the point $\bar{x}(S)$, which is sometimes referred to as the ideal point of the set S . However G is independent of alternatives which do not determine the ideal point: i.e., it is a solution which possesses the following property.

Property 16: Independence of Alternatives Other Than the Disagreement Point and Ideal Point. Suppose that (S, d) and (T, d) are bargaining games such that T contains S , $\bar{x}(S) = \bar{x}(T)$, and $f(T, d)$ is an element of S . Then $f(S, d) = f(T, d)$.

⁹For some related work, see Rosenthal [1976], and Schmitz [1977].

Of course G is not the only solution which possesses this property; for instance the Nash solutions possess it too, since any solution which possesses Property 3 must possess Property 16 as well. The relationship between Properties 3 and 16 suggests consideration of the following independence property as well.

Property 17: Independence of Alternatives Other Than the Ideal Point. If (S, d) and (T, d') are bargaining problems such that $\bar{x}(T) = \bar{x}(S)$ and T contains S , and if $f(T, d')$ is an element of S , then $f(S, d) = f(T, d')$.

Solutions which possess Property 17 have been considered in the literature; e.g., Yu [1973] and Freimer and Yu [1976] consider solutions which select the "closest" point in S to $\bar{x}(S)$, in terms of some metric. Such a solution is Pareto optimal and symmetric, but not independent of equivalent utility representations. In fact, we have the following general result (Roth [1977d]).

Theorem 17: There exists no solution on the class B which is symmetric, Pareto optimal, independent of alternatives other than the ideal point, and independent of equivalent utility representations.

Proof: It will be sufficient to demonstrate that no solution satisfying the conditions of the theorem exists for the case $n = 2$. Suppose to the contrary that f is such a solution, and let T be the convex hull of the points $(0, -2)$, $(-2, 0)$, $(-15, 0)$, $(-15, -15)$, and $(0, -15)$, and let $d' = (-15, -15)$. Then (T, d') is a symmetric bargaining problem whose Pareto surface is the line segment joining the points $(0, -2)$ and $(-2, 0)$. Consequently $f(T, d) = (-1, -1)$.

Let S be the convex hull of the points $(0, -6)$, $(-1/2, -3/2)$, $(-1, -1)$, $(-6, 0)$, and $(-6, -6)$. Then T contains S , $f(T, d') \in S$, and $\bar{x}(S) = \bar{x}(T) = (0, 0)$. So Property 17 requires that $f(S, d) = f(T, d') = (-1, -1)$ for any disagreement point d .

Now let $S' = \{(2x_1, (2/3)x_2) \mid (x_1, x_2) \in S\}$, and let $d'' = (2d_1, (2/3)d_2)$. Then independence of equivalent utility representations implies that $f(S', d'') = (-2, -2/3)$. But T contains S' , which is the convex hull of the points

Roth, Alvin E. *Axiomatic Models of Bargaining*, Springer-Verlag, 1979.
http://kuznets.fas.harvard.edu/~aroth/Axiomatic_Models_of_Bargaining.pdf

$(0,-4)$, $(-1,-1)$, $(-2,-2/3)$, $(-12,0)$, and $(-12,4)$. Also $\bar{x}(S') = \bar{x}(T) = (0,0)$ and $f(T,d') \in S'$, and so Property 17 implies that $f(S',d'') = f(T,d') = (-1,-1)$. This supplies the contradiction needed to complete the proof.

For some related work, see Thomson [1978], and Thomson and Myerson [1978].

Appendix: Summary of the main properties and results

Property 1. (p. 6) **Independence of Equivalent Utility Representations:** For any bargaining game (S, d) and real numbers a_i and b_i for $i = 1, \dots, n$ such that each $a_i > 0$, let the bargaining game (S', d') be defined by $S' = \{y \in R^n \mid \text{there exists an } x \text{ in } S \text{ such that } y_i = a_i x_i + b_i \text{ for } i = 1, \dots, n\}$ and $d'_i = a_i d_i + b_i$ for $i = 1, \dots, n$. Then $f_i(S', d') = a_i f_i(S, d) + b_i$ for $i = 1, \dots, n$.

Property 2. (p. 6) **Symmetry:** Suppose that (S, d) is a symmetric bargaining game—i.e., suppose that $d_1 = d_2 = \dots = d_n$, and that if x is contained in S , then so is every permutation of x . Then

$$f_1(S, d) = f_2(S, d) = \dots = f_n(S, d).$$

Property 3. (p. 6) **Independence of Irrelevant Alternatives:** Suppose that (S, d) and (T, d) are bargaining games such that T contains S , and $f(T, d)$ is an element of S . Then $f(S, d) = f(T, d)$.

Property 4. (p. 7) **Pareto optimality:** For any bargaining game (S, d) , if x and y are elements of S such that $y > x$, then $f(S, d) \neq x$.

Theorem 1 (Nash's theorem) (p. 8): There is a unique solution possessing Properties 1-4. It is the function $f = F$ defined by $F(S, d) = x$ such that $x \geq d$ and $\prod_{i=1}^n (x_i - d_i) > \prod_{i=1}^n (y_i - d_i)$ for all y in S such that $y \geq d$ and $y \neq x$.

Corollary 1.1 (p. 12): Suppose that for any symmetric bargaining game (A, d) such that $d = (0, \dots, 0)$ and $P(A)$ is a subset of the hyperplane $T = \{x \mid \sum x_i = n\}$, that f is a solution such that $f(A, d) = (1, \dots, 1)$. Then if f possesses Properties 1 and 3, f is identical to the Nash solution F .

Property 5. (p. 13) **Individual rationality:** For every bargaining game (S, d) , $f(S, d) \geq d$.

Theorem 2 (p. 14): There are precisely two solutions f possessing Properties 1, 2, 3, and 5. One is Nash's solution, $f = F$, and the other is the disagreement solution $f = D$ defined by $D(S,d) = d$ for all bargaining games (S,d) .

Theorem 3 (p. 16): For each strictly positive vector p with $\sum p_i = n$, there is a unique solution f possessing Properties 1, 3, and 5, such that $f(\bar{A}, \bar{O}) = p$. For any bargaining game (S,d) , $f(S,d) = x$, such that $x \geq d$ and $\Pi(x_i - d_i)^{p_i} > \Pi(y_i - d_i)^{p_i}$ for all y in S such that $y \geq d$ and $y \neq x$.

Property 6 (p. 17) **Strong individual rationality:** For every bargaining game (S,d) , $f(S,d) > d$.

Property 7 (p. 18) **Strong Pareto optimality:** For any bargaining game (S,d) , if x and y are distinct elements of S such that $y \geq x$, then $f(S,d) \neq x$.

Theorem 4 (p. 39): A utility function v is more risk averse than u if and only if, for all riskless outcomes c in C , $v(c) = k(u(c))$, where k is an increasing concave function.

Theorem 5 (p. 44): The utility which Nash's solution assigns to a player in a two-person game increases as his opponent becomes more risk averse. That is, $F_i(\hat{S}, \hat{d}) \geq F_i(S, d)$ where (\hat{S}, \hat{d}) is obtained from (S, d) by replacing player j , ($j \neq i$), with a more risk averse player.

Proposition 5.1 (p. 46): If v_i and u_i are two utility functions such that v_i is more risk averse than u_i , then the risk limit \bar{r}_i for any pair of proposals is lower under v_i than under u_i .

Property 8 (p. 47) **Risk sensitivity:** If a two-person bargaining game (S,d) is transformed into a game (S', d') by replacing player i with a more risk averse player, then $f_j(S', d') \geq f_j(S, d)$.

Theorem 6 (p. 47): If f is a solution for two-player games which possesses Properties 4 and 8, then it also possesses Property 1.

Theorem 7 (p. 50): $F(S^*, d^*) = z^* = (u_1(w_1 + c_1^*), u_2(w_2 + c_2^*))$ such that

$$c_1^* + c_2^* = Q \text{ and } b_1(w_1, c_1^*) = b_2(w_2, c_2^*).$$

Theorem 8 (p. 57): For preferences which obey Conditions 1-5 (pp. 54-56), the

utility θ of playing position k in a bargaining game $(S, \bar{0})$ is $\theta_k(S) = x_k$, where for any $S \in B^1$, $k \in N$, $\theta_k(S) = x_k$ where x is the unique element of S^+ such that $\Pi x_i^{q_i} > \Pi y_i^{q_i}$ for all $y \in S^+$ such that $y \neq x$, where $q = (q_1, \dots, q_n)$ is the vector such that $q_k = c$, $\sum q_i = n$, and $q_i = q_j$ for $i, j \neq k$.

Corollary 8.1 (p. 58): The Nash solution is equal to the utility vector θ reflecting strategic risk neutrality.

Property 9 (p. 69) Independence of ordinal transformations: For any bargaining

game (S, d) in B^* and any continuous, order preserving functions m_i ,

$i = 1, \dots, n$, let the bargaining game (S', d') be defined by

$S' = m(S) \equiv \{y \in R^n \mid y = m(x) \text{ for some } x \text{ in } S\}$ and $d' = m(d)$. Then

$$f_i(S', d') = m_i(f_i(S, d)) \text{ for } i = 1, \dots, n.$$

Theorem 9 (p. 69): No ordinally independent solution exists which also possesses the properties of strong individual rationality and independence of irrelevant alternatives.

Theorem 10 (p. 70): For two-player games, no ordinally independent solution exists which also possesses the property of strong individual rationality.

Corollary 10.1 (p. 74): For any n , there is no solution defined on the class B^* which is both ordinally independent and strongly individually rational.

Property 10 (p. 79): Independence of common scale changes: For any bargaining

game (S, d) and real numbers $a > 0$ and b_i , $i = 1, \dots, n$, let the game (S', d')

be defined by $S' = \{y \in R^n \mid \text{for some } x \text{ in } S, y_i = ax_i + b_i \text{ for } i = 1, \dots, n\}$ and

$$d'_i = ad_i + b_i \text{ for } i = 1, \dots, n. \text{ Then } f_i(S', d') = af_i(S, d) + b_i \text{ for } i = 1, \dots, n.$$

Property 11 (p. 82): Decomposibility: If (S, d) and (T, d) are bargaining games such that T contains S , then $f(T, d) = f(T, f(S, d))$, whenever $(T, f(S, d)) \in B$.

Property 12 (p. 82) **Monotonicity:** If (S,d) and (T,d) are bargaining games such that T contains S , then $f(T,d) \geq f(S,d)$.

Lemma 11.1 (p. 82): If f is a solution which is independent of common scale changes, decomposable, and individually rational, then f is monotonic.

Lemma 11.2 (p. 83): If f is a decomposable solution which is strongly individually rational on the class D , then f is (weakly) Pareto optimal on the class D .

Lemma 11.3 (p. 84): If f is an individually rational solution which is independent of common scale changes and decomposable, and if f is strongly individually rational on the class D , then f is proportional on the class D .

Theorem 11 (p. 86): Let f be an individually rational solution which is strongly individually rational on the class D , and which is decomposable and independent of common scale changes. Then for any game $(S,d) \in B$,
 $f(S,d) = k(S,d)p + d$ where $0 \leq k(S,d) \leq \lambda(S,d)$.

Property 13 (p. 87): **Individual monotonicity.** If (T,d) and (S,d) are games such that T contains S and, for some $i \in N$, T_D^+ intersected with $\{x \in R^n | x_i = d_i\}$ is equal to the intersection of S_D^+ $\{x \in R^n | x_i = d_i\}$, then
 $f_i(T,d) \geq f_i(S,d)$.

Property 14 (p. 88): **Continuity on the class D :** Let $(S_j,d) \in D$ be bargaining games defined for a sequence of sets S_j such that, in the limit as j goes to infinity, S_j converges to S (in the Hausdorff topology), where (S,d) is also in the class D . Then $\lim_{j \rightarrow \infty} f(S_j,d) = f(S,d)$.

Theorem 12 (p. 88): If f is an individually rational solution which is strongly individually rational on the class D , and if f is independent of irrelevant alternatives, independent of common scale changes, individually monotonic, and continuous on the class D , then f is a proportional solution on the class B .

Property 15 (p. 93): Independence of ordinal transformations preserving interpersonal comparisons: Let (S, d) and (S', d') be two-player games in B^* such that $d'_i = t_i(d_1, d_2)$ for $i = 1, 2$ and $S' = \{(t_1(x_1, x_2), t_2(x_1, x_2)) \mid (x_1, x_2) \in S\}$ where $t = (t_1, t_2)$ is a transformation $t: S \rightarrow R^2$ such that for all x, y in S and $i = 1, 2$,

- (i) $t_i(x_1, x_2) \geq t_i(y_1, y_2)$ if and only if $x_i \geq y_i$ and
- (ii) $t_1(x_1, x_2) - t_1(d_1, d_2) \geq t_2(x_1, x_2) - t_2(d_1, d_2)$ if and only if $x_1 - d_1 \geq x_2 - d_2$.

Then $f_i(S', d') = t_i(f(S, d))$.

Theorem 13 (p. 93): There is a unique solution f which is strongly individually rational, strongly Pareto optimal, independent of irrelevant alternatives, and independent of ordinal transformations preserving interpersonal comparisons. It is the solution $f = E$ defined by $E(S, d) = x$ such that $\min\{x_1 - d_1, x_2 - d_2\} > \min\{y_1 - d_1, y_2 - d_2\}$ for all $y \in \bar{P}(S)$ such that $y \neq x$.

Theorem 14 (p. 99): The solution G is the unique solution defined on two player games in the class B which is strongly Pareto optimal, symmetric, independent of equivalent utility representations, and individually monotonic.

Theorem 15 (p. 104): The utility which the solution G assigns to a player in a two-person game increases as his opponent becomes more risk averse. That is, for $i = 1$ or 2 , $G_i(\hat{S}, \hat{d}) \geq G_i(S, d)$, where (\hat{S}, \hat{d}) is obtained from (S, d) by replacing player j ($j \neq i$) with a more risk averse player.

Theorem 16 (p. 105): For bargaining games with 3 or more players, no solution exists which is Pareto optimal, symmetric, and individually monotonic.

Property 16 (p. 107): Independence of alternatives other than the disagreement point and ideal point: Suppose that (S, d) and (T, d) are bargaining games such that T contains S , $\bar{x}(S) = \bar{x}(T)$, and $f(T, d)$ is an element of S . Then $f(S, d) = f(T, d)$.

Property 17 (p. 108): Independence of alternatives other than the ideal point.

If (S, d) and (T, d') are bargaining problems such that $\bar{x}(T) = \bar{x}(S)$ and T contains S , and if $f(T, d')$ is an element of S , then $f(S, d) = f(T, d')$.

Theorem 17 (p. 108): There exists no solution on the class B which is symmetric, Pareto optimal, independent of alternatives other than the ideal point, and independent of equivalent utility representations.

Bibliography

- Anbar, Dan and Ehud Kalai [1978], "A One-Shot Bargaining Problem," International Journal of Game Theory (to appear).
- Arrow, Kenneth J. [1965], Aspects of the Theory of Risk-Bearing, Yrjö Jahnsson Foundation, Helsinki.
- Arrow, Kenneth J. [1971], Essays in the Theory of Risk Bearing, Chicago: Markham.
- Ashenfelter, Orley and George E. Johnson [1969], "Bargaining Theory, Trade Unions, and Industrial Strike Activity," American Economic Review, Vol. 59, pp. 35-49.
- Aumann, Robert J. and M. Kurz [1977a], "Power and Taxes," Econometrica, Vol. 45, pp. 1137-1161.
- Aumann, Robert J. and M. Kurz [1977b], "Power and Taxes in a Multi-Commodity Economy," Israel Journal of Mathematics, Vol. 27, pp. 185-234.
- Bishop, Robert L. [1963], "Game Theoretic Analyses of Bargaining," Quarterly Journal of Economics, Vol. LXXVII, pp. 559-602.
- Bishop, Robert L. [1964], "A Zeuthen-Hicks Theory of Bargaining," Econometrica, Vol. 32, pp. 410-417.
- Bradley, Gordon H. and Martin Shubik [1974], "A Note on the Shape of the Pareto Optimal Surface," Journal of Economic Theory, Vol. 8, No. 4, August, pp. 530-538.
- Butrim, B. I. [1976], "A Modified Solution of the Bargaining Problem," (Russian), Zhurnal Vychislitel'noi Matematiki Matmematicheskoi Fiziki, Vol. 16, pp. 340-350.
- Butterworth, Robert L. [1977], "Relative Acceptability: A Proposed Strategy for the Bargaining Game," Theory and Decision, Vol. 8, No. 4, pp. 337-350.
- Coddington, Alan [1968], Theories of the Bargaining Process, London.
- Comay, Yochanan, Arie Melnik, and Abraham Subotnik [1974], "Bargaining, Yield Curves, and Wage Settlements: An Empirical Analysis," Journal of Political Economy, Vol. 82, pp. 303-313.
- Cross, John G. [1965], "A Theory of the Bargaining Process," American Economic Review, Vol. 40, pp. 67-94.
- Cross, John G. [1969], The Economics of Bargaining, New York: Basic Books.
- Crott, H. W. [1971], "Experimentelle Untersuchung Zum Verhandlungsverhalten in Kooperativen Spielen," Zeitschrift für Sozialpsychologie, Vol. 2, pp. 61-74.
- de Menil, George [1971], Bargaining: Monopoly Power Versus Union Power, Cambridge: MIT Press.
- Edgeworth, F. Y. [1881], Mathematical Psychics, London: Routledge and Kegan Paul.
- Fouraker, Lawrence E. and Sidney Siegel [1963], Bargaining Behavior, New York: McGraw-Hill.
- Freimer, M. and P. L. Yu [1976], "Some New Results on Compromise Solutions for Group Decision Problems," Management Science, Vol. 22, pp. 688-693.

- Harsanyi, John C. [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, Vol. 24, pp. 144-157.
- Harsanyi, John C. [1958], "Notes on the Bargaining Problem," Southern Economic Journal, Vol. 24, pp. 471-476.
- Harsanyi, John C. [1961], "On the Rationality Postulates Underlying the Theory of Cooperative Games," Journal of Conflict Resolution, Vol. V, pp. 179-196.
- Harsanyi, John C. [1963], "A Simplified Bargaining Model for the n-Person Cooperative Game," International Economic Review, Vol. 4, pp. 194-220.
- Harsanyi, John C. [1965], "Bargaining and Conflict Situations in the Light of a New Approach to Game Theory," American Economic Review, Vol. 55, pp. 447-457.
- Harsanyi, John C. [1968], "Games with Incomplete Information Played by 'Bayesian' Players: Part II," Management Science, Vol. 14, pp. 320-334.
- Harsanyi, John C. [1975], "The Tracing Procedure: A Bayesian Approach to Defining a Solution for n-Person Noncooperative Games," International Journal of Game Theory, Vol. 4, pp. 61-94.
- Harsanyi, John C. [1977], Rational Behavior and Bargaining Equilibrium in Games and Social Situations, Cambridge: Cambridge University Press.
- Harsanyi, John C. [1978], "A Solution Concept for n-Person Noncooperative Games," International Journal of Game Theory (forthcoming).
- Harsanyi, John C. and Reinhard Selten [1972], "A Generalized Nash Solution for Two-Person Bargaining Games with Incomplete Information," Management Science, Vol. 18, Part II, pp. 80-106.
- Hart, Sergiu [1977], "The Nash Solution of Two-Person Bargaining Games," mimeographed lecture notes, Stanford.
- Herstein, I. N. and J. W. Milnor [1953], "An Axiomatic Approach to Measurable Utility," Econometrica, Vol. 21, pp. 291-297.
- Hicks, J. R. [1932], The Theory of Wages, London: Macmillan.
- Isbell, J. R. [1959], "Absolute Games," in Contributions to the Theory of Games IV, (A. W. Tucker and R. D. Luce, eds.), Annals of Mathematics Studies, No. 40, pp. 357-396.
- Isbell, J. R. [1960], "A Modification of Harsanyi's Bargaining Model," Bulletin of the American Mathematical Society, Vol. 66, pp. 70-73.
- Kalai, Ehud [1977a], "Nonsymmetric Nash Solutions and Replications of Two-Person Bargaining," International Journal of Game Theory, Vol. 6, pp. 129-133.
- Kalai, Ehud [1977b], "Proportional Solutions to Bargaining Situations: Interpersonal Utility Comparisons," Econometrica, Vol. 45, pp. 1623-1630.
- Kalai, Ehud and Robert W. Rosenthal [1978], "Arbitration of Two-Party Disputes Under Ignorance," International Journal of Game Theory, Vol. 7, pp. 65-72.
- Kalai, Ehud and Meir Smorodinsky [1975], "Other Solutions to Nash's Bargaining Problem," Econometrica, Vol. 43, pp. 513-518.

- Kaneko, Mamoru and Kenjiro Nakamura [1978a], "The Nash Social Welfare Function," Econometrica.
- Kaneko, Mamoru and Kenjiro Nakamura [1978b], "Cardinalization of the Nash Social Welfare Function," mimeo.
- Kihlstrom, Richard E. and Leonard J. Mirman [1974], "Risk Aversion with Many Commodities," Journal of Economic Theory, Vol. 8, pp. 361-388.
- Kihlstrom, Richard and Alvin E. Roth [1979], "Risk Aversion and the Negotiation of Insurance Contracts," mimeo.
- Kihlstrom, Richard, Alvin E. Roth and David Schmeidler [1979], "Risk Aversion and Nash's Solution to the Bargaining Problem," mimeo.
- Krantz, David H., R. Duncan Luce, Patrick Suppes and Amos Tversky [1971], Foundations of Measurement Volume I: Additive and Polynomial Representations, Academic Press, New York.
- Luce, R. Duncan and Howard Raiffa [1957], Games and Decisions: Introduction and Critical Survey, New York: Wiley.
- Myerson, Roger B. [1977], "Two-Person Bargaining Problems and Comparable Utility," Econometrica, Vol. 45, pp. 1631-1637.
- Myerson, Roger B. [1979], "Incentive Compatibility and the Bargaining Problem," Econometrica, Vol. 47, pp. 61-74.
- Nash, John F. [1950], "The Bargaining Problem," Econometrica, Vol. 28, pp. 155-162.
- Nash, John F. [1961], "Noncooperative Games," Annals of Mathematics, Vol. 54, pp. 286-295.
- Nash, John F. [1953], "Two-Person Cooperative Games," Econometrica, Vol. 21, pp. 129-140.
- Nydegger, Rudy V. and Guillermo Owen [1975], "Two-Person Bargaining: An Experimental Test of the Nash Axioms," International Journal of Game Theory, Vol. 3, pp. 239-249.
- O'Neill, Barry [1976], "Variable Threat Bargaining: A Test of the Raiffa-Nash Theory," mimeo.
- O'Neill, Barry [1978], "The Number of Outcomes in the Pareto-Optimal Set of Discrete Bargaining Games," mimeo.
- Owen, Guillermo [1968], Game Theory, Philadelphia: Saunders.
- Perles, Micha and Michael Maschler [1979], "A Superadditive Solution to Nash Bargaining Games," preliminary manuscript.
- Pen, Jan [1952], "A General Theory of Bargaining," American Economic Review, XLII, pp. 24-42.
- Pen, Jan [1959], The Wage Rate Under Collective Bargaining, Cambridge: Harvard University Press.
- Pratt, J. W. [1964], "Risk Aversion in the Small and in the Large," Econometrica, Vol. 32, pp. 122-136.

- Raiffa, Howard [1953], "Arbitration Schemes for Generalized Two-Person Games," in H. W. Kuhn and A. W. Tucker (eds.), Contributions to the Theory of Games II, Annals of Mathematics Studies, No. 28, Princeton: Princeton University Press, pp. 361-387.
- Rapoport, Anatol, Melvin J. Guyer and David G. Gordon [1976], The 2 x 2 Game, Ann Arbor: University of Michigan Press.
- Rapoport, Anatol, Oded Frenkel and Josef Perner [1977], "Experiments with Cooperative 2 x 2 Games," Theory and Decision, 8, pp. 67-92.
- Richter, Marcel K. [1966], "Revealed Preference Theory," Econometrica, Vol. 34, pp. 635-645.
- Rosenthal, Robert W. [1976], "An Arbitration Model for Normal-Form Games," Mathematics of Operations Research, Vol. 1, pp. 82-88.
- Rosenthal, Robert W. [1978], "Arbitration of Two-Party Disputes Under Uncertainty," Review of Economic Studies, forthcoming.
- Roth, Alvin E. [1977a], "Individual Rationality and Nash's Solution to the Bargaining Problem," Mathematics of Operations Research, Vol. 2, No. 1.
- Roth, Alvin E. [1977b], "The Shapley Value as a von Neumann-Morgenstern Utility," Econometrica, Vol. 45, pp. 657-664.
- Roth, Alvin E. [1977c], "Bargaining Ability, the Utility of Playing a Game and Models of Coalition Formation," Journal of Mathematical Psychology, Vol. 16, pp. 153-160.
- Roth, Alvin E. [1977d], "Independence of Irrelevant Alternatives and Solutions to Nash's Bargaining Problem," Journal of Economic Theory, Vol. 16, No. 2.
- Roth, Alvin E. [1978], "The Nash Solution and the Utility of Bargaining," Econometrica, Vol. 46, pp. 587-594, 983.
- Roth, Alvin E. [1979a], "Proportional Solutions to the Bargaining Problem," Econometrica, Vol. 47, pp. 775-778.
- Roth, Alvin E. [1979b], "The Nash Solution as a Model of Rational Bargaining," Extremal Methods and Systems Analysis, Springer Verlag, forthcoming.
- Roth, Alvin E. [1979c], "Interpersonal Comparisons and Equal Gains in Bargaining," mimeo.
- Roth, Alvin E. [1979d], "An Impossibility Result Concerning n-Person Bargaining Games," International Journal of Game Theory, forthcoming.
- Roth, Alvin E. and Michael W. K. Malouf [1979], "Game-Theoretic Models and the Role of Information in Bargaining," mimeo.
- Saraydar, Edward [1965], "Zeuthen's Theory of Bargaining: A Note," Econometrica, Vol. 33, pp. 802-813.
- Saraydar, Edward [1971a], "Uncertainty, the Bargaining Problem, and the Nash-Zeuthen Solution," Theory and Decision, Vol. 1, pp. 309-319.
- Saraydar, Edward [1971b], "A Certainty-Equivalent Model of Bargaining," Journal of Conflict Resolution, Vol. XV, pp. 281-297.

- Schelling, Thomas C. [1956], "Essays on Bargaining," American Economic Review, Vol. 46, pp. 281-306.
- Schelling, Thomas C. [1957], "Bargaining, Communication, and Limited War," Journal of Conflict Resolution, Vol. I, pp. 19-36.
- Schelling, Thomas C. [1958], "The Strategy of Conflict: Prospectus for The Re-orientation of Game Theory," Journal of Conflict Resolution, Vol. II, pp. 203-264.
- Schelling, Thomas C. [1959], "For the Abandonment of Symmetry in Game Theory," Review of Economics and Statistics, Vol. 41, pp. 213-224.
- Schelling, Thomas C. [1960], The Strategy of Conflict, Cambridge: Harvard University Press.
- Schmitz, Norbert [1977], "Two-Person Bargaining Without Threats: A Review Note," Proceedings of the OR Symposium - Aachen,
- Shapley, Lloyd S. [1969], "Utility Comparison and the Theory of Games," in La Decision (G. Th. Guilbaud, ed.), Editions du CNRS, Paris.
- Shapley, Lloyd S. and M. Shubik [1974], "Game Theory in Economics-Chapter 4: Preferences and Utility," Rand Corporation, R-904/4 NSF, December.
- Shubik, Martin [1959], Strategy and Market Structure: Competition, Oligopoly, and the Theory of Games, New York: Wiley.
- Siegel, Sidney and Lawrence E. Fouraker [1960], Bargaining and Group Decision Making: Experiments in Bilateral Monopoly, New York: McGraw Hill.
- Stahl, Ingolf [1972], Bargaining Theory, EFI, The Economic Research Institute, Stockholm.
- Stone, J. J. [1958], "An Experiment in Bargaining Games," Econometrica, Vol. 26, pp. 286-297.
- Thomson, William [1978], "A Class of Solution to Bargaining Problem," mimeo.
- Thomson, William and Roger Myerson [1978], "Monotonicity and Independence Axioms," mimeo.
- von Neumann, John and Oskar Morgenstern [1944], Theory of Games and Economic Behavior, Princeton: Princeton University Press.
- Wagner, Harvey M. [1957], "A Unified Treatment of Bargaining Theory," Southern Economic Journal, Vol. 23, pp. 380-397.
- Wagner, Harvey M. [1958], "Rejoinder on the Bargaining Problem," Southern Economic Journal, Vol. 24, pp. 476-482.
- Wagner, R. Harrison [1979], "On The Unification of Two-Person Bargaining Theory," Journal of Conflict Resolution, Vol. 23, pp. 71-101.
- Yaari, M. E. [1969], "Some Remarks on Measures of Risk Aversion and On Their Uses," Journal of Economic Theory, Vol. 1, pp. 315-329.
- Young, Oran R. [1975], Bargaining: Formal Theories of Negotiation, Urbana: University of Illinois Press.

Roth, Alvin E. *Axiomatic Models of Bargaining*, , Springer-Verlag, 1979.
http://kuznets.fas.harvard.edu/~aroth/Axiomatic_Models_of_Bargaining.pdf

Yu, P. L. [1973], "A Class of Solutions for Group Decision Problems," Management Science, Vol. 19, pp. 936-946.

Zeuthen, Frederik [1930], Problems of Monopoly and Economic Welfare, London: G. Routledge.