Chapter 2

THE NASH SOLUTION TO THE BARGAINING PROBLEM

2.1 Introduction

Policy formation is a process of political interaction among individuals and groups. While some participants in the political process may share similar, or even identical interests, the political preferences of other participants diverge so that conflicts of interest among participating parties are unavoidable. Nevertheless, such conflicts are often resolved in the policy formulation process. The emerging policies reflect the participants' policy preferences as well as their political power. This is the essence of the political-power theory of policy formation. The following major questions arise: What is the nature of the political interaction giving rise to observed economic policy? How are the political conflicts resolved? How can this process be modeled? It is only natural to presume that interactions emerge through bargaining and negotiations among participants.

Among the various approaches to the solution of the inherent bargaining problem, the Nash/Harsanyi (NH) conceptualization provides an internally consistent framework. Along with the strong theoretical foundations of the NH theory it is also a convenient analytical model. The theoretical foundation for this framework is also the basis of Harsanyi's (1962a, 1962b) model of social power. As this particular bargaining theory is widely employed in this book, the present chapter is dedicated to the introduction and exposition of the NH theory. Intuition is emphasized in the presentation as many theoretical details are available elsewhere.[1]

Nash's axiomatic solution to the two-person bargaining problem, given fixed disagreement payoffs, represents the core of the bargaining framework presented in this chapter. In Chapter 3 we will analyze the problem of mutually optimal threat strategies that the two parties may select in order to influence disagreement payoffs and thus indirectly the bargaining outcome. In effect, the models presented in Chapter 3 treat Nash's solution to the bargaining problem, introduced in Section 2.2 as the solution to the second-stage subgame of a more complex, two-stage bargaining problem. Because this more complex problem is solved through backwards induction, it makes sense to initially present the simpler, second-stage problem. It is important to keep in mind that any strengths or limitations of Nash's axiomatic solution to the game discussed in this chapter are directly inherited by solutions to the larger game considered in Chapter 3.

All political-economic analytical frameworks must specify the nature of the interaction between policy makers and special interests. The axiomatic approach of Nash suppresses many details of the decision making process and explains outcomes by identifying conditions that any outcome arrived at by rational decision makers should satisfy a priori. These conditions are treated as axioms, from which the outcome is deduced using set-theoretical arguments.

For the analysis of many political-economic problems the strengths of the axiomatic approach are undeniable. It is important, however, to be aware that the approach also has limitations. For some problems, such as analyses of how politicians and governmental officials might select and implement a given set of policy instruments, these limitations are not generally confronted. For other political-economic problems, such as analyses of the underlying collective choice rules or institutional designs that structure the policy making process, the limitations of the axiomatic approach become serious. In these instances, instead

[1] Readers interested in a more rigorous elucidation of the NH bargaining theory are advised to consult Harsanyi (1977).

of an axiomatic approach, a strategic approach may be required that models constraints on the decision making process itself and generates outcomes by determining the equilibrium non-cooperative strategies of decision makers facing those constraints.

Whether the axiomatic or the strategic approach is justified depends upon a pivotal axiom used by Nash (1950), the so-called "independence of irrelevant alternatives" (IIA) axiom. In Section 2.3 we examine the critical role of this axiom and its relationship with a condition used by Arrow (1951, 1953) in the derivation of his famous impossibility theorem. It is demonstrated that the IIA axiom is largely responsible for both the strengths and the limitations of the Nash axiomatic approach. Four formulations are assessed that belong to the strategic approach to bargaining theory, but "implement" the Nash bargaining solution. In other words, all four formulations result in non-cooperative outcomes that coincide with the Nash axiomatic cooperative solution. Section 2.4 concludes the chapter.

2.2 The Nash Solution to the Bargaining Problem with Fixed Disagreement Payoffs

We first consider the two-person bargaining problem with fixed disagreement payoffs. This is the problem considered by Nash (1950) in a paper that provided the foundation of modern bargaining theory. The Nash two-person solution to this problem can easily be generalized to the n-person case. The general two-person bargaining problem may be stated as follows: Consider two persons, 1 and 2, whose preferences over outcomes are given by the utility functions u_1 and u_2, respectively. The payoff vector, $u = (u_1, u_2)$, is an element of a two-dimensional payoff space, P, (i.e., $u \in P$). P is assumed to be compact and convex.[2] Let H be the set of payoff vectors in P not dominated, even weakly, by any other payoff vector in

[2] Compactness of P implies that the payoff space is bounded and closed. Boundedness is a natural assumption, otherwise a solution need not exist. Closedness is desirable because it ensures that functions defined on P have a maximum, while convexity is required to ensure uniqueness of the solution payoffs. Note that the use of randomized mixed strategies is a sufficient condition for the convexity of P.

P. We shall refer to H as the *upper-right boundary* of P. Obviously, $H \subset P$ is the efficiency frontier of P. Let $t = (t_1, t_2) \in P$ be the vector of disagreement payoffs of person 1 and person 2, respectively, $t_i\,(i = 1, 2)$ being the payoff that person i gets if the parties fail to agree.

Let $P^* = \{u \in P : u_1 \geq t_1, u_2 \geq t_2\}$. Clearly, $P^* \subseteq P$. It is assumed that t is fixed, i.e., t_1 and t_2 are determined by the rules of the game. Let H^* denote the upper-right boundary of P^*. Thus, $H^* \subseteq H$. The bargaining problem is then: Given P and t, what will be the solution, $\bar{u} = (\bar{u}_1, \bar{u}_2)$, that the bargaining parties will eventually reach, assuming all individuals act rationally?[3]

Classical theory, recognizing only ordinal utility functions, is capable of providing only two relevant rationality axioms. These are:

- **Individual rationality (IR):** No person will agree to accept a payoff lower than the one guaranteed to him under disagreement; namely,

$$\bar{u}_i \geq t_i \qquad\qquad (i = 1, 2)$$

  so that $\bar{u}_i \in P^*$.

- **Pareto optimality (PO):** The agreement will represent a situation that could not be improved on to both persons' advantage (because rational participants would not accept a given agreement if some alternative arrangement could make both parties better off or at least one better off with the other no worse off).

These two classical axioms limit the solution, $\bar{u}$, to H^* (the "negotiation set" as defined by Luce and Raiffa (1957)). But the negotiation set, which also happens to be the core of the game, is not a unique solution, $\bar{u}$. Nash (1950) proposed a unique solution to the bargaining game which is based on the two classical axioms plus three additional axioms. These additional Nash axioms are:

[3]See, for instance, Harsanyi (1977).

- **Symmetry (SYM):** Let P^* be "symmetric"; namely, if any vector, $(a, b) \in P^*$, then the vector (b, a) is also in P^*. Then, if P^* is symmetric, $\bar{u}_1 = \bar{u}_2$.

- **Linear invariance (LINV):** Let $\bar{u}$ be the solution of the bargaining game, G. Let G^* be the game that results from G if one party's utility function, u_i, is subjected to an order-preserving linear transformation, T, leaving the other player's utility function, u_j, unchanged. Then the solution $\bar{u}^*$ of the new game, G^*, is the image of $\bar{u}$ under T, i.e., $\bar{u}^* = T\bar{u}$.

- **Independence of irrelevant alternatives (IIA):** Let G be the bargaining game with payoff space P and disagreement payoff t, and let $\bar{u}$ be the solution of G. Let G^* be the game obtained from G by restricting P to $Q \subset P$ such that $t \in Q$ and $\bar{u} \in Q$. Then $\bar{u}$ is also the solution of G^*.

Nash demonstrated that under the above five axioms the solution, $\bar{u} = (\bar{u}_1, \bar{u}_2)$, is the point satisfying

$$(\bar{u}_1 - t_1)(\bar{u}_2 - t_2) = \max_{u \in P} [(u_1 - t_1)(u_2 - t_2)] \tag{2.1a}$$

such that

$$u_i \geq t_i \qquad (i = 1, 2) \tag{2.1b}$$

Nash's solution to the two-person bargaining problem easily generalizes to the n-person case. The simple n-person bargaining game is defined as follows: Let N be the set of n bargaining parties, i.e., $|N| = n$. Let $u = (u_1, u_1, ..., u_n)$ be the vector of individual bargaining parties' utility functions (payoffs). Hence the payoff space P is also n-dimensional; i.e., $P \subset \Re^n$, where $\Re^n$ is the n-dimensional space of real numbers. It is again assumed that

P is compact and convex.[4] Let $H \subset P$ denote the upper right boundary of P, and let $H(u) = 0$ be its equation. Thus, H is a $(n-1)$-dimensional surface in $\Re^n$.

Suppose the vector of the parties' disagreement (conflict) payoffs, $t = (t_1, ..., t_n)$, is given by the rules of the game. Then the bargaining problem is: Given the above description of the simple bargaining game, what are the solution payoffs of the game? As before, classical theory provides only two rationality axioms: (i) individual rationality, which implies $\bar{u}_i \geq t_i$, and (ii) group rationality, which asserts that the individual bargaining parties will not accept any particular solution if another Pareto-superior feasible solution exists. Classical theory, therefore, restricts the solution payoffs to the "negotiation set," a set of payoffs, H^*, contained in H in which all payoffs, $\bar{u}_i \geq t_i$. But the negotiation set does not define a unique solution. As in the two-person bargaining case, a unique solution obtains if one adds n-person analogs to the Nash axioms of *symmetry, linear invariance*, and *independence of irrelevant alternatives*. The full set of axioms indeed yields a unique solution. The unique solution of an n-person simple bargaining game is the particular payoff vector, $\bar{u}$, which maximizes the n-person Nash product,

$$\pi(u) = \prod_{i \in N} (u_i - t_i), \tag{2.2a}$$

subject to the constraints,

$$u \in P \tag{2.2b}$$

and

$$u_i \geq t_i \qquad t_i \text{ constant for all } i \in N.[5] \tag{2.2c}$$

[4] See footnote 2.
[5] See Harsanyi (1977: Theorems 10.1 and 10.2).

2.3 The Pivotal Axiom and Alternative Approaches

In the literature on bargaining theory there are various views on whether an axiomatically derived solution can be implemented non-cooperatively. Nash himself commented as follows on his own demonstration:

> "It is rather significant that this quite different (axiomatic) approach yields the same solution. This indicates that the solution is appropriate for a wider variety of situations than those which satisfy the assumptions we made in the approach via the (strategic) model." (Nash, 1953: 136)

Nash suggests here that his axiomatic approach to bargaining game theory, which he initiated in Nash (1950), is in some sense more "powerful" than the strategic approach, which he initiated in the concluding paragraphs of Nash (1951).[6] Subsequent to the early work of Nash, the more accepted view in the bargaining literature is that the two approaches are complementary. Sutton (1986), for example expresses this view as follows:[7]

> "...the detailed process of bargaining will differ so widely from one case to another that any useful theory of bargaining must involve some attempts to distil out some simple principles which will hold over a wide range of possible processes. What an axiomatic approach attempts to do is to codify some set of principles of this kind. To design such a set of axioms, though, we need at least to carry out some thought experiments, in order to guide our intuition as to what principles are reasonable, or compelling. The easiest way to do this is to imagine some particular process which might be followed, and to ask whether or not the principle will hold good in the case. This motivates the idea of looking at some example(s) of non-cooperative games which correspond to a particular process." (Sutton, 1986: 709)

[6]This is the same paper that introduced the famous solution concept for non-cooperative games, now known as the "Nash equilibrium."

[7]Binmore and Dasgupta (1987) also discuss at length the complementarity of the two approaches.

The collective effort by game theorists to construct such non-cooperative games with the purpose set out by Sutton in mind is commonly referred to as the "Nash program." A third view on the relationship between axiomatic and strategic models stresses the primacy of the latter. If an axiomatic solution concept is appropriate, it should be applied only in collective decision making contexts with specific features, namely, those features revealed by strategic models to yield solutions consistent with its axioms. Accordingly, it is non-cooperative implementations that broaden the scope of application of axiomatically derived solution concepts; not vice versa. Any serious examination of Nash's IIA axiom must be based on the primacy of this strategic model.

It is important to recognize that the IIA axiom has no logical relationship to the condition by the same name used by Arrow (1951, 1963) in the derivation of his famous impossibility result.[8] As demonstrated by Ray (1973), neither implies the other. Arrow's condition concerns irrelevant changes in individual *preferences*, holding the set of alternatives constant. It specifies that the choice made collectively from a given set, S, should not change when individual preferences over alternatives outside of that set change. Nash's axiom, on the other hand, concerns irrelevant changes in the *set of alternatives*, holding individual preferences constant.

Ray notes that if individual rankings over the universal set of alternatives X are aggregated by the so-called rank-order method (a form of weighted voting to choose from any subset S of X), the resulting collective choices satisfy Nash's axiom, but not Arrow's condition. On the other hand, a (somewhat peculiar) choice procedure that selects the maximum of some social welfare function from all proper subsets S of X, but selects the minimum of that function from X itself, would satisfy Arrow's condition, but not Nash's axiom.

[8] The Arrow impossibility result demonstrates that, given some reasonable and mild requirements, individual ordinal rankings cannot provide an acceptable basis for making social decisions about how wealth should be distributed or more generally for resolving conflicts among various interests. This result is not surprising since individual ordinal rankings provide little insight about the relative importance of different interests. As Sen (1970) has shown, if interpersonal comparisons of preference intensity or non-preference information are introduced into the formulation, the impossibility result no longer necessarily holds. Moreover, when constraints such as "single peakedness" and "unidimensionality" are imposed on preference profiles, the impossibility theorem most certainly does not hold.

That the Nash IIA and Arrow IIA axioms are often confused in the literature is perhaps not so surprising if one considers that Arrow himself confused the two, and did so in the very treatise that introduced his own condition. To illustrate a case in which his condition would be violated, Arrow constructs an example in which three voters have to decide on one of four candidates – x, y, z, and w. Voters 1 and 2 have preferences $x \succ y \succ z \succ w$, whereas voter 3 has preferences $z \succ w \succ x \succ y$. Arrow shows that if they aggregate their rankings over all four candidates using the rank-order method, the winner is x. If then the "irrelevant" candidate y is deleted and the voters aggregate their ranking over the remaining three, a tie between x and z results. Clearly, Arrow's example illustrates a violation not of his own condition, but of Nash's axiom.[9]

To illustrate the crucial importance of the Nash IIA axiom, four collective decision making stories are revealing. Each of the stories unfolds under two different scenarios, in the second of which one or more alternatives other than the solution (or the disagreement point) of the first scenario are excluded. The stories are presented in no particular order, so as not to bias the reader either way in deciding whether the outcome under the second scenario will be different from that under the first.

Story 1 happens to be a real-life anecdote related by Aumann (1985) in a discussion of the IIA axiom. The first scenario unfolds as follows:

> "Several years ago I served on a committee that was to invite a speaker for a fairly prestigious symposium. Three candidates were proposed: their names would be familiar to many of our readers, but we will call them Alfred Adams, Barry Brown, and Charles Clark. A long discussion ensued, and it was finally decided to invite Adams."

Under the second scenario, one of the two candidates not chosen under the first scenario is no longer available:

[9]For an alternative view on this point, see Bordes and Tideman (1991).

> "At that point I remembered that Brown had told me about a family trip that he was planning for the period in question, and realized that he would be unable to come. I mentioned this and suggested that we reopen the discussion."

We leave it to the reader to anticipate how the committee reacted to Aumann's suggestion.

Story 2 is fictional, but at the same time realistic enough that it could unfold in most any modern economy the reader might like to imagine. It concerns a government agency charged with regulating a public utility. The agency does not have complete discretion over the price; it is bound by law to limit yearly price increases to no more than 5% over inflation. Under the first scenario, the agency decides, after weighing a plea by executives of the utility for a maximum price increase to help fund new equipment investments, that this year a price increase of 2% over inflation best meets its overall policy objectives. Under the second scenario, all else is equal, but the legal upper bound on price increases is 3% rather than 5%. Will the agency's decision be different?

Story 3 is again a fictional but realistic story on wage negotiations. Under the first scenario, the labor union comes out initially with a demand for a 19% wage increase and the employer offers 4%. Protracted negotiations follow, which result in a stalemate: The employer's absolutely final offer is 9%, but the union refuses to accept anything under 11%. Only after a two-week strike do the two sides finally agree on a 9.9% wage hike. Under the second scenario all else is equal, but the government, in an attempt to fight inflation, has imposed general wage controls. In no industry are wages allowed to increase by more than 10%. Obviously, this will affect the labor union's initial demand. But will it also affect the final agreement?

Story 4 is again real-life, at least under the first scenario. It is the story of the 1992 American presidential campaign, pitting the incumbent, Republican President Bush, against Democratic candidate Clinton and unaffiliated candidate Perot. One of the dramatic events of the campaign was Perot's withdrawal from the race on July 16, claiming that the Republican Party was planning to disrupt both the wedding of his daughter and his business operations. Under the first scenario, Perot re-enters the race on October 1, with just a month to go before election day. Clinton ends up winning the election with 43% of the vote, against Bush's 38% and Perot's 19%. Under the second scenario Perot stays out of the race. Will Clinton still be the winner?

Story 1 is a fairly convincing example of a collective decision making context in which it is reasonable to expect the outcome to be consistent with the IIA axiom. Aumann, in fact, presents the story in defense of the IIA axiom, and recounts the following reaction by the committee members to his suggestion to reopen the discussion after he remembered Brown's prior engagement:

> "The other members looked at me as if I had taken leave of my senses. "What difference does it make that Brown can't come," one said, "since in any case we decided on Adams?" I was amazed. All the members were eminent theorists and mathematical economists, thoroughly familiar with the nuances of the Nash model. Not long before, the very member who had spoken up had roundly criticized IIA in the discussion period following a talk. I thought that perhaps he had overlooked the connection, and said that I was glad that in the interim, he had changed his mind about IIA. Everybody laughed appreciatively, as if I had made a good joke, and we all went off to lunch. The subject was never reopened, and Adams was invited." (Aumann, 1985: 603-604)

The reason the committee thought Aumann had taken leave of his senses is that the committee's choice of Adams is likely to have been based on some standard or set of standards on which all committee members agreed, and according to which Adams was "best" among the three initial candidates. Obviously, if the committee agreed that Adams was better than either Brown or Clark, they also agreed that Adams was better than Clark alone.

The reasoning underlying Aumann's example is implicit also in the justification Nash (1950) himself gives when introducing the IIA axiom. Nash argues that if two rational individuals agree that a given point in a set is a fair bargain, then it should only be easier for them to agree on the same point in a subset Q of P. As Peters and Wakker (1991) point out, implicit in Nash's calling the latter agreement "of lesser restrictiveness" is that bargainers do more than simply agree on some point in the utility possibility set, P. They somehow-axiomatic models do not concern themselves with the how-agree on a collective preference relation over all points in the set, according to which the point chosen is "best." It should indeed then be only easier to choose from a subset Q of P, since there are fewer comparisons to be made.

Formally, this point can be made precise in the following way. Let an n-person bargaining problem be defined by a compact, convex payoff space $P \in \Re^n$ with vector of disagreement payoffs t. Let B denote the set of all bargaining problems. A bargaining solution is a choice function $\bar{u} : B \rightarrow \Re^n$ that assigns to each bargaining problem in B a unique element of P. Suppose now that there exists a (binary) collective preference relation, denoted by $\succeq$, such that for all bargaining problems (P, t) in B

$$\bar{u}(P, t) = \{u \in P : u \succeq v \text{ for every } v \in P\}.$$

Then the preference relation $\succeq$ is said to "represent" the choice function, $\bar{u}$. It is easy to see that the existence of such a relation implies IIA: If the point $\bar{u}(P, t)$ is weakly preferred to all other points in P, then it is clearly also weakly preferred to all other points in any subset Q of P. Peters and Wakker (1991) show that if the choice function is single-valued and if its domain, B, is intersection-closed,[10] the converse also holds: Any single-valued choice function satisfying IIA can be represented by a binary preference relation, $\succeq$. They also establish conditions for the existence of a real-valued collective utility function f on $\Re^n$, that the choice function, $\bar{u}$, maximizes. The following axiom turns out to be important:

- **Continuity (CONT):** For every sequence $(P^0, t), (P^1, t), (P^2, t), (P^3, t), ..., \in B$, if $P^i \rightarrow P$ in the Hausdorff metric, then $\bar{u}(P^i, t) \rightarrow \bar{u}(P, t)$.

[10] B is intersection-closed if, for any payoff spaces P and Q with common disagreement point t, whenever $(P, t) \in B$ and $(Q, t) \in B$, it is also the case that $(P \cap Q, t) \in B$.

This CONT axiom, which is quite weak, essentially prohibits bargaining outcomes from "jumping" discretely when the bargaining set changes slightly.

Peters and Wakker establish that if the outcomes of bargaining between n players are consistent with both the PO and CONT axioms, then the IIA axiom is a necessary condition for those bargaining outcomes to be consistent with the maximization of *any* real-valued function f on $\Re^n$. More precisely, if a given group of n players is presented with a succession of two or more utility-possibility sets $P^1, P^2, ..., \in \Re^n$ with the same disagreement point t, and the respective bargaining outcomes $\bar{u}\left(P^i, t\right)$ satisfy PO and CONT, then those outcomes will maximize a given function $f : \Re^n \rightarrow \Re$ over the P^i only if they satisfy IIA as well. Peters and Wakker also establish that if $n = 2$, the PO, CONT, and IIA axioms are jointly sufficient for the existence of such a function f. For $n > 2$, a stronger condition than IIA is required (again in combination with PO and CONT) for sufficiency.

Based on the above results, the IIA axiom is thereby shown to be directly responsible for the remarkable property of the Nash bargaining solution that it can be calculated as the maximum of a function of the players' utilities. Any solution concept that does not satisfy IIA will not have this convenient property and will, as a result, be much more difficult to use in both theoretical and empirical work. It is important to be aware, however, that this convenience is available only if the IIA axiom is itself justified. Is it reasonable to presume that the players in a bargaining game do not just somehow arrive at a decision, but rather maximize a collective preference relation?

By the "Methodological Individualism" doctrine, such a presumption is reasonable if the ultimate outcome of the bargaining is not a collective decision in the first place, but dictated by a single decision maker. This is why Story 2 is an even more straightforward example of a context in which the IIA axiom is sensible: If a 2% price rise above inflation maximizes the government agency's objective function (which one can think of as balancing, governance-function style, the interest of consumers in low prices against the interests of the public utility), it still does so when the non-binding legal constraint on the price rise is moved but remains non-binding.

In Stories 3 and 4, on the other hand, the parties in the collective decision making problem quite obviously do not have a collective preference ordering according to which the outcome is best. These stories are examples of contexts in which collective decisions are unlikely to be consistent with the IIA axiom. In the wage-bargaining story, the imposition of wage controls in the second scenario is likely to weaken the union's bargaining position relative to that of the employer, resulting in a lower agreed wage increase. In the American presidential-election story, it is at least conceivable that if Perot had stayed out of the race, Bush would have attracted enough Perot supporters to win the election.

The first of two lessons that might be drawn from these stories is that the IIA axiom is reasonable in contexts in which players either act as if they are a single decision maker, or in which there effectively is only a single decision maker; contexts, in other words, that reduce collective choice to simple constrained optimization problems. One might ask if it is ever appropriate to formulate any political economy problem as a single-person decision framework. As noted in Chapter 1, a single-person framework is ubiquitous in the political economy literature; often the policy maker is modeled as acting as a Stackelberg leader, who chooses the optimal level of a policy instrument taking as given the voters' and/or the interest groups' reaction functions.[11]

Just as decision procedures can be introduced into Aumann's story that yield violations of the IIA axiom, decision procedures can also be introduced into wage-bargaining examples that yield outcomes consistent with the IIA axiom. Four such procedures have been developed in the literature. Each of these procedures is a strategic game that yields an outcome consistent not only with the IIA axiom but with all four axioms that characterize the Nash bargaining solution. The games, therefore, implement this solution non-cooperatively.

The first procedure has emerged from the Nash demand game, introduced by Nash (1953) precisely as an example of a non-cooperative implementation of his own solution concept.

[11] For example see de Gorter and Tsur (1991), Hillman (1982), Laffont and Tirole (1991), Peltzman (1976), Swinnen (1994).

The game requires both players to make simultaneous demands y_1 and y_2, without being sure where the boundary of the bargaining set, P, lies. If it turns out that point $y = (y_1, y_2)$ lies within that boundary, so that the demands can be simultaneously satisfied, the players each achieve what they demanded. If not, they receive the disagreement payoff. A continuous function $p\,(y)$ represents the players' common beliefs about the probability that a given point y lies in P. In Nash's specification, this probability equals one everywhere on P, and tapers off rapidly to 0 as y moves away from the set. Nash shows that a sequence of non-cooperative equilibrium demands of this game converges to the Nash bargaining solution as the players' uncertainty over the bargaining set vanishes.

A qualitatively equivalent variant of Nash's specification has been introduced by Binmore (1987). Binmore first normalizes the disagreement point, t, to the origin. This normalization is without loss of generality and, for ease of exposition, it is retained. As in the Nash demand game, it is specified that players expect to receive their simultaneous demands y_1 and y_2 with probability $p\,(y)$ and their disagreement payoffs of 0 with probability $1 - p\,(y)$. To distinguish the Binmore specification from Nash it is assumed that $p\,(y) = 1$ for any $y \in \rho P$, where $0 < \rho < 1$, and that the function tapers off to become 0 exactly on the boundary of P. Also, $p\,(y) = 0$ for any $y \notin P$. For a pair of demands (y_1^*, y_2^*) to be a non-cooperative Nash equilibrium of the game, it must be the case that

$$y_1^* p\,(y_1^*, y_2^*) \geq y_1 p\,(y_1, y_2^*) \tag{2.3}$$

$$y_2^* p\,(y_1^*, y_2^*) \geq y_2 p\,(y_1^*, y_2) \tag{2.4}$$

for any other demands y_1 and y_2 that the players might make. The point $y^* = (y_1^*, y_2^*)$ at which the function M defined by

$$M(u_1, u_2) = u_1 u_2 p(u_1, u_2) \tag{2.5}$$

achieves its maximum value is such a Nash equilibrium.

Whatever the exact form of the function p, this point must lie in the lens-shaped area shown in Figure 2.1 enclosed by the boundary of P and by the level set of the function $u_1 u_2$ tangent to the boundary of ρP. This is because at the tangency point z of this level set with the boundary of ρP the value of M is just $z_1 z_2$, and outside of the lens-shaped area M is always lower. Letting uncertainty about the location of the boundary go to 0 can be modeled as letting ρ go to 1. The lens-shaped area then becomes squeezed between successively higher level sets and the boundary of P, until in the limit it shrinks to a single point, the Nash bargaining solution.

As the uncertainty goes to 0, the players in Nash's game behave increasingly *as if* they maximize a collective utility function $u_1 u_2$ on P. It follows that if the employer and the union in our wage story bargained in the manner specified by Nash, and if they were almost certain about the location of the boundary of their utility possibilities set, they would agree on a wage rise of 9.9%-the wage rise maximizing the product of their utility gains-regardless of any non-binding wage controls.

Under the conventional interpretation of the Nash program, any strategic game that implements his (Nash) bargaining solution is seen as providing an independent justification for this solution, and thereby for the axioms from which the solution was originally deduced. However, this implementation equivalence only suggests a class of collective decision making *contexts* to which the Nash bargaining solution can reasonably be applied. How significant is this class? Some uncertainty about the exact location or shape of the boundary of the bargaining set is undoubtedly a feature of most real-world bargaining contexts. On the other hand, it is difficult to think of a realistic context in which players not only fully agree on the

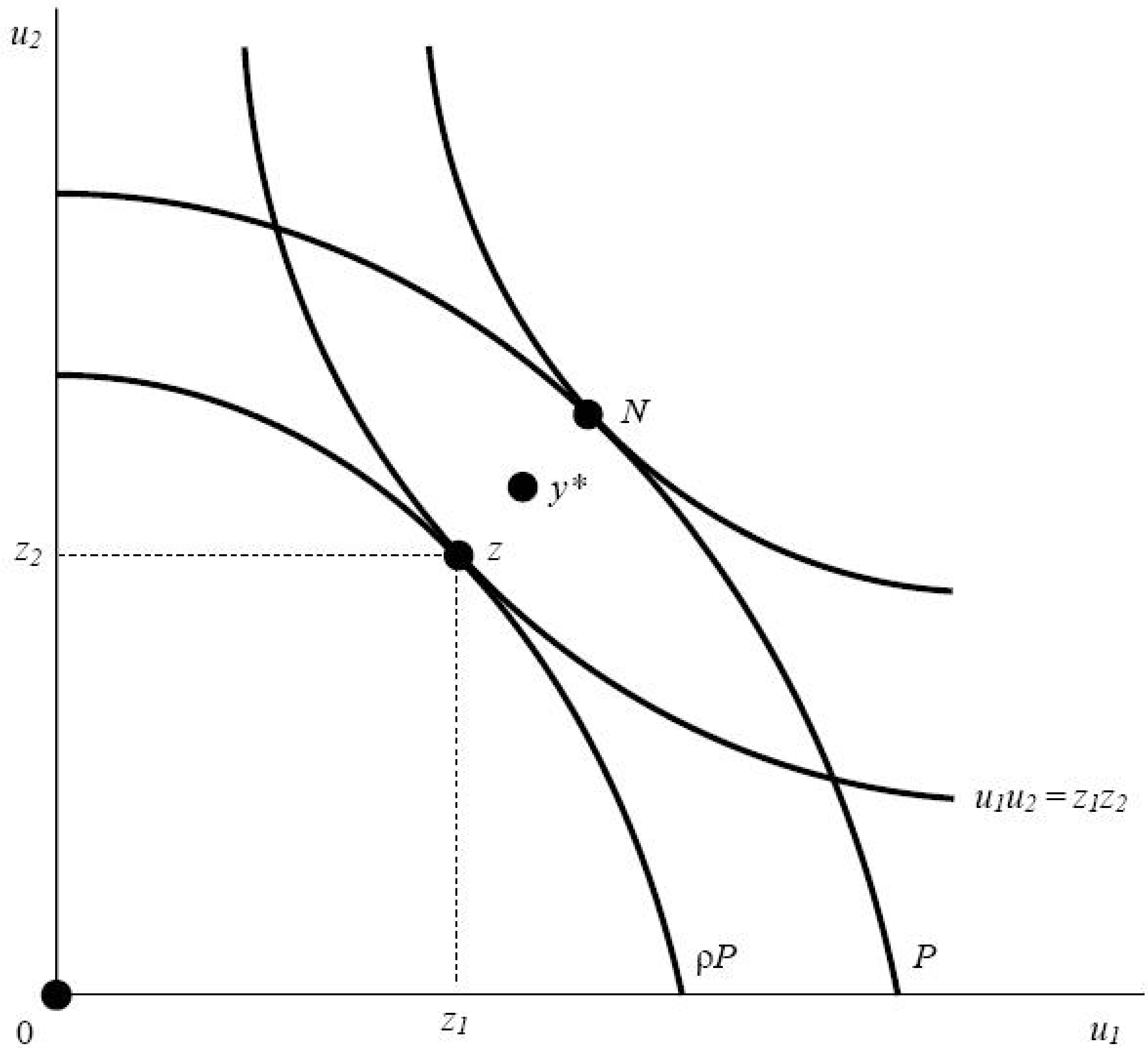

Figure 2.1: Bounds of the solution, y^*, of the Nash demand game

form of that uncertainty but also have to make proposals in the manner of the Nash demand game, with guaranteed permanent breakdown of negotiations if they are too far wrong. Binmore (1987) shows that it is not impossible to dream up such a context. He suggests a context in which the two players place their demands in sealed envelopes which are then passed on to a referee. The referee opens the envelopes after time t, and determines whether the demands are compatible. If the players discount time at the same rate and have the same probability distribution function for the random time t at which the envelop is opened, the game is formally equivalent to the Nash demand game. Clearly, however, Binmore's suggestion only barely qualifies as a counter-argument to the criticism by Luce and Raiffa (1957) that the Nash demand game is a "completely artificial mathematical 'escape'" from the non-uniqueness of non-cooperative Nash equilibria in the game under full certainty.

Luce and Raiffa's criticism applies with equal, if not more, force to a variant on the Nash demand game introduced by Anbar and Kalai (1978) that also implements the Nash bargaining solution. This game also requires players to make simultaneous demands, and also incorporates an element of uncertainty. Here, however, the players are uncertain, not about the size of the bargaining set but about each other's demands. Given any pair of demands y_1 and y_2, they receive their demands if $y \in P$ and the disagreement payoff if otherwise. Anbar and Kalai normalize the disagreement point t to the origin, and also normalize each player i's maximum feasible payoff (each party's "ideal point"), defined by

$$h_i\left(P,t\right) \equiv \max\left\{u_i : u \in P, u \geq t\right\}, \qquad i = (1,2) \tag{2.6}$$

to unity. With these normalizations, the set of feasible demands for each player consists of the interval $[0,1]$. Anbar and Kalai then show that if player i expects player j to demand some y_j according to the *uniform* distribution on $[0,1]$, then player i's expected utility is maximized if he himself demands the payoff $\bar{u}_i\left(P,t\right)$ associated with the Nash bargaining solution. This is most readily seen if we express the upper-right boundary of P (denoted by H) as a function of u_1, by defining the function

$$\phi(u_1) \equiv \max\{u_2 : (u_1, u_2) \in P\}. \tag{2.7}$$

Player 1's expected utility from demanding y_1 is then $y_1 \Pr[y_2 \leq \phi(y_1)]$. Given his uniform prior on player 2's demand, y_2, this expression in turn equals $y_1 \phi(y_1)$. This expected utility is maximized at $y_1 = \bar{u}_1(P, t)$, i.e., at a demand equal to the first coordinate of the Nash bargaining solution. If player 2 has the same prior on player 1's demand, then the outcome of the game will coincide with the Nash bargaining solution. Clearly, however, this result is strongly dependent on the very specific form of the players' beliefs.

Yet another strategic game that implements the Nash bargaining solution was developed by Zeuthen (1930) – long before Nash even formulated his solution – to analyze precisely the wage-bargaining situations specified above. It was later formalized and extended to a more general two-person bargaining situation by Harsanyi (1956) and is now generally referred to as the Harsanyi-Zeuthen bargaining procedure. The procedure is essentially just a rule stipulating which player should make a concession if the players make incompatible "proposals" in utility space, where a proposal consists not just of a utility demand for oneself, but also a utility offer to the other player. Suppose, for example, that player 1 proposes the point $x = (x_1, x_2)$ and player 2 proposes $y = (y_1, y_2)$, where $x_1 > y_1$ and $y_2 > x_2$. In contrast to the Nash demand game or the one-shot bargaining game defined by Anbar and Kalai, such incompatible proposals do not immediately imply that players receive the disagreement payoff (which we again normalize to zero, for simplicity). Instead, the procedure requires players to compare each other's readiness to *risk* disagreement, which one might call their "boldness."[12] This is defined as the maximum probability of complete

[12]By analogy to a concept by the same name defined by Aumann and Kurz (1977). Both concepts are equivalent when x and y are very close to each other.

breakdown of negotiations that players are willing to incur in holding out for their own demand, rather than giving in completely to their opponent's demand. Given that complete breakdown of negotiations yields both players zero payoff, this maximum probability p for player 1 is given by

$$(1-p)\,x_1 = y_1 \tag{2.8}$$

and the maximum probability p' for player 2 is given by

$$x_2 = (1-p')\,y_2 \tag{2.9}$$

If it is then found that $p > p'$, say, or equivalently,

$$\frac{x_1 - y_1}{x_1} > \frac{y_2 - x_2}{y_2} \tag{2.10}$$

then it is player 2 who is least bold. That player is then required to make a concession large enough to reverse the inequality (if possible), and the procedure is then applied anew. It is easy to see why repeated application of the procedure forces the players' proposals to ultimately converge to the Nash bargaining solution. Rewriting the above inequality (2.10) as

$$x_1 x_2 > y_1 y_2 \tag{2.11}$$

shows that the procedure effectively requires players to compare which level set of the function $u_1 u_2$ their respective proposal lies on, and that it is always the player whose proposal lies on the lower level set who is required to concede. As a result, the sequence of proposals must converge to the tangency of the highest attainable level set with the bargaining set P, which occurs at the Nash bargaining solution.

How significant is the class of contexts that share the essential features of the Harsanyi-Zeuthen model? As with the previous games, the answer to this question must depend on how plausible one considers the model to be. On this, Bishop argues:

> "I do not share Harsanyi's insistence (which is even more emphatic than Zeuthen's) that bargaining behavior must be in accordance with this model if it is to escape the charge of being "irrational." Even when the bargaining situation is essentially static and nonrecurrent, and even when the conditions of requisite knowledge are fulfilled (unlikely as this may be in practice), there still seem to me to be ample grounds for resisting the Zeuthen-Harsanyi prescription.
>
> For one thing, it should be noticed that the question of who concedes a little is answered with reference to the expected utilities involved in either making or not making a different and typically much larger concession, all of the way to the other bargainer's current demand. On the other hand, that large hypothetical concession is really not seriously considered at all, except as a means of determining some smaller actual concession. Furthermore, the subjective probabilities of conflict, which motivate the concession, are mechanical ones that the bargainers must arrive at in a uniquely specified way, rather than truly subjective estimates that they might make in any way that they happened to see fit. Thus, even though the theory is ostensibly rooted in a process of successive concessions, as in many instances of realistic bargaining, it really implies a foreordained outcome that the bargainers might just as well establish without any play acting. These aspects of the theory are merely the more obvious ones upon which skepticism

may appropriately be focused." (Bishop, 1964: 412)

A fourth strategic game that can also implement the Nash bargaining solution, at least approximately, is the famous two-person bargaining game analyzed by Rubinstein (1982). In this game players make alternating offers over the division of a pie that shrinks over time because of costs of delay. It was proved independently by Binmore (1987), Maclennan (1982), and Moulin (1982) that the non-cooperative equilibria of this game – there are two, depending on which player makes the first offer – converge to the Nash bargaining solution as the time delay between offers goes to zero.

Figure 2.2 illustrates the equilibria of the Rubinstein game when both players discount payoffs from future rounds at a rate $\delta = \frac{1}{3}$, so that the pie shrinks between rounds at that rate. (A "round" consists of an offer by one player and a decision by the other player to either accept or reject that offer.) In subgame-perfect equilibria, when player 1 is the first to make a demand, player 2 must be offered at least δ times his maximum demand in the second round. Player 1 will want to offer no more than that, so we have

(2.12) $$x_2 = y_2 = \delta y_2.$$

In turn, player 2's maximum demand in the second round must offer player 1 at least δ times her maximum demand in the third round. Since the subgame then reached is identical to the game as a whole – given that the game has an infinite horizon – and since, again, player 2 will want to offer player 1 no more than the minimum required to induce her to accept his demand, we have

(2.13) $$\delta y_1 = y_1 = x_1 = \delta^2 x_1$$

Combining equations 2.12 and 2.13 yields

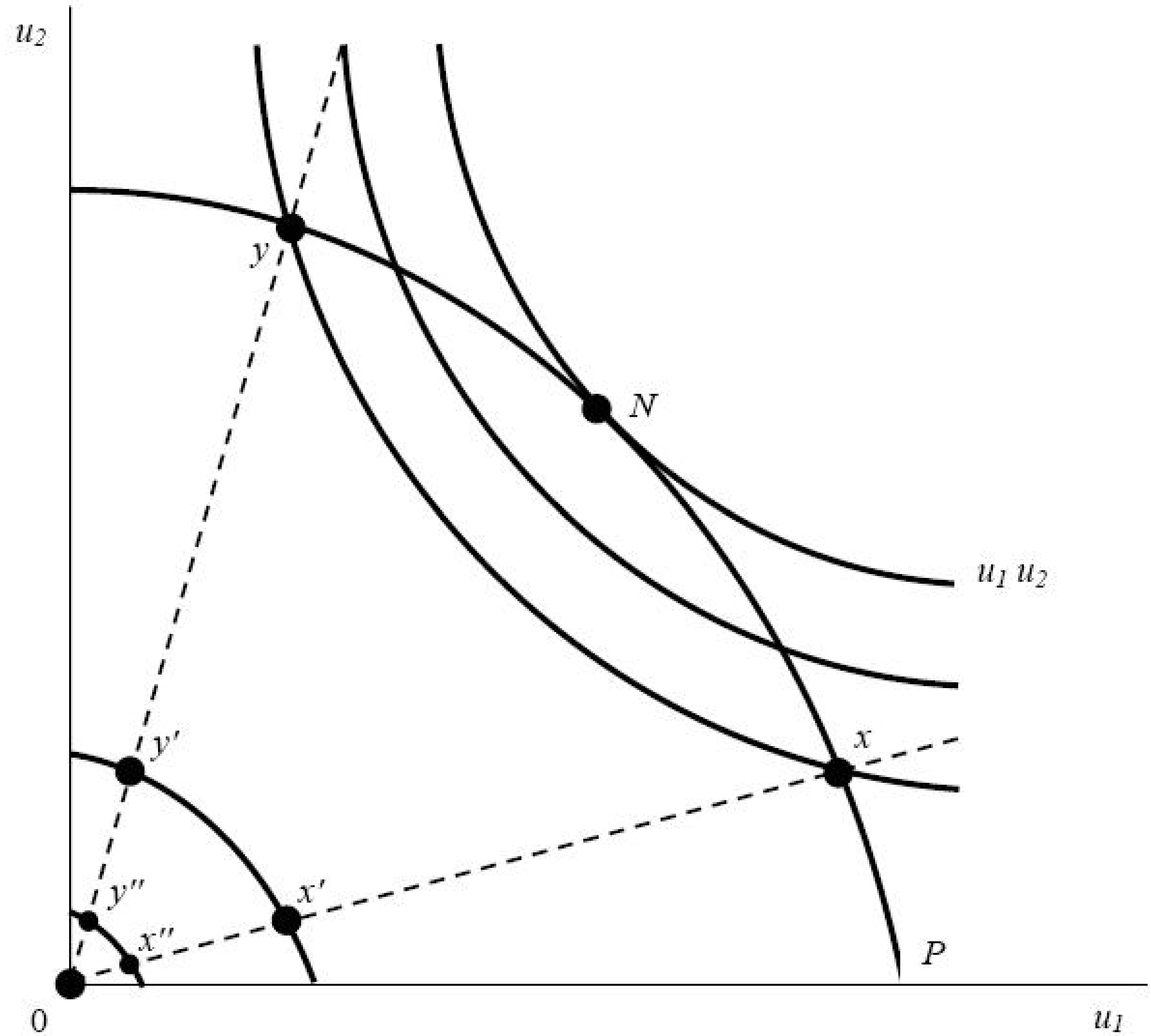

Figure 2.2: The equilibria of the Rubinstein game when $\delta = 1/3$

(2.14) $$\delta x_1 y_2 = x_1 x_2 = y_1 y_2,$$

showing that the two equilibria must lie on the same level set of the function . The two equilibria must also lie on the Pareto frontier of P: The players will not leave any pie undivided. Consider now any pair of points such as a and b in Figure 2.3 that lie on the Pareto frontier and on a higher level set than x and y. For such points, the equation

(2.15) $$\delta' a_1 b_2 = a_1 a_2 = b_1 b_2$$

must hold by analogy to the equation directly above. Although the product $a_1 a_2 = b_1 b_2$ is strictly greater than $x_1 x_2 = y_1 y_2$ for such points, the product $a_1 b_2$ can be no larger than $x_1 y_2$. It must therefore be the case that $\delta' > \delta$. This shows that reducing the delay between rounds and thus increasing δ moves the equilibria of Rubinstein's game to higher level sets of the function $u_1 u_2$. It is also evident from the equations and from Figure 2.3 that when δ goes to unity, implying that the delay between offers goes to zero, the equilibria converge to the Nash bargaining solution. The result thus shows that if it is costs of delay between offers that spur bargainers to come to an agreement – there must be some such spur, or bargaining could go on forever – it is possible that they will agree to a division that lies close to the Nash bargaining solution. Paradoxically, the agreement will be closer to that solution the less important the costs of delay in fact are, either because the players do not discount the future much at all or because the players foresee a rapid exchange of offers and counter-offers if the bargaining process were to advance beyond the first round (which in equilibrium it does not). That in such cases the bargaining outcomes will be close-to-consistent with IIA follows because, for any value of δ, only the two line segments on the bargaining set joining the disagreement point t (the origin) with the two equilibrium points are "relevant" (see

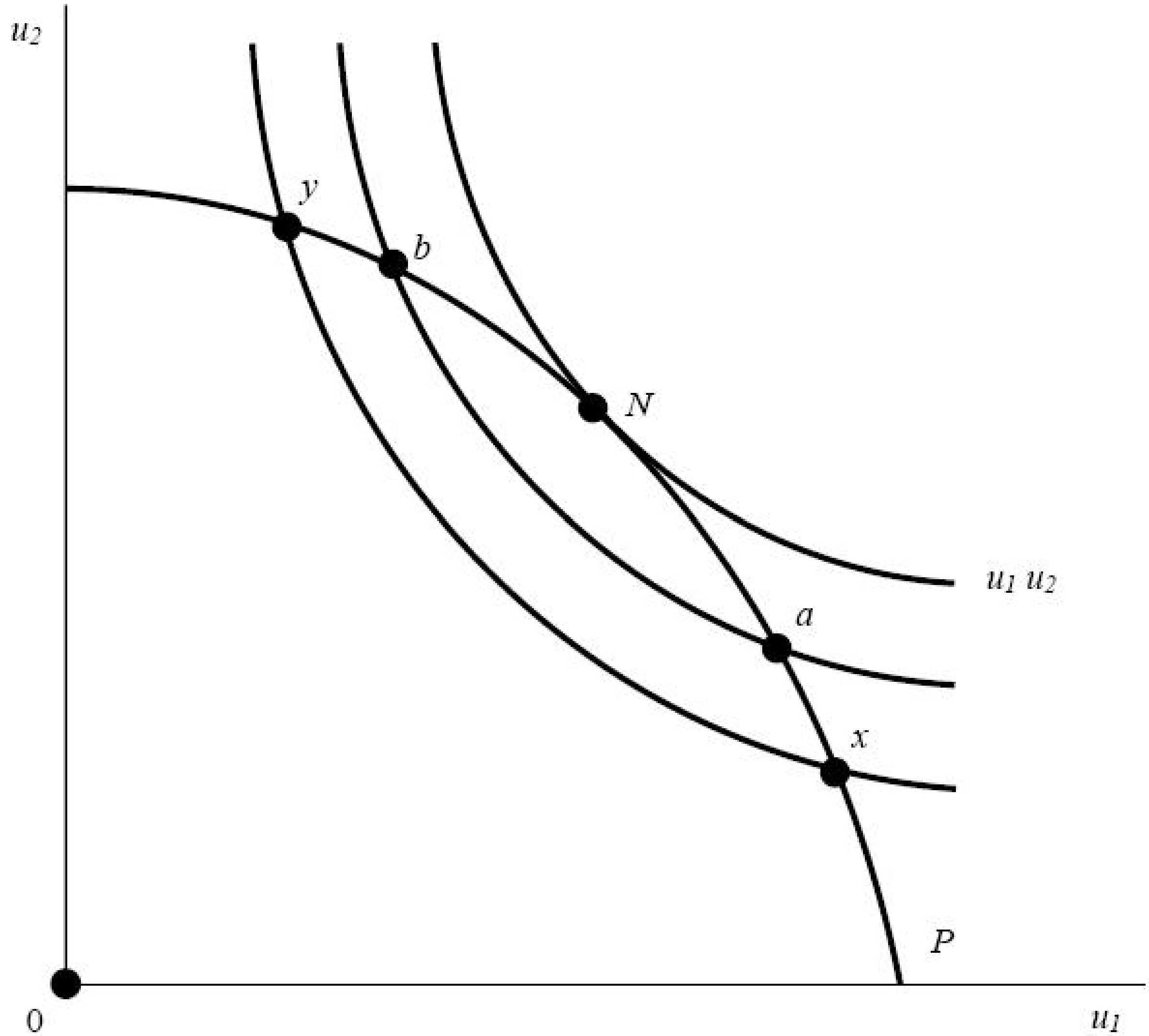

Figure 2.3: The equilibria of the Rubinstein game for different values of δ

Figure 2.2). Removing any part of the set outside of these line segments leaves the outcome unaffected. As δ goes to 1, the two line segments converge to a single line segment joining the disagreement point with the Nash bargaining solution, so that the outcome becomes independent of all alternatives not on that line.[13]

In terms of suggesting a significant class of collective decision making contexts to which the Nash bargaining solution can reasonably be applied as a rough prediction of the outcome, the Rubinstein game looks considerably more promising than any of the foregoing games. The class of such contexts can be broadened further by relaxing some of the model's restrictive assumptions. For example, in contrast to the foregoing games, the Rubinstein game is not restricted to two-person bargaining contexts. Krishna and Serrano (1990) have shown that an n-person version of the Rubinstein game exists that implements the n-person Nash bargaining solution as δ goes to 1. The only feature (with no n-person analog) added in this game is that in any round players are allowed to "walk away" from the bargaining table with the share offered to them by the player whose turn it was to propose an agreement, even if there are other players who reject that same agreement. The remaining players then continue bargaining over the left-over pie.

Another assumption of the Rubinstein model that can be relaxed is that players have equal discount rates. This turns out to be necessary only for consistency of the limit solution with the SYM axiom. If the assumption is dropped, the limit of the Rubinstein game converges to the so-called "asymmetric" Nash bargaining solution, which maximizes the function $u_1^{\alpha} u_2^{(1-\alpha)}$ for some value of $\alpha \in (0,1)$. This solution still satisfies the PO, LINV, and IIA axioms. A last assumption of the model that can be relaxed, or rather reinterpreted, is the infinite-horizon assumption. Bargainers need not literally expect bargaining to go on forever if they keep rejecting each other's offers. It is a well-known fact that discounting future payoffs in a multi-round game at a fixed rate δ is equivalent to assuming a fixed

[13]Note that this line segment is the minimal subset of P to which the IIA axiom applies. Removing alternatives in the interior of the segment is not considered by the theory, which covers only the domain B of bargaining problems with convex bargaining sets.

probability $p = (1 - \delta)$ that the game will end after any given round (with zero payoffs). This implies that the pie may also shrink between rounds because players (credibly) threaten complete breakdown of the negotiations with probability p whenever their current offer is rejected. Such threats certainly appear to be a feature of many real-world bargaining contexts. So are costs of dealing between offers, whether or not imposed deliberately.

Nevertheless, despite the plausibility of many of the Rubinstein model's features, any appeal to the model in support of either the IIA axiom or the Nash bargaining solution runs into serious problems. First of all, there remains the disturbing paradox that players in the Rubinstein model only act as if they maximize a collective utility function when costs of delay and/or threats of complete breakdown simultaneously shrink to insignificance *and* continue to drive the bargaining process. In addition, the essential feature of the model that players have either an infinite or an indeterminate horizon is hard to recognize in political-economic contexts. Many, if not most, such contexts have a horizon that is both finite and determinate: there is a clear deadline before which agreement must be reached. Even when there is no such deadline, costs of delay, far from driving the bargaining process, appear to at best be incidental to it.

As for threats of complete breakdown, these are often simply not credible in political-economic contexts. In most situations where a candidate must be selected, for example, it is clear to all those involved in the selection process that somehow some candidate will eventually emerge as winner; if all else fails, perhaps only by the luck of a draw. What is not clear – particularly of course if lots are to be drawn – is who that candidate will be. Similarly, in the current legislative debates on economic and social policy reform in many countries, it is understood by all sides that some reform plan will eventually be negotiated. Maintaining the status quo is simply not a realistic option in the existing political climate.

2.4 Conclusion

In this chapter the basic, two-person Nash (1950) axiomatic bargaining solution has been presented. Nash showed that the PO, LINV, SYM, and IIA axioms imply a unique bargaining solution. The remarkable simplicity of the Nash bargaining solution, in particular the fact that it can be calculated as the point in the bargaining set that maximizes the product of the players' utility gains relative to a fixed disagreement point, has facilitated its use in both theoretical and empirical work.

The work of Peters and Wakker demonstrates that the "independence of irrelevant alternatives" (IIA) axiom is directly responsible for the convenient property that the Nash bargaining solution can be calculated as the maximum of a function. Nash's axiomatic model posits IIA as a primitive; it is, therefore, appropriate to apply this model to specific contexts only if the validity of the axiom is self-evident in these contexts. It was argued in this chapter that such contexts are limited to those in which players act *as if* there effectively is only a single decision maker-contexts, in other words, that reduce to simple constrained-optimization problems.

The PO, LINV, SYM, and IIA axioms not only imply the Nash bargaining solution, they are also implied by the solution. Consequently, any non-axiomatic justification of the Nash bargaining solution can be interpreted as providing indirect support for the reasonableness of the set of axioms. Nash (1951, 1953) was the first to suggest using strategic models of bargaining to provide such indirect support. He showed that his bargaining solution could indeed be derived, not just from a set of axioms, but also as the approximate non-cooperative equilibrium of a particular strategic bargaining model. This model, and several other strategic models developed since, yield the Nash bargaining solution – and hence the IIA property – as a consequence of their primitives; and it is perfectly possible that these primitives are quite adequate stylizations of political-economic contexts that involve "real" bargaining.