Journal of Risk and Uncertainty, 2: 353-365 (1989)

Risk Aversion and the Relationship Between Nash's Solution and Subgame Perfect Equilibrium of Sequential Bargaining

ALVIN E. ROTH*
Department of Economics, University of Pittsburgh, Pittsburgh, PA 15260

Key words: bargaining, axiomatic models, strategic models, risk aversion

Abstract

This article presents some new, intuitive derivations of several results in the bargaining literature. These new derivations clarify the relationships among these results and allow them to be understood in a unified way. These results concern the way in which the risk posture of the bargainers affects the outcome of bargaining as predicted by Nash's (axiomatic) solution of a static bargaining model (Nash, 1950) and by the subgame perfect equilibrium of the infinite horizon sequential bargaining game analyzed by Rubinstein (1982). The analogous, experimentally testable predictions for finite horizon sequential bargaining games are also presented.

This article has two primary purposes. The first is to present some new derivations of several results in the bargaining literature. These deviations clarify the relationships among these results and allow them to be understood in a unified and intuitive way. These results concern the way in which the risk posture of the bargainers affects the outcome of bargaining as predicted by Nash's (axiomatic) solution of a static bargaining model (Nash, 1950) and by the subgame perfect equilibrium of the infinite horizon sequential bargaining game analyzed by Rubinstein (1982). The second purpose of the article is to derive the similar predictions for finite horizon games, and to consider how these might be experimentally tested.

The three results from the literature, informally stated, are that

1. Nash's solution predicts that risk aversion is disadvantageous in bargaining (Roth 1979; Kihlstrom, Roth, and Schmeidler, 1981).
2. The subgame perfect equilibrium predicts that risk aversion is disadvantageous in bargaining (Roth, 1985).

*This work has been partially supported by a grant from the Russell Sage Foundation. I have also received helpful comments from Ken Binmore, Ariel Rubinstein, Asher Wolinsky, and (especially) Shmuel Zamir.

3. In the limit as the costliness of waiting an additional period goes to zero, the subgame perfect equilibrium of the sequential bargaining model converges to Nash's solution of the static model (or to an asymmetric Nash solution, when the bargainers have different costs) (Binmore, 1987a; MacLennan, 1982; Moulin, 1982).

It will be seen that each of these results is a simple consequence of the relationship of the risk postures of the bargainers at Nash's solution of the static game, and at the perfect equilibrium of the sequential game. Indeed, the analysis will emphasize the similarity of the two solutions in this respect: Nash's solution selects the unique point at which the bargainers have equal willingness to bear a certain kind of infinitesimal risk, and perfect equilibrium of the sequential game selects the unique pair of proposals with respect to which the bargainers have equal willingness to bear the same sort of risk. (This latter point of view will also suggest a new proof of Rubinstein's result that a unique stationary perfect equilibrium exists.[1])

It should again be emphasized that, while some of the subsidiary results presented in deriving these theorems in sections 2–4 are new, these three results themselves are already known. However, the new proofs establish connections among these results that had not been at all apparent. It is these new connections, which highlight the central role played by the risk posture of the bargainers, that are the main content of sections 2–4. Section 5 considers finite horizon sequential games, in connection with the issues that arise in designing experimental tests of these predictions.

1. The basic models

Since the purpose of this article is to establish connections between different results, a very simple model will be presented that allows these connections to be made most clearly. The conclusion of section 4 will briefly consider more general models.

To this end, consider bargaining over the division of a fixed quantity Q of a single commodity (e.g., money), by two bargainers who have expected utility functions u_1 and u_2, defined on the real numbers. The initial wealth of each bargainer is normalized to be 0, and similarly $u_i(0) = 0$ for $i = 1,2$. Any nonnegative division is feasible if both bargainers agree; otherwise, they each receive 0. The utility functions u_i are increasing and concave.

Then the static model to which Nash's solution applies is represented by the set of feasible utility payoffs and disagreement utility given by the pair (S,d), where

$$S \equiv \{(u_1(c_1),u_2(c_2)) \mid c_1 + c_2 \leqslant Q,\ c_1,c_2 \geqslant 0\},$$

$$d = (u_1(0),u_2(0)) = (0,0).$$

In the sequential bargaining model, time is divided into periods, and in odd-numbered periods t (starting at an initial period $t = 1$) player 1 may propose to player 2 any division $(c_1, Q - c_1)$. If player 2 accepts this proposal, then the game ends and player 1 receives a utility of $(\delta_1)^{(t-1)}u_1(c_1)$ and player 2 receives a utility $(\delta_2)^{(t-1)}u_2(Q - c_1)$, where δ_i is a number between 0 and 1 reflecting player i's cost of delay. If player 2 does not accept the offer, then the game proceeds to period $t + 1$, and the roles of the two players are reversed. Following Binmore, Rubinstein, and Wolinsky (1986), we consider a model in which the cost of delay arises from a probability q that each period will be the last, so that the bargainers have equal discount factors $\delta_1 = \delta_2 = (1 - q)$. That is, after any rejection, and before a new proposal can be made, there is a probability q $(0 < q < 1)$ that the game will end and that each bargainer will receive 0 dollars. So if the bargainers adopt strategies that taken together have the effect that all offers made before time t will be rejected, and at time t the division (c_1,c_2) will be accepted (if the game has not ended before time t), the expected utility of bargainer i is $(1 - q)^{(t-1)}u_i(c_i)$. Strategies in which no offer will ever be accepted give each bargainer a utility of 0.

2. Nash's solution of the static model

For pairs (S,d) such that S is a compact convex set containing d and having some $x > d$ in S, Nash's (1950) solution to the bargaining problem is the point $F(S,d) = z$ in S that maximizes the product $(z_1 - d_1)(z_2 - d_2)$. Since we are here taking $d = (0,0)$, Nash's solution is the point z in S with maximum z_1z_2. That is, Nash's solution to the game considered here is the utility payoff vector $\mathbf{z}$ corresponding to the division $(a_1,Q - a_1)$ that maximizes the product $u_1(a_1)u_2(Q - a_1)$.

A notion of risk aversion that captures the information needed to determine Nash's solution here is the concept of *boldness* introduced by Aumann and Kurz (1977a,b). Player i's boldness with respect to a division (c_1,c_2) is defined to be the quantity

$$b_i(c_i) = u_i'(c_i)/(u_i(c_i) - u_i(0)) = u_i'(c_i)/u_i(c_i).$$

To see what is going on, suppose player i has been offered c_i, and now considers a gamble that risks his entire gain c_i against the possibility of a small additional gain h. The maximum probability $\bar{q}_i$ of receiving 0 for which player i is willing to accept the gamble can be considered a measure of his boldness. It turns out that $b_i(c_i)$ equals the maximum probability of getting 0 that player i will accept, per dollar of additional gains, for very small potential gains. That is, $b_i(c_i)$ equals the limit as $h \to 0$ of $\bar{q}_i/h$.

To see this, observe that $\bar{q}_i$ is the probability such that $u_i(c_i) = \bar{q}_i u_i(0) + (1 - \bar{q}_i)u_i(c_i + h) = (1 - \bar{q}_i)u_i(c_i + h)$. Consequently,

$$\frac{\bar{q}_i}{h} = \frac{[u_i(c_i + h) - u_i(c_i)]/h}{u_i(c_i + h)},$$

and so

$$\lim_{h\to 0} \frac{\bar{q}_i}{h} = \frac{u_i'(c_i)}{u_i(c_i)} = b_i(c_i).$$

Aumann and Kurz observed that Nash's solution can be characterized as selecting the point z in S at which the players are equally bold.

Lemma 1. $F(S,d) = z = (u_1(a_1^*), u_2(a_2^*))$ such that $a_1^* + a_2^* = Q$ and $b_1(a_1^*) = b_2(a_2^*)$.

Proof: Nash's solution picks the point that maximizes the product $A(c_1) = u_1(c_1)u_2(Q - c_1)$. Setting $dA/dc_1 = 0$ yields $-u_2'(Q - a_1^*)u_1(a_1^*) + u_1'(a_1^*)u_2(Q - a_1^*) = 0$, which gives the required result.

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,

Lemma 2. If $x < y$, $b_i(x) > b_i(y)$.

Proof: $db_i(c_i)/dc_i = [u_i(c_i)u_i''(c_i) - (u_i'(c_i))^2]/[u_i(c_i)]^2 < 0$, since the denominator is always positive, and the numerator negative.

We can use the characterization of Nash's solution in terms of the bargainers' boldness to give a proof of the first result discussed in the introduction, that is different from the proofs of Roth (1979) and Kihlstrom, Roth, and Schmeidler (1981). Let player $\hat{2}$ be a more risk-averse individual than player 2, i.e., such that $u_{\hat{2}}(a) = k(u_2(a))$ for all $a \geqslant 0$, where k is an increasing, concave function (see Yaari, 1969; Kihlstrom and Mirman, 1974; Roth, 1979). (If k is strictly concave, player $\hat{2}$ will be said to be strictly more risk-averse than player 2.) We begin by showing that player $\hat{2}$ is everywhere less bold than player 2.

Lemma 3. For any $c > 0$, $b_{\hat{2}}(c) \leqslant b_2(c)$, with strict inequality if player $\hat{2}$ is strictly more risk-averse than player 2.

Proof: Without loss of generality, choose a normalization of $u_{\hat{2}}$ such that $u_{\hat{2}}(0) = u_2(0) = 0$ and $u_{\hat{2}}(c) = u_2(c)$. Then $u_{\hat{2}}'(c) \leqslant u_2'(c)$, since k is concave, with strict inequality if k is strictly concave. So $b_{\hat{2}}(c) = u_{\hat{2}}'(c)/u_{\hat{2}}(c) \leqslant u_2'(c)/u_2(c) = b_2(c)$.

Theorem 1. Let $F(S,d) = z = (u_1(a_1^*), u_2(a_2^*))$ and $F(\hat{S},\hat{d}) = \hat{z} = (u_1(\hat{a}_1^*), u_{\hat{2}}(\hat{a}_2^*))$, where (S,d) and $(\hat{S},\hat{d})$ are the bargaining games that arise between players 1 and 2 and players 1 and $\hat{2}$, respectively, where player $\hat{2}$ is more risk-averse than player 2. Then $\hat{a}_2^* \leqslant a_2^*$, with strict inequality if player $\hat{2}$ is strictly more risk-averse than player 2.

Proof: By lemmas 1 and 3, $b_1(a_1^*) = b_2(a_2^*) > b_2'(a_2^*)$. Again by lemma 1, $\hat{a}^*$ is the point such that $\hat{a}_1^* + \hat{a}_2^* = Q$ and $b_1(\hat{a}_1^*) = b_2'(\hat{a}_2^*)$. But since both b_1 and b_2 are decreasing (lemma 2), this implies $\hat{a}_1^* \geqslant a_1^*$ and $\hat{a}_2^* \leqslant a_2^*$, again with all inequalities strict if k is strictly concave.

3. Subgame perfect equilibrium of the sequential game

Rubinstein (1982) observed that a stationary subgame perfect equilibrium of the sequential game must specify the division c^* that player 1 will propose in the subgames in which player 1 proposes, and the division e^* that player 2 will propose in the subgames in which player 2 is the proposer. Such a pair of proposals c^* and e^* are supported by a perfect equilibrium if and only if

$$u_1(e_1^*) = (1 - q)u_1(c_1^*),$$

$$u_2(c_2^*) = (1 - q)u_2(e_2^*).$$

That is, each player must be indifferent between accepting the other bargainer's equilibrium proposal or having a $(1 - q)$ chance of receiving his own equilibrium proposal.[2]

Define for any quantity $a \geqslant 0$ the *risk premium* of player i to be the quantity $h_i^q(a)$ such that

$$u_i(a) = (1 - q)u_i(a + h_i^q(a)).$$

That is, $h_i^q(a)$ is the (minimum) premium in excess of the amount a that player i would have to be guaranteed in period $t + 1$ in order to be willing to reject an offer of a in period t (and bear the risk of getting 0 with probability q).

Note that, if c^* and e^* are equilibrium proposals for players 1 and 2, then

$$c_1^* = e_1^* + h_1^q(e_1^*),$$

$$e_2^* = c_2^* + h_2^q(c_2^*).$$

Since $c_1^* + c_2^* = e_1^* + e_2^* = Q$, this implies that

$$h_1^q(e_1^*) = h_2^q(c_2^*).$$

So, loosely speaking, where Nash's solution picks a division with equal $\bar{q}/h$ for small h (and small $\bar{q}$), the perfect equilibrium picks a pair of divisions which, for fixed q, have equal $h = h_1^q(e_1^*) = h_2^q(c_2^*)$. The following new proposition makes precise the relationship between a bargainer's risk premium and his boldness, both in the limit (part (a)), and for the discrete case. This relationship (particularly part (a)) is at the heart of the connection between Nash's solution and the perfect equilibrium division of the sequential game.

Proposition 1. For all agents i,j, and for all $a > 0$,

a) $\lim_{q \to 0} q/h_i^q(a) = b_i(a)$, and

b) $$\int_a^{a+h_i^q(a)} b_i(x)dx = \int_a^{a+h_j^q(a)} b_j(x)dx = -\ln(1-q)$$

Proof: For any $a > 0$, and any agent i, $u_i(a) = (1-q)u_i(a + h_i^q(a))$. To prove part (a), note that

$$\frac{q}{h_i^q(a)} = \frac{[u_i(a + h_i^q(a)) - u_i(a)]/h_i^q(a)}{u_i(a + h_i^q(a))},$$

and $\lim_{q \to 0} h_i^q(a) = 0$, so

$$\lim_{q \to 0} \frac{q}{h_i^q(a)} = \frac{u'(a)}{u(a)} = b_i(a).$$

To prove part (b), note that for any agent i, $u_i(a)/u_i(a + h_i^q(a)) = 1 - q$, so

$$-\ln(1-q) = \int_a^{a+h_i^q(a)} u_i'(x)/u(x)\,dx = \int_a^{a+h_i^q(a)} b_i(x)\,dx.$$

In order to prove the existence of a unique pair of stationary perfect equilibrium proposals c^* and e^*, and to study some of the properties of perfect equilibrium, it will be convenient to define the following two functions that will let us look at some of the above relationships for nonequilibrium divisions c and e. For any nonnegative division $e = (e_1,e_2)$ define $c(e) \equiv (c_1(e),c_2(e))$ by

$$c_1(e) = e_1 + h_1^q(e_1),$$

$$c_2(e) = Q - c_1(e).$$

That is, $c(e)$ is the division[3] that makes player 1 indifferent between getting e_1 at time t, or getting $c_1(e)$ at $t + 1$ with probability $(1 - q)$.

Similarly, for any nonnegative division e, define $H(e)$ to be the division

$$H(e) = c \qquad \text{such that } h_1^q(e_1) = h_2^q(c_2).$$

That is, $c = H(e)$ is the division[4] such that player 2's risk premium at c equals player 1's risk premium at e. So c is the division with the property that the premium required by player 1 to turn down e equals the premium required by player 2 to turn down c.

In terms of the functions $c(e)$ and $H(e)$, lemma 4 summarizes the conditions for c^* and e^* to be perfect equilibrium proposals.

Lemma 4. c^* and e^* are perfect equilibrium proposals if and only if $c^* = c(e^*) = H(e^*)$.

Before proving that there exists a unique pair of perfect equilibrium proposals c^* and e^*, we show that the risk premium is an increasing function. (Note the parallel with lemma 2, which shows that boldness is a *decreasing* function, since b_i is related to q/h_i^q.)

Lemma 5. If $a < b$, then $h_i^q(a) < h_i^q(b)$.

Proof: By proposition 1,

$$\int_a^{a+h_i^q(a)} b_i(x)dx = \int_b^{b+h_i^q(b)} b_i(x)dx.$$

If $a + h_i^q(a) < b$, then lemma 2 implies that the integrand on the right is everywhere smaller than the integrand on the left, so the interval of integration must be longer, i.e., $h_i^q(b) > h_i^q(a)$. If $a + h_i^q(a) > b$, then the interval $[b, a + h_i^q(a)]$ can be deleted from both integrals, and the remaining interval on the right must again be longer than that on the left, i.e., $b + h_i^q(b) - (a + h_i^q(a)) > b - a$, which again implies $h_i^q(b) > h_i^q(a)$. This completes the proof.

Now we can prove the following.

Theorem 2. There is a unique pair of stationary perfect equilibrium divisions c^* and e^*.

We will first show that a pair of perfect equilibrium divisions exist, and then show that there is only one such pair.

Proof of existence: Since we are considering c^* and e^* together, and are hence treating players 1 and 2 symmetrically, we can without loss of generality suppose $h_1^q(Q/2) \geqslant h_2^q(Q/2)$. Consider the (nonequilibrium) division $e = (0,Q)$. Then

$$c_1(e) = e_1 + h_1^q(e_1) = 0 < Q = H_1(e) \qquad \text{since } H(e) = (Q,0).$$

Now consider the (nonequilibrium) division $e = (Q/2, Q/2)$. Then

$$c_1(e) = e_1 + h_1^q(e_1) = Q/2 + h_1^q(Q/2) > Q/2 \geqslant H_1(e).$$

(The last inequality follows from $h_1^q(Q/2) \geqslant h_2^q(Q/2)$. To see this, let $c' = H(e)$, so $h_2^q(c_2') = h_1^q(e_1) = h_1^q(Q/2)$. But by lemma 5, $h_1^q(Q/2) \geqslant h_2^q(Q/2)$ implies $c_2' \geqslant Q/2$, so $H_1(e) = c_1' \leqslant Q/2$.) But the continuity of the u_i implies that the functions $c(e)$ and $H(e)$ are both continuous, and so there exists an e_1^* in the interval $(Q/2, Q)$ such that $c_1(e^*) = H_1(e^*)$, i.e., such that the divisions e^* and c^* are the required equilibrium divisions.[5]

Proof of uniqueness: We have just shown that there exists a pair of divisions e^* and c^* such that $c^* = c(e^*) = H(e^*)$. To see that there cannot be another pair e and c with this property, we will show that, as e moves away from e^*, $c(e)$ and $H(e)$ move in opposite directions from each other, so there is no e different from e^* for which they coincide.

As e_1 increases, $h_1(e_1)$ increases, by lemma 5. So if $c = H(e)$, $h_2(c_2)$ increases (since $h_2(c_2) = h_1(e_1)$, which implies that $c_2 = H_2(e)$ increases. But as e_1 increases, $c_1(e) = e_1 + h_1^q(e_1)$ increases, so $c_2(e)$ decreases. This completes the proof.

Next we consider how the perfect equilibrium divisions c^* and e^* react to changes in the risk aversion of the bargainers. Suppose that player 2, say, is replaced by a more risk-averse bargainer, player $\hat{2}$, i.e., by a player $\hat{2}$ whose utility for money is given by $u_{\hat{2}}(a) = k(u_2(a))$ for all $a \geqslant 0$, where k is an increasing concave function. Then we can give the following parallel to lemma 3.

Lemma 6. For any $c > 0$, $h_{\hat{2}}^q(c) \geqslant h_2^q(c)$, with strict inequality if player $\hat{2}$ is strictly more risk-averse than player 2.

Proof: Normalize $u_{\hat{2}}$ so that $u_{\hat{2}}(0) = u_2(0) = 0$ and $u_{\hat{2}}(c) = u_2(c)$. Then $u_{\hat{2}}(c + h_2^q(c)) \leqslant u_2(c + h_2^q(c))$ with strict inequality if player $\hat{2}$ is strictly more risk-averse than player 2, so $(1 - q)u_{\hat{2}}(c + h_2^q(c)) \leqslant u_2(c) = u_{\hat{2}}(c)$, so $h_{\hat{2}}^q(c) \geqslant h_2^q(c)$.

The following, alternate proof of lemma 6 shows how it follows directly from lemma 3 and proposition 1.

Alternate proof of lemma 6. If player $\hat{2}$ is more risk-averse than player 2, proposition 1 implies

$$\int_a^{a+h_{\hat{2}}^q(a)} b_{\hat{2}}(x)dx = \int_a^{a+h_2^q(a)} b_2(x)dx.$$

Since $b_{\hat{2}}(x) \leqslant b_2(x)$ for all $x > 0$ (lemma 3), it follows that $h_{\hat{2}}^q(a) \geqslant h_2^q(a)$.

We can now give the following proof of the second result stated in the introduction.[6]

Theorem 3. Let c^* and e^* be the perfect equilibrium proposals of players 1 and 2, and let $\hat{c}^*$ and $\hat{e}^*$ be the perfect equilibrium proposals of players 1 and $\hat{2}$, where player $\hat{2}$ is more risk-averse than player 2. Then $\hat{e}_2^* < e_2^*$ and $\hat{c}_2^* < c_2^*$, with strict inequality if player $\hat{2}$ is strictly more risk-averse than player 2.

Proof: Assume that player $\hat{2}$ is strictly more risk-averse than player 2. It will be clear where strict inequalities must be replaced by weak inequalities in the other case. Define the function $\hat{H}(e) = c$ such that $h_1^q(e_1) = \hat{h}_2^q(c_2)$. Let $\hat{c} = \hat{H}(e^*)$. Then $\hat{h}_2^q(\hat{c}_2) = h_1^q(e_1^*) = h_2^q(c_2^*)$, which implies that $\hat{c}_2 < c_2^*$ (by lemmas 6 and 5) and thus $\hat{c}_1 > c_1^*$. So $c_1(e^*) < \hat{c}_1 = \hat{H}_1(\dot{e})$. Since $c_1(e)$ is an increasing function of e_1 and $\hat{H}_1(e)$ is a decreasing function of e_1 (see the proof of uniqueness for theorem 2), this implies that the unique pair of equilibrium proposals $\hat{e}^*, \hat{c}^*$ are such that $\hat{e}_1^* > e_1^*$ and (consequently) $\hat{c}_1^* = \hat{e}_1^* + h_1^q(\hat{e}_1^*) > c_1^*$, which imply the conclusion of the theorem.

Theorem 3 shows that the qualitative effect of risk aversion predicted by the subgame perfect equilibrium of this sequential bargaining model is the same as that predicted by Nash's solution when bargaining is over the (riskless) division of a commodity.[7] The next result[8] presents a stronger connection: it shows that as the probability q that each period will be the last becomes small, the perfect equilibrium divisions c^* and e^* both converge to the division corresponding to Nash's solution for the corresponding game.

Theorem 4. $\lim_{q\to 0}(c_1^*,c_2^*) = \lim_{q\to 0}(e_1^*,e_2^*) = (a_1^*,a_2^*)$ such that $F(S,d) = (u_1(a_1^*), u_2(a_2^*))$.

Proof: Since $h_1^q(e_1^*) = h_2^q(c_2^*)$, $q/h_1^q(e_1^*) = q/h_2^q(c_2^*)$, and so proposition 1 implies that as q goes to 0, $b_1(e_1^*)$ goes to $b_2(c_2^*)$. But c^* and e^* both converge to a single division a^*, and so by lemma 1 this is the division corresponding to Nash's solution.

We conclude this section by remarking on directions in which these results, and the connections between them, can be generalized beyond the simple model considered here. As mentioned earlier, the decision to model the initial wealths of the players as $w_1 = w_2 = 0$, and to set $u_1(0) = u_2(0) = 0$, are both simply normalizations. Nothing essential would change for arbitrary w_i and $u_i(w_i)$, except that in many calculations terms of the form $u_i(c_i)$ would be replaced by terms of the form $u_i(c_i) - u_i(w_i)$.

Similarly, it is not necessary to suppose that the bargaining is over the division of a fixed quantity of a single commodity. The same kind of analysis can be performed for appropriately smooth utility frontiers however they arise (see, e.g., Binmore (1987b) for a discussion of perfect equilibrium of sequential bargaining in this way, and Harsanyi (1956) for a discussion of Nash's solution). Indeed, the maximum probabilities $\bar{q}_i$ that we considered here play a similar role in Harsanyi's (1956) treatment, which considers proposals made in terms of utility payoffs. Harsanyi shows that Zeuthen's (1930) analysis of the bargaining problem, in terms of the maximum risk of conflict a player is willing to face, leads to the same conclusions at Nash's. For those familiar with that work, the connection to the present discussion is most easily seen by noting that if c^* and e^* are equilibrium proposals for the sequential game considered here, it follows that $u_1(e_1^*)u_2(e_2^*) = u_1(c_1^*)u_2(c_2^*)$. That is, both e^* and c^* correspond to the same Nash product, and we have seen why in the limit as q goes to zero they converge to the division that maximizes that product and yields Nash's solution.

Finally, as noted above, Rubinstein's original treatment of the sequential bargaining game allows the players to have different discount factors. In the context of the model of probabilistic termination discussed here, the same kind of results, although not exactly the same arithmetic, can be obtained from a model in which the probability that the game terminates following a rejection by player 1 is different from the probability of termination following a rejection by player 2.

4. Finite sequential games: some experimentally testable predictions

The prediction that Nash's solution makes about the effect of risk aversion on the outcome of bargaining (theorem 1) is also made not only by the perfect equilibrium of the infinite horizon sequential game, but by a wide family of static bargaining models different from Nash's (see Roth, 1979; Kihlstrom, Roth, and Schmeidler, 1981). While many of the other predictions made by these various static models have not been supported by experimental studies (see, e.g., Roth, 1987), the prediction about risk aversion has received experimental support (see Murnighan, Roth, and Schoumaker, 1988).

Since infinite horizon games cannot be directly implemented in a laboratory environment, the purpose of this concluding section is to observe how the prediction embodied in theorem 4 extends to the case of finite horizon sequential games, which can be studied in the laboratory. Preliminary experimental studies focusing on other implications of perfect equilibria in such games suggest that, as in the case of static models, many of these implications may also not be supported by experimental data (see Ochs and Roth, 1989).[9] It will consequently be of considerable interest to see the results of studies concerned with the effect of risk aversion on the outcome of these games.

The rules of the finite horizon game differ from those of the infinite horizon game only in that there is some last period T, such that if the offer made in period T is rejected, each player receives 0. Let $c^*(T) = (c_1^*, c_2^*)$ denote the perfect equilibrium division of the finite horizon game whose last period is at time $t = T$, and in which player 1 makes the first proposal at time $t = 1$. Then $c^*(T)$ can be computed backwards from the end of the game, with $c^*(1) = (Q,0)$, and $c^*(T)$ given by

$$u_2(c_2^*(T)) = (1 - q)u_2(e_2^*(T - 1)),$$

$$c_1^*(T) = Q - c_2^*(T),$$

where $e^*(T - 1)$ is the perfect equilibrium division of the subgame in which player 2 makes the first proposal following a rejection of player 1's (first) offer.

For a given final period T, let player $\hat{1}$ be a player who is strictly more risk-averse than player 1, and let player $\hat{2}$ be strictly more risk-averse than player 2. Let $\hat{c}(\hat{1}) \equiv (\hat{c}_{\hat{1}},\hat{c}_2)$ denote the perfect equilibrium division when player $\hat{1}$ bargains with player 2, and $\hat{c}(\hat{2}) \equiv (\hat{c}_1,\hat{c}_{\hat{2}})$ denote the perfect equilibrium division when player 1 bargains against $\hat{2}$. Then the finite horizon game yields the following analogue to theorem 3.

Theorem 5. a) For $T = 1$, $c^* = \hat{c}(\hat{1}) = \hat{c}(\hat{2}) = (Q,0)$,
b) For $T = 2$, $c^* = \hat{c}(\hat{1})$, $\hat{c}_{\hat{2}} < c_2^*$,
c) For $T \geqslant 3$, $c_i < c_i^*$ for $i = 1,2$.

The proof is not difficult (although it requires the consideration of several cases) and will be left to the reader.

Note that, for $T \geqslant 3$ (i.e., for games of three or more periods), the perfect equilibrium division makes the same qualitative prediction as in the infinite horizon case about the (disadvantageous) effect of risk aversion. However, when $T = 1$, the risk aversion of the bargainers is predicted to have no effect, and when $T = 2$, only the risk aversion of player 2 influences the perfect equilibrium division. Neither the finite horizon game nor the infinite horizon game are symmetric between the players, but in the infinite horizon game we could treat the players symmetrically by considering c^* and e^* together, since there is symmetry between the subgames at which players 1 and 2 propose. However, as we see when $T = 2$, even this symmetry is absent in the finite horizon case, since the initial game (in which player 1 proposes) is not the same as the subgame in which player 2 proposes, since the two games do not have the same length. It is therefore noteworthy that the qualitative predictions of theorem 5 for $T \geqslant 3$ are nevertheless the same as in the infinite horizon game.

However, as noted above, recent experiments have raised a number of questions about the descriptive accuracy of other predictions of perfect equilibrium in games of this sort. When more such experiments have been done, the situation may come to resemble in this respect that of the various static models that have been subjected to experimental investigation, which were also found to lack descriptive accuracy in some important respects. It proved necessary, in designing experiments concerning the predicted effects of risk aversion in these models, to take account of some of the systematic but unpredicted effects observed in previous experiments (Murnighan, Roth, and Schoumaker, 1988). Similarly, it seems likely that persuasive experimental tests of the perfect equilibrium predictions about risk aversion in the sequential games considered here will have to wait until there is a better understanding of the descriptive accuracy of other predictions of perfect equilibrium in sequential games of this kind.

Notes

1. Rubinstein (1982) also showed that there is no nonstationary perfect equilibrium of the game in question. The present argument is thus not an alternate proof of his result.
2. If either player preferred the other bargainer's proposal to a $1 - q$ chance of receiving his own, the proposals could not be supported by a subgame perfect equilibrium because at such an equilibrium the player cannot credibly threaten to refuse a smaller offer. If either player preferred a $1 - q$ chance of receiving his own proposal, which is accepted at equilibrium, then again no subgame perfect equilibrium can support the two proposals, since now it is in the player's interest to reject the other bargainer's proposal, and hence in the other bargainer's interest to make a larger offer. However, when both bargainers are indifferent, the strategy pair in which each player always refuses any offer that is less than the other bargainer's indicated proposal, and accepts any offer of at least as much, and in which each bargainer proposes the indicated proposal, is a subgame perfect equilibrium.
3. Note that $c_2(e)$ may be negative.
4. Note that $c_1 = H_1(e)$ may be negative.
5. Note that while $c(e)$ and $H(e)$ need not be nonnegative for arbitrary e, $c^* = c(e^*) = H(e^*)$ must be, since $c_1(e)$ and $H_2(e)$ are positive for positive e.
6. As with theorem 1, the result here is stated in terms of players $\hat{2}$ and 2, but the symmetry of the problem (since each subgame has the same structure as the original game) ensures the symmetric result (with the same proof) if these were everywhere exchanged for players $\hat{1}$ and 1.
7. Of course, in the sequential bargaining model, we have interpreted q as a probability that the game will end, so risk is explicitly present in the model. However, the same result would of course apply if we interpret q as the (common) discount factor of the players, unrelated to any probabilistic risk. In that case the theorem continues to state that risk aversion is disadvantageous in bargaining. I find this an intuitive result, even though the underlying agreements are themselves riskless, since it is arguably the risk of disagreement or delay that forces bargainers to come to a particular agreement, and so a more risk-averse bargainer can be expected to offer better terms to reduce this risk, etc. However, note that not all researchers agree: Binmore, Rubinstein, and Wolinsky (1986) state simply (p. 179) that "... there is no apparent reason that a party's attitudes toward risk should affect his bargaining position in a riskless environment." My own feeling is that this is best treated as an empirical issue. When concave transformations of utility are interpreted as changes in risk aversion, testable predictions result, and their truth or falsity is an empirical matter that can be addressed, for example, with experimental methods. Of course, as Binmore et al. emphasize, changes in concavity can be given other interpretations as well, so when the utility functions are not expected utility functions, theorem 3 can be interpreted differently.
8. This was originally proved independently by Binmore (1987), MacLennan (1982), and Moulin (1982), using arguments different from one another and from the argument that follows. For another proof, see Binmore, Rubinstein, and Wolinsky (1986).
9. For some related studies that reach a variety of conclusions on these matters, see Binmore, Shaked, and Sutton (1985), Guth, Schmittberger, and Schwartz (1982), Guth and Tietz (1987), and Neelin, Sonnenschein, and Spiegal (1988).

References

Aumann, Robert J. and M. Kurz. (1977a). "Power and Taxes," *Econometrica* 45, 1137–1161.
Aumann, Robert J. and M. Kurz. (1977b). "Power and Taxes in a Multi-Commodity Economy," *Israel Journal of Mathematics* 27, 185–234.

Binmore, Ken. (1987a). "Nash Bargaining Theory II." In K. Binmore and P. Dasgupta (eds.), *The Economics of Bargaining*. Oxford: Basil Blackwell, pp. 61–76.

Binmore, Ken. (1987b). "Perfect Equilibria in Bargaining Models." In K. Binmore and P. Dasgupta (eds.), *The Economics of Bargaining*. Oxford: Basil Blackwell, pp. 77–105.

Binmore, Ken, Ariel Rubinstein, and Asher Wolinsky. (1986). "The Nash Bargaining Solution in Economic Modelling," *Rand Journal of Economics* 17, 176–188.

Binmore, Ken, Avner Shaked, and John Sutton. (1985). "Testing Noncooperative Bargaining Theory: A Preliminary Study," *American Economic Review* 75, 1178–1180.

Guth, Werner, R. Schmittberger, and B. Schwarz. (1982). "An Experimental Analysis of Ultimatum Bargaining," *Journal of Economic Behavior and Organization* 3, 367–388.

Guth, Werner, and Reinhard Tietz. (1987). "Ultimatum Bargaining for a Shrinking Cake—An Experimental Analysis," mimeo.

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* 24, 144–157.

Kihlstrom, Richard E. and Leonard J. Mirman. (1974). "Risk Aversion with many Commodities," *Journal of Economic Theory* 8, 361–388.

Kihlstrom, R., A.E. Roth, and D. Schmeidler. (1981). "Risk Aversion and Solutions to Nash's Bargaining Problem." In O. Moeschlin and D. Pallaschke (eds.), *Game Theory and Mathematical Economics*. Amsterdam: North-Holland, pp. 65–71.

MacLennan, Andrew. (1982). "A General Noncooperative Theory of Bargaining," mimeo, University of Toronto.

Murnighan, J. Keith, Alvin E. Roth, and Francoise Schoumaker. (1988). "Risk Aversion in Bargaining: An Experimental Study," *Journal of Risk and Uncertainty* 1, 101–124.

Moulin, Herve. (1982). "Bargaining and noncooperative implementation," mimeo, Ecole polytechnique.

Nash, John F. (1950). "The Bargaining Problem," *Econometrica* 28, 155–162.

Neelin, Janet, Hugo Sonnenschein, and Matthew Spiegel. (1988). "A Further Test of Noncooperative Bargaining Theory," *American Economic Review* 78, 824–836.

Ochs, Jack and Alvin E. Roth. (1989). "An Experimental Study of Sequential Bargaining," *American Economic Review,* 79, 355–384.

Roth, Alvin E. (1979). *Axiomatic Models of Bargaining*. Berlin: Springer.

Roth, Alvin E. (1985). "A Note on Risk Aversion in a Perfect Equilibrium Model of Bargaining," *Econometrica* 53, 207–211.

Roth, Alvin E. (1987). "Bargaining Phenomena and Bargaining Theory." In A.E. Roth (ed.), *Laboratory Experimentation in Economics: Six Points of View*. Cambridge: Cambridge University Press, pp. 14–41.

Rubinstein, Ariel. (1982). "Perfect Equilibrium in a Bargaining Model," *Econometrica* 50, 97–110.

Yaari, M. E. (1969). "Some Remarks on Measures of Risk Aversion and On Their Uses," *Journal of Economic Theory* 1, 315–329.

Zeuthen, Frederik. (1930). *Problems of Monopoly and Economic Welfare*. London: G. Routledge.