Cowles Foundation Paper 173
Reprinted from Journal of Conflict Resolution, 6(1), 1962

Bargaining in ignorance of the opponent's utility function[1]

JOHN C. HARSANYI

Department of Economics, Wayne State University

I

The Zeuthen–Nash theory of two-person bargaining games, like most of existing game theory,[2] is based on the assumption that the two parties *know* each other's utility functions (Nash, 1950, 1953; Zeuthen, 1930, ch. 4; Harsanyi, 1956, 1961). This essentially means that they know each other's *preferences* as well as each other's attitudes towards *risk*. On the basis of this assumption and certain rationality postulates, the theory defines optimal strategies for both parties, and also undertakes to uniquely predict the outcome of bargaining between two rational bargainers. In this paper, I propose to discuss the more general case where the two parties do *not* know (and know they do not know) each other's utility functions. I shall also indicate some interesting problems our analysis raises for empirical research.

II

In bargaining, and more generally in all non-trivial game situations, the behavior of a rational individual will depend on what he expects the *other* party will do. Party 1 will ask for the best terms he expects *party 2* to accept. But party 1 will know that the terms party 2 will accept in turn depend on what terms party 2 expects *party 1* to accept. Thus, party 1's behavior will depend on what may be called his *second-order* expectations, i.e., on party 1's expectations concerning party 2's *expectations* about party 1's behavior. These again will depend on party 1's *third-order* expectations, i.e., on his expectations concerning party 2's *second-order* expectations, etc.

To be able to keep track conceptually of all these compound expectations of various orders, we introduce the following notation. Let A_1 and A_2 be the two parties' *concession points*, i.e., the *least favorable* terms that party 1 and party 2 respectively would accept rather than having to face a conflict situation. Let the operator e_1 stand for "party 1's estimate (or expectation) of . . ."

[1] The original version of this paper was written at the Cowles Foundation for Research in Economics at Yale University, under Contract Nonr–358 (01), NR 047–006 of the Cowles Foundation with the Office of Naval Research. It was distributed as *Cowles Foundation Discussion Paper* No. 46. A revised version of this paper was presented at the Princeton Conference on *Game Theory and Negotiations* in October, 1961. The paper has benefited from comments by the other participants of the Conference.

[2] But of course there are exceptions. For a model dealing with the case where the players do *not* know one another's utility functions, see (Luce and Adams, 1956).

and e_2 stand for "party 2's estimate (or expectation) of . . . ."

Then our previous argument can be restated as follows. Party 1 will choose his own concession point A_1 on the basis of his estimate of party 2's concession point A_2. That is, A_1 will depend on e_1A_2. But party 2 in turn will choose A_2 on the basis of his estimate of A_1. Hence A_2 will depend on e_2A_1.

As party 1 will know that A_2 depends on e_2A_1, his own estimate of the former will depend on his estimate of the latter. Hence e_1A_2 will depend on $e_1e_2A_1$. For similar reasons e_2A_1 will depend on $e_2e_1A_2$.

As party 1 will know that e_2A_1 depends on $e_2e_1A_2$, his own estimate of the former will again depend on his estimate of the latter. Hence $e_1e_2A_1$ will depend on $e_1e_2e_1A_2$. For similar reasons, $e_2e_1A_2$ will depend on $e_2e_1e_2A_1$, etc.

To sum up, A_1 will depend on e_1A_2

e_1A_2 on $e_1e_2A_1$

$e_1e_2A_1$ on $e_1e_2e_1A_2$, etc.

That is, the analysis of party 1's choice of his concession point A_1 leads to an infinite sequence $\{A_1(n)\}$ of compound expectations where the *odd* members of the sequence have the form

(1) $$A_1(n) = A_1(2k-1) = (e_1e_2)^{k-1}A_1$$

whereas the *even* members have the form

(2) $$A_1(n) = A_1(2k) = (e_1e_2)^{k-1}e_1A_2$$

But the trouble is that in this infinite sequence each member $A_1(n)$ depends on the *next* member $A_1(n+1)$ — and this clearly leads to an *infinite regress*. Obviously, the analysis of party 2's choice of A_2 would yield a similar conclusion.

III

One of the important achievements of game theory in general, and of the Zeuthen–Nash theory of bargaining games in particular, has been to show how this infinite regress of compound expectations can be analytically resolved. This resolution is achieved essentially by formulating certain *equilibrium conditions* or *consistency requirements* that rational bargainers' expectations must satisfy concerning each other's bargaining behavior (Harsanyi, 1961, esp. pp. 183–90).

More specifically, the Zeuthen–Nash theory makes use of the fact that a bargaining party faced with a presumably rational opponent cannot *rationally expect* this opponent to make a concession in a situation where he himself, following his *own* criteria of rational behavior, would refuse to make a concession. This imposes a strong symmetry requirement on the bargaining strategies that can be rationally chosen by two bargainers who expect each other to act rationally. This symmetry postulate, together with some other very natural postulates of rational behavior, then selects a unique solution (equilibrium agreement point) A^* for each particular bargaining game Γ. (For Nash's own axioms, see [Nash, 1950, pp. 156–9]. For an alternative but equivalent set of axioms, based directly on consistency requirements concerning the two parties' expectations, see [Harsanyi, 1961].)

Mathematically the Zeuthen–Nash solution can be defined as follows. Let u_1 and u_2 be the two parties' utility functions. Let C be the conflict situation which would arise if the parties reached no agreement. Let $\mathcal{A}$ denote the set of all possible terms (agreement points) A the two parties could agree upon. Then the solution point A^* is that particular agreement point A which maximizes the product

(3) $$\pi_0 = \pi(u_1, u_2) = [u_1(A) - u_1(C)] \cdot [u_2(A) - u_2(C)]$$

subject to the conditions:

$$A \in \mathcal{A}$$

$$u_i(A) \geqq u_i(C) \qquad [i = 1, 2]$$

Assuming that the two parties do know each other's utility functions, the infinite regress in compound expectations now disappears. This is so because both parties, if they follow the Zeuthen–Nash rationality postulates, will *accept*, and will also *expect* each other to accept, the solution point A^* as their agreement point. Therefore we shall have

$$(4) \quad A_1 = e_2A_1 = e_1e_2A_1 = e_2e_1e_2A_1 = \ldots = A^*$$

$$(5) \quad A_2 = e_1A_2 = e_2e_1A_2 = e_1e_2e_1A_2 = \ldots = A^*$$

IV

But, whereas the Zeuthen–Nash solution resolves the problem of compound expectations in the case where the two parties *know* each other's utility functions, the problem immediately reappears in a somewhat different form if we relax the assumption of mutually known utility functions.

Suppose that both bargainers follow, and also expect each other to follow, the rationality postulates of the Zeuthen–Nash theory. But let us assume that they do *not* know each other's utility functions. However, suppose that each party will form *single-valued* estimates of the other party's utility function and of other variables unknown to him. Of course, in general these estimates will not coincide with the *true* values of these variables.[3]

Then, party 1 may *attempt* to choose his concession point A_1 by maximizing the product $\pi_0 = \pi(u_1, u_2)$ defined by equation (3). But, as in fact the function u_2 is unknown to him, all he *can* do is to maximize the product $\pi_1 = \pi(u_1, e_1u_2)$, where his own more or less inaccurate *estimate* e_1u_2 replaces the *true* utility function u_2 of his opponent. Thus *in fact* A_1 will be the point where the product $\pi(u_1, e_1u_2)$ is maximized. We can regard this point A_1 as party 1's *estimate* of the *true* Zeuthen–Nash solution A^* and can write

$$(6) \quad A_1 = e_1A^*$$

Similarly, party 2's concession point A_2 will be the point where the product $\pi_2 = \pi(e_2u_1, u_2)$ is maximized, and we can interpret this point A_2 as

$$(7) \quad A_2 = e_2A^*$$

But more sophisticated bargainers can do better than that. Party 1 knows that party 2 *in fact* chooses his concession point A_2 by maximizing the product π_2, rather than by maximizing the theoretically "correct" product π_0 itself. Therefore party 1 will try to estimate the functions e_2u_1 and u_2 occurring in π_2, and then will maximize the product $\pi_{21} = \pi(e_1e_2u_1, e_1u_2)$ which uses his estimates $e_1e_2u_1$ and e_1u_2 of these two functions. The point $A_1(2)$ where π_{21} is maximized will represent party 1's estimate of *party 2's* concession point $A_2 = e_2A^*$. At the same time it will also serve as party 1's *own* (adjusted) concession point because, being party 1's estimate of A_2, it will indicate the concessions that party 1, in his own assessment of the situation, has to make on his own part in order to avoid a conflict. Thus party 1's *second-order* adjusted concession point will be

$$(8) \quad A_1(2) = e_1A_2 = e_1e_2A^*$$

which will be the point where the product π_{21} is maximized.

Similarly, party 2 can follow the more sophisticated strategy of choosing his own *second-order* adjusted concession point as

[3] We would obtain essentially the same conclusions if we made the more general assumption that, instead of forming *single-valued* estimates, each party would assign a subjective *joint probability distribution* to the variables unknown to him, and then would try to maximize his *expected utility* in terms of this probability distribution.

(9) $A_2(2) = e_2A_1 = e_2e_1A^*$

which will be the point where the product $\pi_{12} = \pi(e_2u_1, e_2e_1u_2)$ is maximized.

Of course, a still higher degree of sophistication will require party 1 to choose his *third-order* concession point by trying to estimate party 2's *second-order* concession point $A_2(2)$. This *third-order* concession point can be written as

(10) $A_1(3) = e_1e_2A_1 = e_1e_2e_1A^*$

and will be the point where the product $\pi_{121} = \pi(e_1e_2u_1, e_1e_2e_1u_2)$ is maximized, etc.

In general, party 1's *nth-order* concession point will have the form

(11) $A_1(n) = A_1(2k-1) = (e_1e_2)^{k-1}e_1A^*$

if n is *odd*, and

(12) $A_1(n) = A_1(2k) = (e_1e_2)^kA^*$

if n is *even*.

The only fully satisfactory concession point for party 1 will be the limit $\bar{A}_1$ of this sequence $\{A_1(n)\}$ — provided that such a limit exists. $\bar{A}_1$ will be called party 1's *final* concession point. Party 2's final concession point $\bar{A}_2$ will be defined in a similar way. Of course, in view of the definition of $\bar{A}_1$ and $\bar{A}_2$ as *limits* of the sequences $\{A_1(n)\}$ and $\{A_2(n)\}$ we must have

(13) $\bar{A}_1 = e_1\bar{A}_2$ and $\bar{A}_2 = e_2\bar{A}_1$.

That is, $\bar{A}_1$ as party 1's final concession point must be *consistent* with his expectations concerning party 2's final concession point, and vice versa.

The resemblance of equations (11) and (12) to the equations (1) and (2), obtained in the case of mutually *known* utility functions, is obvious. But there is the important difference that now the nth member $A_1(n)$ of the sequence does *not* depend on the $(n+1)$th member $A_1(n+1)$, but rather depends on party 1's higher-order expectations (i.e., on his expectations about his opponent's expectations) of the appropriate order, with respect to the utility functions u_1 and u_2. Thus the problem of an infinite regress does not arise. But the problem of compound expectations does arise in the following somewhat different form.

V

In real life, bargaining situations where the parties concerned have very little reliable information about each other's utility functions (i.e., about each other's preferences and/or attitudes towards risk) are fairly common. But usually both parties still manage to choose *some* bargaining strategies, and in particular to choose *some* concession points $\bar{A}_1$ and $\bar{A}_2$, consistent with their *expectations* about each other's behavior. This seems to show that for the two parties the sequences $\{A_1(n)\}$ and $\{A_2(n)\}$ usually do *converge* to some reasonably well-defined points $\bar{A}_1$ and $\bar{A}_2$.

Indeed, quite often the two parties do reach an *agreement*, i.e., avoid insisting on mutually incompatible demands. This means that they manage to choose not only well-defined but also mutually *compatible* concession points $\bar{A}_1$ and $\bar{A}_2$. This seems to happen much more often than mere chance would allow. These two facts lead to the following two questions:

1. What are the mechanisms that tend to ensure that for each party i ($i = 1, 2$) the sequence $A_i, A_i(2), \ldots, A_i(n), \ldots$ will actually *converge* to a well-defined limit $\bar{A}_i$, which then can serve as this party's actual concession point during the bargaining process (or during some particular stage of the bargaining process)?

2. What are the mechanisms that tend to ensure that the two parties' concession points $\bar{A}_1$ and $\bar{A}_2$, representing the limits of

the sequences $\{A_1(n)\}$ and $\{A_2(n)\}$, will be mutually *compatible*?

The answer to question 1 seems to be relatively easy. If nothing else, the mere limitations in each party's information-processing and computing abilities will ensure that he will not proceed beyond the first few steps in separately estimating what his opponent's utility function is, what his opponent's estimate of his own utility function is, what his opponent's estimate of his own estimate of his opponent's utility function is, etc. Formally this means that the sequence $\{A_i(n)\}$ will become a constant sequence after, say, the first k members, and will therefore necessarily converge.

Moreover, as will become obvious from our subsequent argument, the mechanisms instrumental in making the two parties' final concession points $\bar{A}_1$ and $\bar{A}_2$ mutually *compatible* also help to make each party's expectation sequence $\{A_i(n)\}$ *converge* to a well-defined concession point $\bar{A}_i$. Thus our answer to question 2 will also provide a part of the answer to question 1.

Going over to question 2, two main types of possible mechanisms seem to suggest themselves. On the one hand it is conceivable that in a given society with well-established cultural traditions people tend to *enter* bargaining situations with more or less consistent expectations about each other's utility functions. It may happen that *all* members of a given society are expected to have essentially the *same* utility function. Or, more realistically, we may assume that at least persons of a *given* sex, age, social position, education, etc. are expected to have similar utility functions of a specified sort. Any utility function u^0 attributed by public opinion to persons of a given description we shall call a *stereotype utility function*. Under this assumption, in a given bargaining situation, say, party 1 will know that party 2 will expect him to display a certain stereotype utility function u_1^0, and party 2 will know that party 1 will expect party 2 to entertain this particular expectation, etc. Likewise, party 2 will be expected to display (a possibly different) stereotype utility function u_2^0, etc. This mechanism we shall call *mechanism I*.

Alternatively, we may assume that compatibility between the two parties' final concession points $\bar{A}_1$ and $\bar{A}_2$ is *not* the result of their already *entering* the bargaining situation with mutually consistent expectations, but is rather the result of mutual *adjustment* of their expectations *during* the bargaining process itself. This hypothetical adjustment process we shall call *mechanism II*. Clearly, the common sense interpretation of the bargaining process is essentially in terms of mechanism II. Bargaining differs from a mere exchange of "ultimatums" in that the bids made, up to the very last bid of each party, are not final but are merely tentative, and serve to *test out* the opponent's attitudes.

But on more careful analysis it turns out, it seems to me, that the common sense view tends to overstate the importance of mechanism II, and that our analysis must make use of *both* mechanisms and indeed, in many cases, must rely primarily on mechanism I.

This is so because mechanism II can operate only if *bluffing* (i.e., gross overstatement by either party of his true final demands—in other words, gross misrepresentation of the true position of his own concession point) can be brought under effective control; and it seems to be very questionable whether this condition is satisfied to a sufficient extent in most real-life bargaining situations.

To show this point, consider a bargaining situation where bluffing is fully permissible and is not subject to any penalties whatever. Suppose that parties 1 and 2 enter this bargaining situation with the mutually *incompatible* concession points $\bar{A}_1$ and $\bar{A}_2$. That is, party 1 plans not to retreat beyond $\bar{A}_1$ because he mistakenly expects that party 2 will accept $\bar{A}_1$ in the end. On the other hand, party 2 plans not to retreat beyond $\bar{A}_2$ because he mistakenly expects that party 1 will accept $\bar{A}_2$. Although these expectations are mutually inconsistent, the two parties will never have a chance of revising these expectations during the bargaining process—before it will be too late and the negotiations will break down. Of course, each party will notice that during the bargaining process his opponent is showing a tougher attitude than the first party anticipated. But he will have no reason to regard this as more than mere bluffing, and up to the last minute will expect his opponent to reverse his attitude in the end. Therefore, so long as unrestricted bluffing is permitted, mechanism II cannot make the two parties' concession points compatible if they have not been compatible from the very beginning.

In contrast, mechanism I can be fully effective in this case. If party 2 attributes the stereotype utility function u_1^0 to party 1, party 1 will have strong incentives to act *as if* u_1^0 *were* his true utility function, even if his true utility function u_1 is *in fact* rather different. For, if his true utility function suggested *tougher* attitudes than appropriate for the stereotype utility function u_1^0 ascribed to him then, for the reasons already stated, he would not be *able* to convince his opponent of this fact. His opponent would go on interpreting his behavior in terms of the stereotype utility function u_1^0, and would regard his apparent toughness as mere bluff.

On the other hand, if party 1's true utility function u_1 suggested a *less tough* attitude than appropriate for the stereotype utility function u_1^0 attributed to him, then in general it would not *pay* him to communicate this information to his opponent, because he would obtain better terms from his opponent if the latter thought him to be tougher in his attitudes than he really was.

In the case of bargaining about money (or some other value admitting of quantitative measurement), under the Zeuthen–Nash theory a given individual i will show more readiness to take *risks* and will show *tougher* bargaining attitudes the *more steeply* the slope of his utility function (i.e., his marginal utility for money) *increases* in the relevant range.[4] What our last result says is that, however the shapes of his stereotype utility function u_1^0 and of his true utility function u_1 compare in this respect, he will have to follow his stereotype utility function u_1^0 rather than his true utility function u_1.[5,6]

[4] That is, the larger on the average the *second* derivative of his utility function is within this range.

[5] Under this model, the only case where it can be *both* practicable *and* profitable to change the opponent's views about one's true utility function is this. The disclosure of the relevant information about one's true utility function must shift the equilibrium agreement point $\bar{A}$ to a point $\bar{A}'$ yielding *both* parties higher utilities than $\bar{A}$ did. Then the change will be profitable to the party initiating it, and will be at the same time acceptable also to the other party. But it does not follow that it always pays for a given party i to point out the *existence* of mutually preferable potential agreement points $\bar{A}'$ to his opponent. This is so because disclosure of this information may weaken party i's *relative* bargaining position to such an extent that the new equilibrium agreement point will not be some point $\bar{A}'$ yielding higher utilities to *both* parties but rather will be some point $\bar{A}''$ yielding a higher utility to his opponent but yielding a lower utility to party i *himself*.

[6] Of course, if it happens very often that persons of a given social group have to act *as if* they did actually possess the stereotype util-

To sum up, in the case where bluffing is unrestricted, if the two parties ascribe each other the stereotype utility functions $u_1{}^0$ and $u_2{}^0$, although their true utility functions are in fact u_1 and u_2, we shall have

(14) $e_2u_1 = e_1e_2u_1 = e_2e_1e_2u_1 = \ldots = u_1{}^0$ even if $u_1 \neq u_1{}^0$ and

(15) $e_1u_2 = e_2e_1u_2 = e_1e_2e_1u_2 = \ldots = u_2{}^0$ even if $u_2 \neq u_2{}^0$.

Accordingly, both sequences $\{A_1(n)\}$ and $\{A_2(n)\}$ will converge to a common limit $A = \bar{A}$, which will be the point where the product $\pi^0 = \pi(u_1{}^0, u_2{}^0)$ is maximized. (More particularly, whereas the *first* member of each sequence may be different from $\bar{A}$, *all other* members will be simply equal to $\bar{A}$.)

VI

Now we shall consider the case where "excessive" bluffing is subject to certain penalties.

A simple way of introducing penalties for bluffing is this.[7] The bargaining process is divided into n stages. At any given stage, say, at stage k, both bargainers know that the negotiations will break up (because, e.g., at least one of them will leave the negotiating table) if their bids $B_1(k)$ and $B_2(k)$ will be further apart than some specified distance $D(k)$. (If the bargaining is about a money price, then $D(k)$ can be measured in terms of the difference between the price asked for and the price offered, etc.) At the first stage of the bargaining process even large differences between the two parties' bids are tolerated, i.e., $D(1)$ is large. But at later stages (with larger k) $D(k)$ gradually decreases until at the last stage $D(n) = 0$, i.e., the bargainers must come to accept the same terms if they are to reach an agreement.

Under this model it will be still true that at any given moment, say, party 1's *true* concession point $\bar{A}_1$ will be at that point where he expects his *opponent's* concession point to be, i.e., $\bar{A}_1 = e_1\bar{A}_2$ (see equation [13]). But at stage k party 1 will know that he can safely overshoot $\bar{A}_2$ by the distance $D(k)$. Hence his actual bid $B_1(k)$ will tend to go beyond $\bar{A}_1 = e_1\bar{A}_2$ by the amount $D(k)$. To this extent he will be bluffing. But his opponent, party 2, will be fully aware of this, and will also know that party 1's *true* concession point is at a distance $D(k)$ from his actual bid $B_1(k)$. Hence, *after* party 1 has made his bid $B_1(k)$, i.e., during stage $(k+1)$, party 2 will know where party 1's concession point $\bar{A}_1$ *was* at the previous stage, i.e., at stage k. Likewise party 1 will know where party 2's true concession point was at the previous stage.

Of course the two parties will never know exactly at any given stage where each other's true concession points are at *that* stage, because neither party will know how his own last bid has shifted his opponent's true concession point. Yet, knowing how the opponent's concession point moved during all the preceding stages of the negotiations, they will be able at least to make well-informed guesses, and presumably by the nth stage, if not earlier, will understand reasonably well what terms they can ask for in their last bids without too much risk of the opponent's rejecting them.

Thus, under this model, mechanism II would work quite effectively. Even if the two parties started out with mutually inconsistent expectations, there would be a good chance of their reaching mutually consistent expectations by the end of the

ity function attributed to them, this will probably tend in real fact to assimilate their *true* utility functions to these stereotypes as a result of certain well-known psychological mechanisms.

[7] This model was suggested to me by Professor Robert J. Aumann.

bargaining process. Of course, the availability of mechanism I would still help. That is, if the parties have already started with not-too-divergent expectations, there will be less danger that at the first stage their differences will exceed the permitted tolerance limit $D(1)$, and it will presumably take fewer steps for them to reach full agreement.

However, the full effectiveness of mechanism II strongly depends on the actual assumptions of our model. In most bargaining situations there probably is something like a tolerance limit $D(k)$ for the differences between the two parties, and negotiations tend to break up if these differences are too great and persist in being too great. But the precise value of the tolerance limit $D(k)$ is hardly ever known to the two parties, and probably should be treated as a random variable subject to chance influences.

Under such conditions the actual bids of the two parties will allow only very vague inferences about their true concession points, and will convey much less information about the parties' true attitudes than was the case in the model we have discussed. Consequently, there will be much less opportunity during the bargaining process for mutual adjustment towards greater consistency in the two parties' expectations. Hence, all the greater will be the importance of already *starting* with fairly consistent expectations, based on stereotype utility functions accepted by both parties.[8]

VII

In a world where people would *know* each other's utility functions, there would be no need for bargaining in the usual sense because they would not have to *test out* each other's utility functions by the laborious procedure of successive bids and counterbids. Both parties could simply state their final terms independently of each other (e.g., in writing) and then compare their bids. If these turned out to be compatible, each party would receive what he had asked for. If they turned out to be incompatible, of course a conflict would result since by assumption both parties' bids would be regarded as *final*, i.e., as having the nature of an ultimatum. However, in fact, between rational bargainers a conflict would never arise because each party would know what his own payoff would be under the solution of the game and would only ask for that particular payoff, so that the two parties' bids would never contain incompatible demands.

Even in a world where the parties would *not* know each other's utility functions, but where *bluffing* would be unrestricted, bargaining would make little sense because the parties could not obtain any useful information about each other's unknown utility functions by observing each other's bargaining behavior—at least if both parties acted rationally. (Only if one of the par-

[8] Another variant of mechanism II seems to operate between two parties who repeatedly become involved in bargaining situations with each other. In such cases a mutual understanding may develop to exclude bluffing, or at least the grosser forms of it. This understanding tends to be observed by both parties in order to maintain their good relations, which both parties are interested to preserve. For instance, as has been pointed out by Dean G. Pruitt, in most cases very little bluffing and misrepresentation occurs in informal diplomatic negotiations between representatives of friendly governments, even in cases involving serious conflicts of interests between the parties. A different mechanism that may help the parties to arrive at mutually consistent expectations has been suggested (Schelling, 1957). It is based on certain psychologically *salient* features of the situation, which may catch both parties' attention and may suggest some particular agreement point to both of them.

ties made a *mistake* and gave away information *weakening* his own bargaining position, could the other party obtain useful information during the bargaining process.) Of course, bargaining could still persist as a form of entertainment or as a custom serving ceremonial purposes.[9] There would be scope also for negotiations for purposes other than bargaining, in particular for communicating factual information, e.g., about possibilities of cooperation, or about each party's strategical possibilities in case of a conflict, etc.

In the real world, however, there *is* scope for true bargaining because the parties in general do not fully know each other's utility functions, and because they have *some* incentives to avoid excessive bluffing so that observing the opponent's bids does allow them some inferences about the latter's true utility function. Moreover, there is always the possibility that the opponent will make a tactical mistake and will lay bare some of his weak points during the bargaining process.

Assuming that we have been right in arguing that the amount of useful information the parties can in fact learn about each other's attitudes during the bargaining process is rather limited (and if it is true that experienced bargainers will seldom unnecessarily disclose information about their own weak points), then in fact bargaining, as distinguished from an exchange of single final bids, is a practice of much *lesser* usefulness than may appear to be the case at first sight.

VIII

In summation, even if the parties *know* each other's utility functions, the problem of choosing a rational bargaining strategy leads to the problem of *compound expectations*, as the parties must try to estimate each other's concession points, as well as each other's estimates of these estimates, etc. The Zeuthen–Nash theory of bargaining games resolves this problem by stating certain *consistency requirements* to be satisfied by rational bargainers' expectations about each other's behavior.

But the problem of compound expectations reappears if we relax the assumption of mutually known utility functions. The problem now becomes, in cases where they do reach agreements, what mechanisms secure convergence of the two parties' expectations to some common limit.

We have considered two possible mechanisms. One was the existence of *stereotype utility functions*, which tend to make the members of the same society and culture *enter* bargaining situations more or less with already mutually consistent expectations (mechanism I). The other was mutual *adjustment* of the two parties' expectations *during* the bargaining process itself, as a result of their testing out each other's attitudes by means of tentative bids and counterbids (mechanism II).

We have seen that mechanism II can operate only if *bluffing* by the parties is brought under control. We have argued that this condition is rather imperfectly met in most real-life situations, which restricts the usefulness of bargaining for the purpose of testing out the opponent's true attitudes.

It may be added that ignorance and mis-

[9] In the real world, bargaining processes—or at least some parts of these processes—do sometimes seem to serve a ceremonial function. In collective bargaining on the labor market, and in certain types of diplomatic or political bargaining, protracted negotiations are sometimes made necessary by the fact that the constituents of the negotiators would think they did not try hard enough to obtain better terms if they reached an agreement without negotiations of some length.

judgment of each other's utility functions is one of the basic reasons why conflicts can occur between rational individuals.[10] The two mechanisms we have discussed are clearly not effective in all cases.

One case where the effectiveness of both mechanisms tends to be greatly reduced is that of bargaining between members of two different societies or cultures. In this case the two parties are likely to entertain different and mutually inconsistent stereotypes about each party's utility function. At the same time, they are also likely to have different ideas about how much bluffing they themselves may indulge in and how much bluffing they will meet with on the part of their opponents. Thus, both mechanisms I and II are likely to operate in a much weaker form if they operate at all. In the case of bargaining between Communist and non-Communist powers all of these difficulties seem to appear in rather extreme form.

We propose to conclude by suggesting some interesting problems for empirical research.

1. Is there any empirical evidence for the existence of *stereotype utility functions*, attributed in a given society to members of various social groups?

2. What is the relationship between these *stereotype utility functions* (if such exist) and the *true* utility functions of the relevant individuals?

3. How *definite*, how *realistic*, and how *consistent* with each other are two bargainers' ideas, usually, before the negotiations, about the terms they can eventually achieve?

4. What advantages do bargainers usually *expect* from step-by-step bargaining? Do they expect to obtain a good deal of information about their opponent's real attitudes? What criteria, if any, do they use to distinguish bluffs from serious bids? Do they expect the opponent to make the mistake of disclosing information which tends to weaken his own position?

5. How much information do bargainers *in fact* obtain about their opponent's true attitudes during the negotiations? To what extent will the opponent's bids *shift* their own expectations about the final outcome, and their own readiness to make concessions, under various conditions?

6. In all these respects, what are the differences between experienced and inexperienced negotiators, and again between negotiators dealing with opponents who come from the same cultural background, and negotiators dealing with opponents who come from a different culture?

[10] Another reason is mutual *distrust*, i.e., lack of belief in the other party's willingness to keep explicit agreements and/or tacit understandings, in particular in cases where no independent agencies exist to effectively *enforce* agreements.

REFERENCES

HARSANYI, JOHN C. "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 (1956), 144–57.

———. "On the Rationality Postulates Underlying the Theory of Cooperative Games," *The Journal of Conflict Resolution*, 5 (1961), 179–96.

LUCE, R. DUNCAN and ADAMS, E .W. "The Determination of Subjective Characteristic Functions in Games with Misperceived Payoff Functions," *Econometrica*, 24 (1956), 158–71.

NASH, JOHN F. "The Bargaining Problem," *ibid.*, 18 (1950), 155–62.

———. "Two-Person Cooperative Games," *ibid.*, 21 (1953), 128–40.

SCHELLING, THOMAS C. "Bargaining, Communication, and Limited War," *The Journal of Conflict Resolution*, 1 (1957), 19–36.

ZEUTHEN, FREDERIC. *Problems of Monopoly and Economic Warfare.* London: George Routledge & Sons, 1930.