61e Rejoinder

1. In his *Reply*, Roth (1986) adheres to the position he took in (1980), in which he said that for $0 < p < 1/2$, "(1/2, 1/2, 0) is the unique outcome ... consistent with the hypothesis that the players are rational utility maximizers." We find this position untenable. The case for (1/2, 1/2, 0) depends on a kind of mutual reliance between Players 1 and 2 that goes far beyond ordinary individual utility maximization, and that, because of its riskiness, may be totally unreasonable. This was explained in our *Comment* (Aumann (1985)), using informal as well as formal arguments; we will not repeat them here.

2. Roth complains that no one of our formal bargaining models predicts the value. We never claimed that they do; indeed, we emphasized that "the quantitative results (of the formal bargaining models) cannot be considered particularly significant ..." (Aumann (1985, p. 673)). These models were meant primarily to rebut Roth's position that (1/2, 1/2, 0) is the *only* reasonable outcome of these games; taken together, they also lend qualitative support to the value. Roth has no response to this, nor to any of the less formal reasoning that forms the bulk of our *Comment*.

3. Our own position is not dogmatic. For $0 < p < 1/2$, the unique value is (1/3, 1/3, 1/3), the unique core point (1/2, 1/2, 0). For small p the core looks reasonable, the value strange. As p grows, the value becomes more and more reasonable, the core stranger and stranger. For large p it is the value that is reasonable, the core strange. The Harsanyi solution (Hart (1985)) which yields $(1/2 - p/3,\ 1/2 - p/3,\ 2p/3)$ seems in these games more "sensible"—less "extreme"—than either the core or the value. Nevertheless, both the core and the value reflect important qualitative features of the games; one would not want to dispense with either one. The different outcomes that different solution concepts yield represent different approaches or viewpoints; they illuminate the problem from various angles.

4. When $p = 0$, we agree unequivocally that (1/3, 1/3, 1/3) is inappropriate. But just at this point, where Roth's case appears finally to become transparent, it vanishes. The outcome (1/3, 1/3, 1/3) is then no longer *the* value; it is only *a* value, another one being precisely the outcome (1/2, 1/2, 0) preferred by Roth. One can't base a compelling argument against the value on this kind of multivalent situation.

Roth objects that the value (1/2, 1/2, 0) is supported by weights (1, 1, 0); he says that zero weights are a "technical flaw," and implies that they

This paper originally appeared in *Econometrica* 54 (1986): 985–989. Reprinted with permission.

should not be used in conceptual discussion.[1] Zero weights are indeed associated with degeneracies; but they are in the game, not in the value. When $p = 0$, Player 3 can benefit no one but himself by joining a coalition. If one views the weight of a player as an endogenous measure of his importance and influence, "the weight he pulls" in Society, then in such a game it is quite natural to assign him weight 0. (See also (8) below.)

In brief, zero weights are associated with degenerate games, which it is best to avoid in conceptual discussion. But when they do occur, they cannot be dismissed, and are indeed quite natural.

5. When $p < 0$, the game is not superadditive; again, one cannot base compelling counterintuitive examples on such games. Roth calls them "perfectly playable." But if one does play them, then each player can assure 0 to himself, without the help of anyone else. In practice, therefore, each of the coalitions $\{1,3\}$ and $\{2,3\}$ can obtain (0, 0). With a coalitional form (characteristic function) that reflects this, (1/2, 1/2, 0) reappears as a value, and Roth's argument disappears.

6. Next, Roth brings up the old Maschler–Owen example, which involves completely different issues. We welcome the opportunity to discuss this.

Two variants, which we call V_0 and V_1, are at issue. In V_0, there are three players; by himself each one can get 0. If Players 1 and 3, or 2 and 3, form a coalition, then both get 0; if 1 and 2 form a coalition, then 1 gets a payoff of 1, and 2 gets nothing. All three together can share 1 in any way they please. The unique NTU value of this game is (1/2, 1/2, 0). This seems strange because Player 3 fulfills an important function in enabling 1 and 2 to share the amount 1 between them; while he need not get the same payoff as they, it certainly appears that he should get *something* for his services.

Before going on, we call attention to a four-person TU game, communicated to us by S. Zamir, in which the core behaves somewhat like the value of V_0, but seems even stranger. Define v on the player set $\{1,2,3,4\}$ by

$$v(S) = \begin{cases} 3 & \text{if } |S| = 4, \\ 0 & \text{if } |S \cap \{1,2,3\}| = 0 \text{ or } 1, \\ 2 & \text{otherwise.} \end{cases}$$

Intuitively, v is obtained from the majority game on $\{1,2,3\}$ by adding Player 4, who brings with him resources enabling all players together to get 3 rather than 2. The core of v consists of the single point (1, 1, 1, 0).

1. But when it suits his purposes, he himself uses zero weights with impunity; they appear explicitly in (Roth (1980)), as an important component of his attack on Harsanyi's solution.

This is even stranger than the value of V_0. There, one may feel that the additional player should be compensated for his services in enabling 1 to transfer payoff to 2. Here, the additional player brings tangible resources, which actually increase total revenue by a significant amount. Yet he gets no part of this revenue.

What is happening is that without Player 4, the game is coreless. The three players must either vie for a spot on a two-player coalition, knowing that one of them will be left in the cold, or agree to a compromise that yields any two of them less than what they can get for themselves. Player 4's contribution is just enough to make the core nonempty; it is totally gobbled up by 1, 2, and 3, on the pretext that any two of them thereby get no more than what they could have gotten previously. But only *two* of the three could previously have gotten 1; to conclude for this reason that now *each one* of them should get 1 strikes us as absurd.

We do not wish to disparage the core. The usefulness of a solution concept is measured not by its behavior in contrived examples, but by the insights it yields into social models of some generality. In this respect, both the core and the value have rather good track records.

But in fact, the value of V_0 is not all that strange. Suppose you and your brother are bequeathed a house located in your town. You wish to take possession, and to send your brother half its value; to this end, you ask your bank to make the transfer. What should the bank's fee be? Most people would suggest a relatively small fixed sum, or perhaps a few promil of the amount to be transferred. Very few would suggest anything substantial.

It may be objected that there is competition among the banks, so they cannot take too much; at worst, you *could* bring the money by hand. But even if there is only one bank, and you are somehow prevented from bringing the money by hand, many people would be appalled if the bank took any substantial proportion. This view is expressed by the value, which is a measure of a player's contribution to the social product (see (8) below). The transfer under discussion does not change the total social product; enabling it makes no substantial, measurable contribution, such as is made by Player 4 in Zamir's game v. While this is not the only possible view, it is not an unreasonable one. On the other hand, the core of v makes no sense at all to us.

Even if one insists that the bank should get a positive proportion, it is not clear how much. The game has no obvious symmetries; any positive amount seems possible. Thus at worst, the value appears as an extreme or limiting point of the "reasonable outcomes." That can't be considered a compelling counterintuitive example.

We come next to V_1, which differs from V_0 only in that the coalition $\{1, 2\}$ gets $(2, -1)$ rather than (1, 0). As in V_0, the value is (1/2, 1/2, 0). Roth argues that the payoffs to $\{1, 2\}$ are individually irrational, and should therefore be ignored. What remains is symmetric in all players, so the value "should" be (1/3, 1/3, 1/3).

This argument applies not to the value, but to formulations of the coalitional form that involve individual irrationalities. If one insists that individual irrationalities cannot occur in practice and that their prospect can have no effect on the final outcome, one should exclude them to start with. This procedure is quite standard in Game Theory; applied to V_1, it indeed yields the value (1/3, 1/3, 1/3). That presents no problem for the value.

There is, however, another view, in which individual irrationalities do play a significant role. The statement that "Player 2 can guarantee himself 0" can be interpreted to refer to the *beginning* of bargaining only. During the course of bargaining, he may make commitments which, while undertaken with the expectation of profit, *can* also lead to loss; and if they do, he cannot at that time renege and go back to 0. For example, consider a two-person bargaining game in which the individually rational levels are 0, and the two players together can get either $(2, -1)$ or $(-1, 2)$. In the "standard" solution, each player has an expected payoff of 1/2, based on a coin-toss between $(2, -1)$ and $(-1, 2)$; the price of the positive *expectation* is a commitment to accept a negative, individually irrational payoff if the toss goes against you. Similarly, in the course of bargaining, players may wish to commit themselves to joining certain coalitions under certain circumstances; or they may find it advantageous to forego certain options that they have. "Strategic risks" of this kind are in principle quite similar to the coin-toss in the above example. They are risky because one cannot be sure how the other players will respond; and while made in the expectation of profit, they may lead to loss, and even to individually irrational outcomes.

In V_1, suppose 2 commits himself to joining a coalition with 1, on pain of paying a fine in the amount of 1. Then 2's individually rational level drops to -1, and V_1 is transformed into a game V_0^* that is strategically equivalent to V_0. Reasoning as in V_0, we conclude that in V_0^*, Player 3 "should" enable 1 to make transfers to 2, without expecting a substantial fee. Thus 1 and 2 may expect to share the amount of $2 + (-1) = 1$. Since each one can get 0 at the beginning of the game, it seems reasonable for them to share this amount 1/2 – 1/2; i.e., for 2 to extract, in return for his commitment, a promise from 1 to transfer to him the amount 3/2, if (or when) 3 will allow it. This yields precisely the value.

To sum up, in analyzing V_1 one must first decide whether to exclude individual irrationalities at the outset, or to recognize them as important elements of the dynamics of bargaining. In the former case, we agree that the outcome should be (1/3, 1/3, 1/3), and indeed this is the value of the appropriately adjusted form of V_1. In the latter case, Roth's symmetry argument vanishes, and the value (1/2, 1/2, 0) appears as reasonable as in V_0.

7. The last example in the *Reply* is Shafer's, which we have already discussed (Aumann (1985, Section 8)). Roth finds it strange that there are exchange economies in which the value allocation yields one of the traders more of every commodity than his endowment. This is certainly interesting, but on reflection, not so strange. The utility function of the trader in question is "flatter" than that of the others; he is less particular, less risk-averse. Merchants like that have parlayed shoestrings into fortunes; the very willingness to take risks confers an advantage, even when not actually taken.[2] It is a strength of the value, not a weakness, that it reflects subtleties such as the effects of the utility function on bargaining strength.

8. Unable to make a compelling case with examples, Roth turns to the *definition* of value as such. He argues that since it allows Transfers of Utility, the endogenous TU game appearing in the definition has no clear relevance to the original NTU game.

This sounds cogent at first, but it does not survive examination. A TU game v is a function that associates with each coalition S a real number $v(S)$. One interpretation of $v(S)$ is as a sum of money (or other transferable good) that S may divide among its members in any way it wishes; this gives rise to the appellation "TU." But it is not the only interpretation. Even when utility is not transferable, one can take $v(S)$ simply as an appropriate numerical measure of the worth of S. That is what the NTU value does.

Measures or indexes like this are quite common in economics, accounting, and indeed all walks of life. A person's (or firm's) net worth is a sum of worths of assets that are very different from each other, e.g. in their liquidity; the total dollar figure is operationally meaningless, yet conveys important information. Similar remarks hold for GNP, national debt, price indexes, the mean of a distribution, the Gini index of inequality, and so on. Economic theory (e.g. taxation or growth theory) often uses social welfare functions—simple sums of utilities—even though util-

2. Suppose two agents must divide 6 dollars, with conflict payoffs of 0. If their utility functions are x and $\sqrt{x}$, the Nash bargaining solution yields 4 dollars to the risk neutral agent, and 2 dollars to the risk averse one, even though no overt risks are taken by any one.

ity is not transferable; nobody bats an eyelash. In Game Theory, the most frequent application of the TU value is to voting, committees and so on; coalitions S are assigned worth 1 if they can win, 0 if they can't, even though utility is not transferable.

The value of a player is meant as a measure of his social productivity, his contribution to the total social product. The explicit formula involves expected contributions to coalitions S that he may join. None of this requires transferable utility to make sense.

For the measure of S's worth, the maximum of a weighted sum of utilities—a kind of social welfare function—seems eminently reasonable. The weights are chosen so that the resulting value is feasible; an infeasible result would indicate that some people are overrated (or underrated), much like an imbalance between supply and demand indicates that some goods are overpriced (or underpriced). Note that zero weights fit very naturally into this picture, much like zero prices.

We do not claim that this is the only possible formulation of the intuitive concept in question. But it is reasonable and natural enough, and certainly does not suffer from the ills that Roth attributes to it.[3]

9. To conclude, not one single "clean," convincing counterintuitive example to the value has been adduced. The one remaining oddity—for indeterminately small positive p—comes nowhere near in sharpness and force to existing counterintuitive examples to other solution concepts such as the core, and cannot be considered a serious challenge to the value. The attacks on the definition of the value as such also fail to hold up.

From all this, the NTU Shapley value emerges stronger than ever. In applications, it often has substantial intuitive content; not infrequently, it yields important, unexpected insights (see the references of our *Comment*). It has several quite different characterizations, all of them intuitively meaningful; and it enjoys many important relationships with other solution concepts of economics and game theory. Its domain is very broad—it is almost always nonempty, in political contexts as well as all kinds of economic ones. On the other hand, it is "sharp," almost always containing only finitely many points, often quite few. And as we have seen, it is associated with unusually few anomalies or conceptual difficulties.

In brief, the value emerges as an eminently robust and useful cooperative solution concept. Al Roth's thoughtful probing has contributed significantly to this development.

3. Myerson (1986) provides a very pretty alternative rationale, both for the general definition of the NTU value, and for its realization in the Maschler–Owen game V_0.

Acknowledgments

This work was supported by National Science Foundation Grant SES83-20453 at the Institute for Mathematical Studies in the Social Sciences (Economics) at Stanford University, by the Institute for Mathematics and its Applications at the University of Minnesota, and by the Mathematical Sciences Research Institute, Berkeley. Technical services (typing) on some earlier versions were graciously provided through the good offices of A. Roth, and funded by the Office of Naval Research under Contract N 00014-84-K-0263 at the University of Pittsburgh. We are very grateful to S. Zamir for permission to publish his counterintuitive example to the core (Section 6); and to D. Kreps, the co-editor in charge of this paper, for several helpful suggestions, and especially for pointing out an error in an earlier version. We alone are responsible for all remaining errors.

References

Aumann, Robert J. (1985): "On the Non-transferable Utility Value: A Comment on the Roth–Shafer Examples," *Econometrica*, 53, 667–677 [Chapter 61c].

Hart, S. (1985): "Non-transferable Utility Games and Markets: Some Examples and the Harsanyi Solution," *Econometrica*, 53, 1445–1450.

Myerson, R. (1986): "An Introduction to Game Theory," in *Studies in Mathematical Economics*, ed. by S. Reiter, Vol. 25 of MAA Studies in Mathematics. Washington, D.C.: Mathematical Association of America, 1–61.

Roth, A. (1980): "Values for Games Without Side Payments: Some Difficulties with Current Concepts," *Econometrica*, 48, 457–465 [Chapter 61a].

———(1986): "On the Non-transferable Utility Value: A Reply to Aumann," *Econometrica*, 54, 981–984 [Chapter 61d].