GAMES AND ECONOMIC BEHAVIOR **24**, 181–210 (1998)
ARTICLE NO. GA970612

On the State of the Art in Game Theory: An Interview with Robert Aumann*

Eric van Damme, Interviewer

Q: We could take your talk yesterday on ''Relationships'' as a starting point. That talk is a follow-up to your paper ''What is Game Theory Trying to Accomplish,''[1] which you presented in Finland some years ago. Selten acted as a discussant for that paper. He agreed with some of the points you made, but disagreed with others. It would be nice to make an inventory of what has happened since that time. Have your opinions gotten closer together or have they drifted further apart? Do you agree that this is a good starting point for the interview?

A: Fine.

Q: Let us recall your position. Yesterday you said that science is not a quest for truth, but a quest for understanding. The way you said this made it clear that you have to convince people of this point; not everybody agrees with your point of view, maybe even the majority does not agree. Is that true?

A: You are entirely right. The usual, naive, view of science is that it is a quest for truth, that there is some objective truth out there and that we are looking for it. We haven't necessarily found it, but it is there, it is independent of human beings. If there were no human beings, there would still be some kind of truth. Now I agree that without human beings there

* This interview took place one day after the close of the ''Games '95'' Conference, which was held in Jerusalem, June 25–29, 1995. There are several references in the interview to events occurring at the conference, in particular to Paul Milgrom's lecture on the spectrum auction on June 29, to Aumann's lecture on ''Relationships'' on June 29, and to Aumann's discussion of the Nobel Prize winners' contributions on Sunday evening, June 26. The interviewer was Eric van Damme, who combined his own questions with written questions submitted by Werner Güth. Most of the editing of the interview was done at the State University of New York at Stony Brook, with which Aumann has a part-time association. Editing was completed on June 19, 1996. This interview originally appeared in *Understanding Strategic Interaction*, *Essays in Honor of Reinhard Selten* (Wulf Albers, Werner Güth, Peter Hammerstein, Benny Moldovanu, and Eric van Damme, Eds.), published by Springer-Verlag, Berlin. We are grateful to Springer-Verlag for their permission to reprint it in *Games and Economic Behavior*.

[1] In *Frontiers of Economics* (K. Arrow and S. Honkapohja, Eds.), Oxford: Basil Blackwell, 1985, pp. 28–76.

0899-8256/98

would still be a universe, but not the way we think of it in science. What we do in science is that we organize things, we relate them to each other. The title of my talk yesterday was ‘‘Relationships.’’ These relationships exist in the minds of human beings, in the mind of the observer, in the mind of the scientist. The world without the observer is chaos, it is just a bunch of particles flying around, it is ‘‘Tohu Vavohu’’—the biblical description of the situation before the creation.[2] It is impossible to say that it is ‘‘true’’ that there is a gravitational force: the gravitational force is only an abstraction, it is not something that is really out there. One cannot even say that energy is really out there, that is also an abstraction. Even the idea of ‘‘particle’’ is in some sense an abstraction. There certainly are individual particles, but the idea that there is something that can be described as a particle—that there is a class of objects that are all particles—that is already an abstraction, that is already in the mind of the human being. When we say that the earth is round, roundness is in our minds, it does not exist out there, so to speak. In biology, the idea of ‘‘species’’ is certainly an abstraction, and so is that of an individual. So all of science really is in our minds, it is in the observer’s mind. Science makes *sense* of what we see, but it is not what is ‘‘really’’ there.

Let me quote from my talk yesterday: Science is a sort of giant jigsaw puzzle. To get the picture you have to fit the pieces together. One might almost say that the picture *is* the fitting together, it is not the pieces. The fitting together is when the picture emerges.

Certainly, that is not the usual view. Most people do not see the world that way. Maybe yesterday’s talk gave people something to think about. The paper ‘‘What is Game Theory Trying to Accomplish?’’ had a focus that is related, but a little different.

Q: What I recall from that paper is that you also discuss ‘‘applied’’ science. You argue that from understanding follows prediction and that understanding might imply control, but I don’t recall whether you use this latter word. Perhaps you use the word engineering. Your talk yesterday was confined to understanding: what about prediction and engineering?

A: We heard a wonderful example of that yesterday: Paul Milgrom talking about the work of himself and Bob Wilson and many other game theorists in the United States—like John Riley, Peter Cramton, Bob Weber, and John Macmillan—who were asked by the FCC (Federal Communications Commission) or one of the communications corporations to consult on the big spectrum auction that took place in the US last year,

[2] Genesis 1, 2. ‘‘Tohu’’ is sometimes translated as ‘‘formless.’’ It seems to me that the idea of ‘‘form’’ is in the mind of the observer; there can be no form without an observer.

and that netted close to 9 *billion* dollars.[3] This is something big and it illustrates very well how one goes from understanding to prediction and from prediction to engineering. Let me elaborate on that. These people have been studying auctions both theoretically and empirically for many years. (Bob Wilson wrote a very nice survey of auctions for the Handbook of Game Theory.[4]) They are theoreticians and they have also consulted on highly practical auction matters. For example, Wilson has consulted for oil companies bidding on off-shore oil tracts worth upwards of 100 million dollars each. These people have been studying auctions very seriously, and they are also excellent theoreticians; they use concepts from game theory to predict how these auctions might come out, and they have gotten a good feel for the relation between theory and what happens in a real auction. Milgrom and Weber[5] once did a theoretical study, looking at what Nash equilibrium would predict on a qualitative basis for how oil lease auctions would work out, and then Hendricks and Porter[6] did an empirical study and they found that a very impressive portion of Milgrom's predictions (something like 7 out of 8) were actually true; moreover these were not trivial, obvious predictions. So these people make predictions, they have a good feel for theory, and they have a good feel for how theory works out in the real world. And then they do the engineering, like in the design of the spectrum auction. So one goes from theory, to prediction, to engineering. And as we heard yesterday, this was a tremendously successful venture.

Another example of how understanding leads to prediction and to engineering comes from cooperative game theory; it is the work of Roth and associates on expert labor markets. This is something that started theoretically with the Gale−Shapley[7] algorithm, and then Roth found that this algorithm had actually been implemented by the American medical profession in assigning interns to hospitals.[8] That had happened *before*

[3] See P. Milgrom, *Auction Theory for Privatization,* Oxford: Oxford University Press (forthcoming).

[4] R. Wilson, (1992) "Strategic Analysis of Auctions," in *Handbook of Game Theory, Vol.* 1, (R. Aumann and S. Hart, Eds.) Amsterdam: Elsevier, pp. 227−279.

[5] "The Value of Information in a Sealed Bid Auction," *Journal of Mathematical Economics* **10** (1982), 105−114; see also R. Engelbrecht−Wiggins, P. Milgrom, and R. Weber (1983), "Competitive Bidding and Proprietary Information," *Journal of Mathematical Economics* **11**, 161−169.

[6] "An Empirical Study of an Auction with Asymmetric Information," *American Economic Review* **78** (1988), 865−883.

[7] D. Gale and L. Shapley (1962), "College Admissions and the Stability of Marriage," *American Mathematical Monthly* **69**, 9−15.

[8] A. Roth (1984), "The Evolution of the Labor Market for Medical Interns and Residents: A Case Study in Game Theory," *Journal of Political Economy* **92**, 991−1016.

Gale and Shapley published their paper. It had happened as a result of 50 years of development. So there was an empirical development, something that happened out there in the real world, and it took 50 years for these things to converge; but in the end, it did converge to the game theoretic solution, in this case the core. Now this is amazing and beautiful, it is empirical game theory at its best. We're looking at something that it took smart, highly motivated people 50 years to evolve. That is something very different from taking a few students and putting them in a room and saying, OK, you have a few minutes to think about what to decide. These things are sometimes quite complex and take a long time to evolve; people have to learn from their mistakes, they have to try things out.

This was the beginning of a very intensive study by Roth and collaborators, Sotomayor and others,[9] into the actual working of expert labor markets. And now they are beginning to consult with people and say, Listen, you guys could do this and that better. So again we have a progression from theory to empirical observation to prediction and then to engineering.

Q: From these examples, can one draw some lessons about the type of situations in which one can expect game theory to work in applications?

A: What one needs for game theory to work, in the sense of making verifiable (and falsifiable!) predictions, is that the situation be structured. Both the auctions and the markets that Roth studies are highly structured. Sometimes when people interview me for the newspapers in Israel, they ask questions like, can game theory predict whether the Oslo agreement will work or whether Saddam Hussein will stay in power. I always say, those situations are not sufficiently structured for me to give a useful answer. They are too amorphous. The issues are too unclear, there are too many variables. To say something useful we need a structured situation. Besides the above examples, another example is the formation of governments. For years I have been predicting the government that will form in Israel once the composition of the Israeli parliament is known after an election. That is a structured situation, with set rules. The parliament has 120 seats. Each party gets a number of seats in proportion to the votes it got in the election. To form a government, a coalition of 61 members of parliament is required. The president starts by choosing someone to

[9] See A. Roth and M. Sotomayor (1990), *Two-Sided Matching*: *A Study in Game-Theoretic Modeling and Analysis,* Econometric Society Monograph Series, Cambridge: Cambridge University Press; A. Roth (1991), "A Natural Experiment in the Organization of Entry-Level Labor Markets: Regional Markets for New Physicians and Surgeons in the United Kingdom," *American Economic Review* **81**, 415–440; and A. Roth and X. Xing (1994), "Jumping the Gun: Imperfections and Institutions Related to the Timing of Market Interactions," *American Economic Review* **84**, 992–1004.

initiate the coalition formation process. (Usually, but not necessarily, this ‘‘leader’’ is the representative of the largest party in parliament.) The important point is that the process of government formation is a structured situation, to which you can apply a theory. That is true also of auctions and of labor markets like those discussed above. They work with fixed rules, and this is ideal for application. Now, if you don’t have a structured situation, that doesn’t necessarily mean that you cannot say anything, but usually you can only say something qualitative. For example, one prediction of game theory is the formation of prices; that comes from the equivalence theorem.[10] That kind of general prediction doesn’t require a clear structure. On the other hand, the Roth example and the spectrum auction are examples of structured situations, and they are beautiful examples of applications.

Q: Could you explain what exactly you mean by a structured situation? Is it that there are players with well-defined objectives and well-defined strategy sets, that there is perhaps even a timing of the moves? In short, is a structured situation one in which the extensive or strategic form of the game is given?

A: No, that is not what is meant by ‘‘structured.’’ It means something more general. A structured situation is one that is formally characterized by a limited number of parameters in some definite, clear, totally defined way. That implies that you can have a theory that makes predictions on the basis of this formal structure, and you can check how often that theory works out, and you can design a system based on those parameters. That holds for auctions, and for Roth’s labor markets, and for the formation of a governmental majority in a parliamentary democracy. In each of those cases, important aspects of the situation are described by a formal structure, a structure with clear, distinct regularities. Of course, there are also aspects that are not described by this formal structure, but at least *some* of the important aspects are described by this formal structure.

That is not so for the Oslo agreement and it is not so for Saddam Hussein. Those situations are in no sense repeatable, there is no regularity to them, there is nothing on which to base a prediction, and there is no way to check the prediction if you make it. Of course the prediction could be right or wrong, but there is no way to say *how often* it’s right, because each prediction is a one-time affair, there’s no way to generalize.

[10] See, e.g., G. Debreu and H. Scarf (1963), ‘‘A Limit Theorem on the Core of a Market,’’ *International Economic Review* **4**, 235–246; R. Aumann (1964), ‘‘Markets with a Continuum of Traders,’’ *Econometrica* **32**, 39–50; R. Aumann and L. Shapley (1974), *Values of Non-Atomic Games*, Princeton: Princeton University Press; and P. Champsaur (1975), ‘‘Cooperation versus Competition,’’ *Journal of Economic Theory* **11**, 394–417.

For instance, in the governmental majority matter, one can set up a parliament as a simple game in the sense of Von Neumann and Morgenstern's cooperative game theory, where we model the party as a player; we get a coalitional worth function that attributes to a coalition the worth 1 when it has a majority and the worth 0 otherwise. And then one can work out what the Shapley values are; the structure is there, it is clear, and one can make predictions. Now there are all kinds of things that are ignored by this kind of procedure, but one can go out and make predictions. Then, if the predictions turn out correct, you know that you were right to ignore what you ignored.

No matter what you do you are going to be ignoring things. This is true not only in game theory, it is true in the physical sciences also; there are all kinds of things that you are ignoring all the time. Whenever you do something out there in the real world, or observe something in the real world, you are ignoring, you are sweeping away all kinds of things and you are trying to say: Well, what is really important over here is this and that and that; let's see whether that's significant, let's see whether that comes out.

Q: In this example of coalition formation, you make predictions using an algorithm that involves the Shapley value. Suppose you show me the data and your prediction comes out correct. I might respond by saying that I don't understand what is going on. Why does it work? Why is the Shapley value related to coalition formation? Is it by accident, or is it your intuition, or is it suggested by theory?

A: There are two answers to that. First, certainly, this is an intuition that arises from understanding the theory. The idea that the Shapley value does represent power comes from the theory. Second, for good science it is not important that you understand it right off the bat. What is important, in the first instance, is that it is correct, that it *works*. If it works, then that in itself tells us that the Shapley value is relevant.

Let me explain this a little more precisely. The theory that I am testing is very simple, almost naive. It is that the leader—the one with the initiative—tries to maximize the influence of his party within the government. So, one takes each possible government that he can form and one looks at the Shapley value of his party within that government; the intuition is that this is a measure of the power of his party within the government. This maximization is a nontrivial exercise. If you make the government too small, let's say you put together a coalition government that is just a bare majority with 61 members of parliament—a minimal winning coalition—then it is clear that any party in the coalition can bring down the government by leaving. Therefore, all the parties in the government have the same Shapley value. So the hypothesis is that a wise leader won't do that. That is also quite intuitive, that such a government is unstable, and it finds its expression in a low Shapley value for the leader. On the other hand, too large a coalition is also not good, since then the leader doesn't have sufficient punch in the government; that also finds its expression in his Shapley value. Consequently, the hypothesis that the leader aims to maximize his Shapley value seems a reasonable hypothesis to test, and it works not badly. It works not badly, but by no means a hundred percent. For example, the current (June 95) government of Israel is very far off from that, it is basically a minimal winning coalition. In fact, they don't even have a majority in parliament, but there are some parties outside the government that support it; though it is really very unstable, somehow it has managed to maintain itself over the past 3 years. But I have been looking at these things since 1977, and on the whole, the predictions based on the Shapley value have done quite well. I think there is something there that is significant. It is not something that works 100% of the time, but you should know that nothing in science works 100% of the time. In physics also not. In physics they are glad if things work more than half the time.

Q: Would you like to see more extensive empirical testing of this theory?

A: Absolutely. We have tried it in one election in the Netherlands, where it seems to work not badly; but we haven't examined that situation too closely. The idea of using the Shapley value is not just an accident, the Shapley value has an intuitive content and this hypothesis matches that intuitive content.

Q: Are game theoretical applications to unstructured situations doomed to fail?

A: No, we already discussed the example of competitive equilibrium, the idea of price formation. It is successful although it is not structured. Even to diplomatic negotiations, which we discussed above, one can make contributions, in that one can point out certain things to be aware of. Let's consider an example. We have these negotiations with Syria. Now, the president of Syria, Mr. Assad, has stated publicly again and again that he is not sure that he is at all interested in reaching any kind of accommodation with Israel. The reason, he says, is that the status quo is not too bad for Syria; it is OK. And he is right, because we are not talking about a population in the Golan Heights, there are almost no Syrian people there. The few Syrian villages that were there and that were displaced in the 1967 war were resettled in Syria; no problems there, it was a small population. Most of the current inhabitants of the Golan Heights are Israeli Jews, all except for two or three villages right near the Syrian border, and there are no problems. These people are essentially satisfied. There is nothing that is burning under Assad’s pants over there, and really under anybody’s pants. So, Assad quite openly says, the status quo is not bad for me. Now, from the game theoretic point of view we have to ask ourselves, is the status quo in the set of feasible agreements? It could be that the status quo is outside the convex set of all possible pairs of utilities to agreements. It might be beyond the northeast boundary of that set. In that case there is no sense in pursuing a possible accommodation. Assad is quite happy the way it is; and while we would like to have a peace treaty with Syria, obviously that should not be at any cost. So the question is, is there a possible peace treaty that is worthwhile for Assad, and that simultaneously is acceptable to us? I don’t think that anybody in our government has answered that question, or even asked it. Professional diplomats don’t think in those terms. They think in terms that are totally different; they don’t have to do with payoffs, with outcomes. A diplomat will think in terms of vaguely defined objectives, like building trust between leaders, or peace, or signing a document, or making a statement to the press. This kind of thing is a totally different world. In this case, while game theory cannot make a prediction, it can say, let us think in certain terms, in certain directions.

By the way, here again the discussion centers on the cooperative theory. It is the idea of a set of possible accommodations, and a status quo point, and where is that status quo point in relation to the set of possible agreements. That is cooperative game theory.

That brings to mind another point. I don’t remember who it was, perhaps Kissinger, somebody in the early fifties said that a major contribution of game theory is just the idea of a strategy (or payoff) matrix. Just have the matrix in front of you. Just say, look, we can do something, and they can do something; let’s write down everything that we can do, and everything that they can do, and then let’s look it over from that point of view. Whereas before game theory, people had only been thinking about what *we* can do. The example of Syria and Israel is similar in that respect, because it says, can we reach an accommodation that we can buy and they can also buy? Or, among all the accommodations that will be acceptable to them, will there be one that is acceptable to us? So we are putting ourselves in both shoes at the same time. That is already a giant step forward. So in this kind of unstructured situation, game theory does have something to contribute. But it is conceptual: ways of thinking, approaches —not specific predictions.

Q: I think that one of the persons who wrote in the fifties that game theory’s most important contribution was the introduction of the payoff matrix was Thomas Schelling, and perhaps he provides a nice, natural way to move to the next topic. Specifically, I would like to move on to disagreements between you and Reinhard Selten, differences of viewpoint.

Yesterday you stressed the importance of relationships, of establishing links between the complex and the simple. Now one might argue that, perhaps, relationships are the more easy to establish the more abstract the setting in which one is working. If I interpret Selten's discussion of your earlier paper correctly, then he fears somewhat that by moving to a more abstract level one is going to lose the connection with the real world. The reason I thought this relates to Schelling is that Schelling has argued in the past that game theory should stand on two strong legs, one is a strong empirical leg and the other is a deductive, formal leg. You are stressing very much this second leg, of establishing formal relationships, whereas Schelling and Selten would argue that we need to develop more the empirical leg. Do you see a conflict there?

A: No, that is a misunderstanding. The term ‘‘relationship’’ does not at all imply that we are talking about theory only. Empirics are vital, and I agree entirely with the view that you just attributed to Schelling. Up to now in this interview we have discussed *only* empirics, so the importance that I attribute to that side should be clear.

Let me point out that at least three major areas of my talk yesterday touched on the importance of empirics. The very first item in my talk was an empirical item. The Naval Electronics Problem[11] was my entrance into decision theory and from there into game theory. The Naval Electronics Problem is a highly empirical problem with both legs in the real world; we are talking about real pieces of naval equipment worth hundreds of thousands of 1955 dollars each. We are talking altogether about billions of 1995 dollars, and we went in there and we used methodology from decision theory, and from linear programming, and some methodology that Joe Kruskal and I developed for this very purpose ourselves, which later turned out to be connected with utility theory. We went in there with these tools, which are fairly sophisticated, to solve this real empirical problem; we did an engineering job, and I think it was a useful one. Yesterday I stressed the relationships within this engineering job, to go from simple hypothetical assignment problems to complex real ones. So that is the first empirical item.

The second one was my analysis of the concept of Nash equilibrium. You will recall that I used three slides. The black slide on the bottom, which had just the payoff matrix and the word ‘‘equilibrium’’ written on it, and two red slides, which I superimposed on the black slide, one after the other. The first red slide explained the payoff matrix in terms of conscious decision making, and the other one explained it in terms of evolution. In

[11] R. Aumann and J. Kruskal (1959), ‘‘Assigning Quantitative Values to Qualitative Factors in the Naval Electronics Problem,’’ *Naval Research Logistics Quarterly* **6**, 1–16.

one case the equilibrium was strategic, and in the other case it was a population equilibrium: two completely different concepts with the same mathematical structure. Now the evolutionary interpretation of Nash equilibrium is certainly closely related to empirics. Let me mention in this connection the astoundingly beautiful work of Selten's associate Peter Hammerstein,[12] who did this study of spiders in the American Southwest: a game theoretic study of the behavior of spiders in which mixed strategies turned up in precisely (or almost precisely) the right proportion when fitnesses were measured in terms of egg-mass; a mindblowing study! This is real empirics and that is empirical game theory.

The third item from yesterday's talk that is really empirically rooted is the study of the Talmud,[13] using the idea of consistency and using the idea of the nucleolus. There we take something that happened 2000 years ago and happened in the real world; law is part of the real world. We are talking about a document that was written by people who had no idea of the discipline of game theory. They had something in mind, for 2000 years nobody understood it, and now it is understood. That is empirics. They were talking about a real bankruptcy problem, they had a real decision problem. This is a relationship, but not a theoretical relationship, it is something that is really happening out there. So that is the third area.

We're doing pretty well, when you take into account also all the other examples that we have mentioned, like the work about auctions, and about matching people to jobs, and the emergence of prices and many items we have not mentioned, like signalling,[14] and labor relations,[15] and other matters. The connection with the real world is indeed very important, and we have it. We have it and we are developing it further and it is closely associated with the theory; far from rejecting that, I embrace it. Many of the relations I discussed yesterday were real world relations.

Having said that, I'll say something else also. At some level the theory has to be several stages in advance of empirics. So we could be doing very

[12] P. Hammerstein and S. Riechert (1988), ''Payoffs and Strategies in Territorial Contests: ESS Analysis of Two Ecotypes of the Spider *Agelenopsis Aperta*,'' *Evolutionary Ecology* **2**, 115–138.

[13] R. Aumann and M. Maschler (1985), ''Game Theoretic Analysis of a Bankruptcy Problem from the Talmud,'' *Journal of Economic Theory* **36**, 195–213.

[14] See, e.g., D. Kreps and J. Sobel (1994), ''Signalling,'' in *Handbook of Game Theory*, *Vol*. 2 (R. Aumann and S. Hart, Eds.), Amsterdam: Elsevier, pp. 849–867.

[15] See, e.g., J. Kennan and R. Wilson (1993), ''Bargaining with Private Information,'' *Journal of Economic Literature* **31**, 45–104; and R. Wilson (1994), ''Negotiation with Private Information: Litigation and Strikes,'' Nancy L. Schwartz Lecture, J. L. Kellogg School of Management, Evanston, IL: Northwestern University.

complicated theory, let us say equilibrium refinement a la Mertens[16] with cohomology and all that, and at the same time be doing empirical studies and empirical engineering like Milgrom was discussing (the spectrum auction). It is very important that we have those two levels, because it is not possible to do the things that Milgrom was discussing without having the Mertens cohomology at the same time. This is a somewhat subtle idea, so let me explain it. There has to be a body of knowledge out there that has been internalized by a body of people. Perhaps Milgrom, as an individual, does not necessarily have to understand Mertens's cohomology-based refinement theory, but the body of science has to develop in such a way that there can exist a Milgrom. (As it happens, Milgrom and Wilson are both very strong theorists; with all his involvement in the spectrum auction, Wilson has just written a highly theoretical piece about Mertens refinement.[17]) In some sense it wouldn't be possible for him to work out these very practical rules without knowing something like Mertens's cohomology refinement; perhaps not necessarily that, but something of similar depth. You have to swim in it. You are not going to be able to swim from the shore to an island that is a kilometer away if it's the first time you are in the water. It has to be a part of your background. As I said yesterday, these things are all hitched together; they are part of a milieu. You have to have a lot of experience, and this experience has to be both practical and theoretical. You can't keep your nose too close to the ground. There has to be some ‘‘spiel,'' some room, and there has to be some background, and the theory always has to be far in advance.

Q: The second point of difference of view might perhaps be the following. You seem to be saying that every relationship is good, the more relations—the tighter the web—the better. But some of these relationships might actually be misleading. For example, take the rationalistic and evolutionary interpretations of Nash equilibrium. Maybe one interpretation applies in one context, while the other applies in a completely different context. Insights relevant in one context need not necessarily be relevant in the other context. Nevertheless, the relation might be used to import insights from one context to the other. One shouldn't mix these things, one should keep them separate. Selten advocates a sharp distinction between descriptive game theory and rationalistic game theory, where

[16] J.-F. Mertens (1989), ‘‘Stable Equilibria—A Reformulation, Part I: Definition and Basic Properties,'' *Mathematics of Operations Research* **14**, 575–625; and J.-F. Mertens (1989), ‘‘Stable Equilibria—A Reformulation, Part II: Discussion of the Definition and Further Results,'' *Mathematics of Operations Research* **16**, 694–753.

[17] S. Govindan and R. Wilson (1996), ‘‘A Sufficient Condition for the Invariance of Essential Components,'' *Duke Journal of Mathematics* **81**, 39–46; see also R. Wilson (1992), ‘‘Computing Simply Stable Equilibria,'' *Econometrica*, **60**, 1039–1070.

the former is based on less strong rationality assumptions. Hence, this is related to the issue of bounded rationality in general. Could you comment on these points?

A: You say "one shouldn't mix these things . . . Selten advocates a sharp distinction." Well, I disagree. When there is a formal relationship between two interpretations, like the rationalistic (conscious maximization) and evolutionary interpretations of Nash equilibrium,[18] then one can say, "Oh, this is an accident, these things really have nothing to do with each other." Or, one can look for a deeper meaning. In my view, it's a mistake to write off the relationship as an accident.

In fact, the evolutionary interpretation should be considered the fundamental one. Rationality—Homo Sapiens, if you will—is a *product* of evolution. We have become used to thinking of the "fundamental" interpretation as the cognitive rationalistic one, and of evolution as an "as if" story, but that's probably less insightful.

Of course, the evolutionary interpretation is historically subsequent to the rationalistic one. But that doesn't mean that the rationalistic interpretation is the more basic. On the contrary, in science we usually progress, in time, from the superficial to the fundamental. The historically subsequent theory is quite likely to *encompass* the earlier ones.

The bottom line is that relationships like this are often trying to tell us something, and that we ignore them at our peril.

Having said this, let me add that game theorists argue too much about interpretations. Sure, your starting point is some interpretation of Nash equilibrium; but when one is doing science, you have a model and what you must say is: What do we *observe*? Do we observe Nash equilibrium? Or don't we observe it? Now, *why* we observe it, that's a problem that is important for philosophers, but less so for scientists. It's like the old joke of the town where they wanted to build a bridge over the river. They started arguing in the town council. The merchants wanted the bridge because it would be good for commerce. People from each side wanted it because it would be easier to go to the other side. The religious people wanted the bridge because then they could reach the synagogue more easily. The lovers wanted it because they could make romance on a full moon night. They started arguing and fighting with each other over why to build the bridge, and in the end they couldn't agree and they did not build the bridge.

So, the story of *why* we have Nash equilibrium, it's an important philosophical problem, but as science it is not that important. You are

[18] R. Aumann (1987), "Game Theory," in *The New Palgrave, Vol. II* (J. Eatwell, M. Milgate, and P. Newman, Eds.), London: Macmillan, pp. 460–482. See specifically Subsection ii of the section entitled 1970–1986, on page 477, near the bottom of the left column.

quite right, when we had a discussion this week, I was a little surprised when Reinhard said, ‘‘Yes, we must distinguish . . . .’’ I don’t quite understand why.

As far as descriptive game theory is concerned, I certainly agree that we must describe what we actually observe. But to divorce this from the theory would be counterproductive. After all, the theory is meant to *account* for what we observe. Imagine a theoretical physics that is divorced from observational physics—that would be absurd! There may be—indeed there *are*—discrepancies between theory and observation, both in physics and in game theory; but fundamentally, the aim must be to bring theory and observation together, not to distinguish between them. And as we have already discussed, in game theory we are doing quite well in this regard.

By the way, the evolutionary interpretation does not support only subgame perfect equilibrium, or, for that matter, trembling hand perfection. It also supports nonperfect equilibria. One beautiful example of that is in the ultimatum game. One can understand what is observed in the ultimatum game[19] on the basis of nonperfect equilibria. This is not an empty prediction. As you know, of course, every possible offer is a part of an equilibrium in the ultimatum game, so one could say that this does not mean very much; but that is not correct, because when you say that you have an equilibrium, then you say that this is in a sense a norm of behavior. That *can* be checked, and it *was* checked in the work of Roth, Prasnikar, Okuno-Fujiwara, and Zamir,[20] and they found regularities of behavior in various different places in the world, which were *different* in the different places. That is a beautiful verification of the evolutionary interpretation of Nash equilibrium. Far from the ultimatum game being some kind of rejection of game theory, it is a beautiful verification of it, a corroboration!

Q: You said yesterday that you thought that this was the direction to go in strategic game theory, that is, to develop further the evolutionary approach. At present, the models that we have are based on agents who do

[19] W. Güth, R. Schmittberger, and B. Schwarze (1982), ‘‘An Experimental Analysis of Ultimatum Bargaining,’’ *Journal of Economic Behavior and Organization* **3**, 367–388. In this experiment, two players were asked to divide a considerable sum (varying as high as DM 100). The procedure was that P1 made an offer, which could be either accepted or rejected by P2; if it was rejected, nobody got anything. The players did not know each other and never saw each other; communication was a one-time affair via computer. The only subgame perfect equilibria of this game lead to a split of 99–1 or 100–0. But there are many Nash equilibria, of the form ‘‘P1 offers x, P2 rejects anything below x.’’

[20] ‘‘Bargaining and Market Behavior in Jerusalem, Ljubljana, Pittsburgh, and Tokyo: An Experimental Study,’’ *American Economic Review* **81** (1991), 1068–1095. Roughly speaking, it was found in this experiment that in each of the four different venues, the accepted offers clustered around some specific x, which differed in the different venues.

not have any cognition at all. Do you view this as a first step towards other models in which there is limited cognition, towards richer models of bounded rationality?

A: No, actually not. That is muddying the waters. The evolutionary interpretation is best understood as standing on its own legs, not as a stepping stone leading to rational or cognitive interpretations. That would not be a good way to go. We have here two separate—but completely parallel, completely isomorphic—interpretations of the *same* mathematical system. A population equilibrium in the evolutionary interpretation corresponds *precisely* to a strategic equilibrium in the cognitive interpretation. In a population equilibrium, each genotype that constitutes a positive proportion of the population is *best possible* given the environment. Similarly, in a strategic equilibrium, each pure strategy with positive probability is *best possible* given the mixed strategies of the others. In both cases we have *full* maxima; there is no way that the equilibria in the evolutionary interpretation could somehow be improved if we made them ‘‘more rational.’’

In brief, one should not try to transform an evolutionary model first into a boundedly rational model and then into something like a full rationality model; that’s not the way to go. One should have both models in mind. Both are important.

Actually, the formal similarity between the two models probably reflects an important conceptual relationship. Rational thinking may have evolved *because* it is evolutionarily efficient. Let me clarify that. Evolutionary biologists keep asking about everything: What is the evolutionary function of this, what is the evolutionary function of that? One question, raised by Dawkins in his book[21] is: What is the evolutionary function of consciousness? What I’d like to suggest is that consciousness evolved because it serves a very important evolutionary function, namely that it enables people to think rationally in the sense of utility maximization. You are not only a computing machine; you have goals to achieve, utilities to maximize. For that you must be conscious, you must be able to experience, you must be able to say this is pleasurable, this is painful.

Just an example of that: What is the function of taste in food? Why must it have a taste? Well, it must have a taste because if it did not taste good we would not be motivated to eat it and then we would starve to death. We have to be conscious for that, we can’t taste something without being conscious. The consciousness enables us to experience pleasure and pain. It has long been recognized that pain has a very important evolutionary function. I’m suggesting that pleasure also does.

[21] *The Selfish Gene*, Oxford: Oxford University Press, 1976.

Another example is, of course, sex. Why is sex pleasurable? What is the evolutionary function of that? Not of sex itself, that's obvious, but of the *pleasure* in sex? It is to induce the organism to have sex. On a deeper level, then, it is quite possible that the combination of the logical thinking in the brain and the consciousness which enables us to have utility, to have desires, is an outcome of an evolutionary process. So rational thinking, purpose oriented, utility maximizing behavior is an *outcome* of the evolutionary process. Not the other way around—like we usually say that evolution is ‘‘as if’’ organisms were thinking! No, the thinking comes *because* of the evolution. What I'm trying to say is that in the end the two interpretations of Nash equilibrium really could be two sides of the same coin.

In his book, *The Growth of Biological Thought*,[22] Ernst Mayr distinguishes between two kinds of explanation in biology, corresponding to the questions ‘‘how’’ and ‘‘why.’’ Sometimes people confuse these terms. For example, the question ‘‘why do we see’’ might be answered in two ways. One is, ‘‘we see because we have eyes with lenses and retinas and neural connections to the brain.’’ The other is, ‘‘we see because it helps us enormously in getting along in the world.’’ Mayr says that the first answer really responds to the question ‘‘how do we see’’; the term ‘‘why’’ is more appropriate to the second answer. ‘‘how’’ refers to the mechanism, ‘‘why’’ to the function. Needless to say, both questions are legitimate, and both answers are correct.

What I'm suggesting is that to the extent that it really occurs, conscious, rational utility maximization is a *mechanism*, an answer to the ‘‘how’’ question. Consciousness, the desire to maximize, the ability to calculate—these are traits that have evolved, like eyes. Their *function* is to enable us human beings to adapt, in a flexible way, to the very complex environment in which we operate. Because of its complexity and variability, it would be much more difficult to operate in this environment with genetically predetermined, ‘‘hard-wired’’ modes of behavior.

Q: Let us now turn to the questions that Werner Güth prepared in writing. The first one is this: One can distinguish two major tasks of game theory, namely

—formally to represent situations of strategic interactions, i.e., to define adequate game forms, and

—to develop solution concepts specifying the individually rational behavior in such games.

A: Well, let me take issue with that right away. It is not only strategic interaction. Yesterday we were talking about cooperative and noncoopera-

[22] Cambridge, MA: Belknap Press, 1982.

tive game theory and I said that perhaps a better name for cooperative would be ‘‘outcome oriented’’ or ‘‘coalitional,’’ and for noncooperative ‘‘strategically oriented.’’ The way Güth’s question is set up already points to the strategic direction. That, of course, is a very important part of game theory, but it is just one side of it. Game theory develops not only solution concepts that specify rational behavior of individuals, but also solution concepts that specify outcomes. It is not just a question of behavior, but also of outcomes. And this is very very important, because in order to do game theory you would be very restrictive if you confined yourself to situations that are strategically well defined. In applications, when you want to do something on the strategic level, you must have very precise rules; for example, like in an auction. An auction is a beautiful example of this, but it is very special. It rarely happens that you have rules like that. What other situations do you have that are strategically well-structured? OK, sometimes you have a parliament where you have rules of order and rules of introducing bills; Ferejohn[23] has actually done some work on applications of noncooperative theory to political science in political situations that are strategically well-defined. Another example is strategic voting.[24] Those situations are basically strategic and there really the noncooperative side is very important. But such situations, with very precisely defined rules, are relatively rare.

Cooperative theory allows the rules of the game to be much more amorphous. If you wanted to do the Roth labor market, you could not do that noncooperatively, the rules are not sufficiently well specified. If you wanted to do a noncooperative analysis of coalition formation, you can’t do it, it is not sufficiently well specified. Who talks first, who talks second? It matters enormously in noncooperative game theory.

To sum it up, it is not just behavior that matters; it is also outcomes, and at least as much outcomes as behavior. So I disagree with the phrasing of the question.

Let’s go on and read the second sentence of Werner’s question.

Q: ‘‘In the social sciences one presently relies nearly exclusively on noncooperative game models.’’

A: But that’s absolutely incorrect. I can’t disagree more. There have been important analyses of noncooperative models, but as the discussion up to now has shown, many of the models that we have been talking about are cooperative and many of them are successful ones. Roth’s work on

[23] See, e.g., J. Ferejohn and J. Kuklinski (Eds.) (1990), *Information and Democratic Processes*, Urbana, IL: University of Illinois Press.

[24] Of course, as we mentioned, voting also has a cooperative side.

labour markets[25] is mindblowing, this is game theory at its best. If anything, the cooperative theory has played a *more* important role in applications than the noncooperative theory.

Q: Maybe social sciences should be narrowed down to economics and the question has been motivated by the fact that there are books about game theory that don't refer to the cooperative branch at all.

A: That is true, but it is unfortunate. The fact that there are books published doesn't mean that that is what the world is about. The people who write these books are missing some very important sides of game theory; they are representing only their own knowledge and their own interests. These books do not give a balanced summary of what game theory has accomplished.

By the way, there is an excellent book, recently published, by Osborne and Rubinstein.[26] These authors made most of their theoretical contributions on the strategic side, and yet they devote a nice portion of their book to cooperative game theory. I recommend this book highly, it is beautifully done, and it recognizes the importance of the cooperative theory.

Incidentally, the Rubinstein alternating offers model[27] is an example of a ‘‘bridge’’: a noncooperative, behavioral model leading to the Nash bargaining solution, which, as we know, is a cooperative, outcome-oriented concept. And this is true not only in TU-situations; it also leads to the Nash solution in an NTU-setup, as Binmore has shown.[28]

Summing up, one should look at what is significant, not at what is fashionable.

Q: Perhaps we can now move to the third and last part of this question. It reads: ‘‘Although the noncooperative focus brought about a very thorough analysis of many institutional aspects (e.g., of economic markets), it seems that nearly every phenomenon can be explained by designing an adequate model (e.g., by relying on an infinite time horizon as in the case of the so-called Folk Theorems).’’ Do you see this more as a problem (we need more selective predictions!) or as an advantage (we have to find out the relevant institutional details!)? Is there some hope for a revival of cooperative game theory, which avoids the specification of many facets?

[25] Op. cit. (Footnotes 8 and 9).

[26] *A Course in Game Theory*, Cambridge, MA: MIT Press, 1995.

[27] A. Rubinstein (1982), ‘‘Perfect Equilibrium in a Bargaining Model,’’ *Econometrica* **50**, 97–109.

[28] ‘‘Nash Bargaining Theory II,’’ in *The Economics of Bargaining* (K. G. Binmore and P. Dasgupta, Eds.), Oxford: Basil Blackwell, 1987, pp. 61–76.

A: There are several things to say to that. First, a word about institutions. Institutions can be treated as exogenous or endogenous. Güth's question (''we have to find out the relevant institutional details'') treats them as exogenous: Given this or that institution, what are the equilibria? Though this has some interest, it is more interesting to treat institutions *endogenously*. *Why* did a given institution come about? What are the underlying forces that led to its formation?

This dichotomy is discussed, for example, in the work by Jacques Dreze and me about rationing.[29] Before this work, rationing was treated exogenously: Given the institution of rationing, what parameter values would bring the economy into equilibrium? Dreze and I asked a different question: What led to the institution of rationing? If it is a matter of excess demand or supply, why specifically rationing, and what form of rationing? Could other institutions handle this?

An even more basic example is the equivalence theorem[30] itself. Rather than taking the institutions of money and prices as given, the equivalence theorem predicts the emergence of these institutions.

Note that both these examples come from cooperative game theory, so we are led naturally to your question about a ''revival'' of the cooperative theory. Cooperative theory is actually doing quite well. I've already said in this interview that many of the most interesting applications of game theory come from the cooperative side. In his invited talk at this conference,[31] Mike Maschler discussed over 30 significant contributions to the cooperative theory that have been produced over the last few years. Yes, I suppose one could call that a ''revival.'' And I agree with you that an important advantage of the cooperative theory is that it takes a broader, more fundamental view, that it is not so obsessed with procedural details; its fundamental parameters are the *capabilities* of players and coalitions.

Adam Brandenburger, who teaches at the Harvard Business School, has told me that the students there consider the cooperative theory a lot more relevant to business than the noncooperative theory. In my opinion, both are important, perhaps for different kinds of applications.

Let's come now to your remark, ''nearly every phenomenon can be explained by designing an appropriate model.'' First of all, as you say, that happens less with cooperative models. But beyond that, the importance of a model depends on the sum total of its applications, on how the specific

[29] ''Values of Markets with Satiation or Fixed Prices,'' *Econometrica* **54** (1986), 1271–1318.

[30] That is, the equivalence between competitive equilibrium in markets and game theoretic concepts like core and value.

[31] ''Games '95,'' Jerusalem.

kind of model of this kind has been applied in other places. One has to get a feel for the applications that are covered by a certain kind of model. Perhaps you could design a model for anything, but that does not mean that it is an interesting model. It is interesting if you design it, and then it has applications A, B, C, D; it applies to labor, it applies to auctions, it applies to search, it applies to signalling, it applies to discrimination. When you have a lot of applications, then you can start saying, well, this is an important model. This is what I tried to point out yesterday in my lecture on relationships, that when you get something which gives you important things in many different applications, like the Shapley value, or the nucleolus, then it gets some credibility.

Q: Your answer brings up a related issue. In the survey that you wrote for The New Palgrave[32] you describe the game theoretic approach as being very different from the classical approach in economics. The game theoretic approach is unifying: the same concept is applied in various contexts. In contrast, in the economic approach, a model and solution concept is tailored to the situation at hand. What we currently see in game theoretic applications to economics is something like a half-way house: There is a unifying solution concept—Nash equilibrium—but there is a great deal of freedom in constructing the extensive or strategic form model. There is a lot of flexibility.

A: Well, that applies to the noncooperative theory. On the cooperative side, there are three or four central solution concepts—value, core, nucleolus, stable sets—but much less flexibility in constructing the model. The model is much better defined.

Even on the noncooperative side, it’s not clear that Nash equilibrium can be called a ‘‘unifying’’ solution concept. What about all the refinements and other variations? We still have a way to go, the smoke still has not cleared on the refinement battlefield, or maybe it is not a battlefield, but a refinement factory. I don’t know what the product there is. We have a lot of refinements; which is the right one? Eric, actually what do *you* think? Now it is 1995, we are more than a decade into the refinement business, almost two decades. Do you see one or two or three refinements emerging as the accepted ones? What do you think?

Q: I think that some, like backward induction, subgame perfect equilibrium, are there to stay. Others, like proper equilibria, probably aren’t, since they don’t satisfy properties like invariance. As far as the variety of stability-like concepts that were introduced by Kohlberg and Mertens[33] is concerned, it is still unclear, the dust hasn’t settled yet. For

[32] Op. cit. (Footnote 18).

[33] ‘‘On the Strategic Stability of Equilibrium,’’ *Econometrica* **54** (1986), 1003–1034.

example, there is still discussion about the admissibility requirement, is it necessary for strategic stability or not? Should one really look at strategy perturbations as, for example, Nash did in his original work on bargaining?[34] The discussion is still open, I think. And then there is the concept of persistent equilibria, due to Kalai and Samet,[35] which I think belongs to a different category; it might show up in a learning context, in what Nash calls the ‘‘mass-action’’ interpretation of equilibria. That is, a learning process might converge to a persistent equilibrium.

A: Your answer is very useful; it underscores the difference between your viewpoint and mine, as representatives of different schools of thought.

For example, in discussing refinements just now, you were concerned about assumptions: admissibility is too strong, too weak, and so on. That is not my concern. I have never been so interested in assumptions. I am interested in conclusions. Assumptions don’t have to be correct; *conclusions* have to be correct. That is put very strongly, maybe more than I really feel, but I want to be provocative. When Newton introduced the idea of gravity, he was laughed at, because there was no rope with which the sun was pulling the earth; gravity is a laughable idea, a crazy assumption, it still sounds crazy today. When I was a child and was told about it, I could not believe it. It does not make any sense; but it does happen to yield the right answer. In science one never looks at assumptions; one looks at conclusions. Where do these refinements *lead*? Take an application, what do Kohlberg/ Mertens[36] have to say about this application? What do Kreps/ Wilson[37] have to say about this application? What do Kalai/ Samet[38] have to say about this application? What do Cho and Kreps[39] have to say about it? I don’t care what the assumptions are. So, I get a feel for what the concept says not by its definition but by where it leads. This is a fundamental difference between what I am trying to promote and the school that you are representing (which, of course, also has its validity).

Beyond that, I agree with you that what’s here to stay is subgame perfect equilibrium. That says a lot and we know what that says, but we don’t know it from the definition. We know it from the applications. What else is here to stay, I don’t know. I agree entirely with what Güth says that, given the

[34] ‘‘The Bargaining Problem,’’ *Econometrica* **18** (1950), 155–162.

[35] ‘‘Persistent Equilibria in Strategic Games,’’ *International Journal of Game Theory* **13** (1984), 129–144.

[36] Op. cit. (Footnote 33).

[37] D. Kreps and R. Wilson (1982), ‘‘Sequential Equilibrium,’’ *Econometrica* **50**, 863–894.

[38] Op. cit. (Footnote 35).

[39] ‘‘Signaling Games and Stable Equilibria,’’ *Quarterly Journal of Economics* **102** (1987), 179–221.

situation, there is a lot of room for constructing a game tree, and it matters very much how you construct the game tree and I think that that is one of the problems. And that limits the applicability of the strategic approach altogether. It does not eliminate it, it has very important applications, but it limits it.

Q: As far as applications are concerned, one can point to signalling games. For example, take Spence's original model of education as a signal.[40] The game model of that situation has many equilibria, but the stable equilibrium outcome corresponds with the unique one that was originally singled out by Spence. I believe we can empirically verify the predictions of this equilibrium, for example, the overinvestment in the signal. I think we could also verify these predictions in other signalling contexts, for example, in financial markets.

A: That's the kind of result I am looking for. When you ask what are the right refinements, you have to say, what are the refinements that give us the Spence signalling equilibrium? That's what we have to ask.

Q: Well, in order to get that answer you have to apply strong solution concepts, you need basically the full strength of the Kohlberg/ Mertens[41] 1986 Econometrica stability concept.

A: Fine, we need more applications like that. We need to have applications out there, not just Beer−Quiche; that's very important, but it is an example. We need a class, a kind of model to which to apply these things. By the way, there is an excellent article of Kreps and Sobel[42] on signalling in the Handbook. There are a lot of important applications out there, there is no question about it. But, as I said, the smoke has not cleared yet. It hasn't cleared on the assumptional side and it hasn't cleared on the conclusional side.

Q: What is definitely a drawback of some of these solution concepts is that they are very difficult to work with. Just as it is difficult to work with Von Neumann−Morgenstern stable sets in the cooperative theory, it is also difficult to see what the implications of the Mertens[43] stability concept are.

A: That is a very important point and it is one that I endorse entirely. It is something that I also stressed in the article ‘‘What is Game Theory Trying to Accomplish?’’: It is important to have an applicable model. It sounds a little like the man who had lost his wallet and was looking under

[40] A. Spence (1974), *Market Signaling*, Cambridge, MA: Harvard University Press.
[41] Op. cit. (Footnote 33).
[42] Op. cit. (Footnote 14).
[43] Op. cit. (Footnote 16).

the lamppost for it. His friend asked him: Why do you look only under the lamppost? And he answered: That's because there is light there, otherwise I wouldn't be able to see anything. It sounds crazy, but when you look at it more closely it is really important. If you have a theory that somehow makes a lot of sense, but is not calculable, not possible to work with, then what's the good of it? As we were saying, there is no ''truth'' out there; you have to have a theory that you can work with in applications, be they theoretical or empirical.

Incidentally, there is a man whom I love and respect very much and with whom I have worked closely and extensively, but who disagrees with me on this issue, and that is Mike Maschler. He has the opposite opinion. He wants to know what the ''truth'' is, and if it is difficult to calculate, he doesn't care, he has to find the truth. I don't have this notion of truth. It is really odd to see him wanting to find the ''right'' bargaining set. He has this idea that some of the bargaining sets are wrong and some of them are right. I don't have this idea. I have the idea of usefulness, not right or wrong. So, he would disagree with me. He says, if it is hard to calculate, it is just too bad, but we must find out what it is. My own viewpoint is that inter alia, a solution concept must be calculable, otherwise you are not going to use it.

Q: Whereas Reinhard Selten has tried to refine the equilibrium concept, as originally proposed by Cournot and Nash, you have developed your well-known coarsening idea of correlated equilibria.[44] Do you think that solution concepts should allow for more behavioral possibilities rather than select more and more specifically?

A: Well, my response to that is that the word ''should'' is out of place. I can't answer the question because I disagree with its formulation. This is again a somewhat different viewpoint from that of many people. Game theory is not a religion. In religion one says, one should observe the Sabbath, one should give charity. But game theory is not a religion, so the word ''should'' is out of place. I love correlated equilibria and I also love subgame perfect equilibria; I have written papers, which I hope the world will enjoy, on both subjects. A few years ago I wrote about correlated equilibria, and recently I published a paper[45] called ''Backward Induction and Common Knowledge of Rationality.'' Backward induction, of course, is a subgame perfect equilibrium. So I do both, and I think that one ''should'' not do this or that exclusively, but one should develop both concepts and see where they lead. There are important things to say about correlated

[44] ''Subjectivity and Correlation in Randomized Strategies,'' *Journal of Mathematical Economics* **1** (1974), 67–96.

[45] *Games and Economic Behavior* **7** (1995), 6–19.

equilibria and there are important things to say about subgame perfect equilibria. This idea that correlated equilibria represent the truth, or that subgame perfect equilibria represent the truth, is an idea that I reject.

Q: If game theory is not religion, is it a tool?

A: Yes, absolutely, it is science. It is a tool for us to understand our world.

Q: Let us now move on to the notion of consistency, which played an important role in your talk yesterday. It is a property that is satisfied by many solution concepts, among them Nash equilibrium: If we fix some of the players in the game at their equilibrium strategies, then the combination of equilibrium strategies for the remaining players constitutes an equilibrium of the reduced game. When writing your paper for The New Palgrave[46] you were apparently already aware that this property could be used to axiomatize the concept. I would like to discuss this property in connection with equilibrium selection. There is a paper[47] that shows that no solution concept that assigns a unique equilibrium to each finite strategic game can be consistent. I don't know how to deal with this.

A: Well, that is certainly a very interesting result, and it is also important. But I can't say that it is distressing. If there is one thing that we have learned from axiomatics, it is that if you write down all things that are obviously something that you would want, then you almost always find a contradiction. So let me start by saying that I won't lose sleep over that.

Having said that, let me respond substantively. Equilibrium selection in the style of Harsanyi and Selten[48] is a beautiful venture, it is very important. It is important for two reasons. One is that, from one point of view, equilibrium selection is an important facet of the background of equilibrium. Game theory is not a religion, and, equilibrium does not have to be ‘‘justified’’; but for those people who do want to justify it, including Harsanyi and Selten, selection plays a fundamental role. (Let me say that though I hold certain methodological viewpoints, that does not mean that I reject the importance of other people's methodological viewpoints. I have said that assumptions are not important, but I can understand also people who argue that assumptions *are* important, and who want to justify equilibria, and who do think that game theory is a way of life, that's OK also.) So in that respect—to ‘‘justify’’ equilibria—selection theory *is*

[46] Op. cit. (Footnote 18).

[47] H. Norde, J. Potters, H. Reynierse, and D. Vermeulen (1996), ‘‘Equilibrium Selection and Consistency,’’ *Games and Economic Behavior* **12**, 219–225.

[48] J. Harsanyi and R. Selten (1988), *A General Theory of Equilibrium Selection in Games,* Cambridge, MA: MIT Press.

important. And it has other important sides to it, by-products: risk dominance is a very important by-product of selection theory, and so is the tracing procedure.

On the other hand, selection theory has a sort of fantastic aura to it. Hundreds of pages, filled with extremely complex instructions as to how one picks one equilibrium in each game. And, you know, for 20 years before they published the book, they had a different version every year or two. And now that they *have* published the book, they still come out with new versions. It is a beautiful abstract structure, and it has important concrete implications, but it has also this fantastic Rube Goldberg aura to it.

In brief, that one cannot find a consistent selection is not terribly distressing or surprising. If anything, it is an additional criticism of selection theory, not of consistency. Though it is not devastating, it *is* something that you have to chalk up against selection theory.

Q: In your foreword to the book of Harsanyi and Selten, you write that one rationale of Nash equilibrium, that a normative theory that advises people how to play has to prescribe an equilibrium since otherwise it is self-defeating, essentially relies on the theory making a unique prediction. Hence, one needs a selection to justify the equilibrium concept in this way.

A: Yes, that is precisely what I meant when I just said that selection plays a fundamental role in justifying equilibrium. Thank you for clarifying that. What was your question?

Q: It was a remark that if you need selection for this ‘‘justification’’ of the equilibrium concept and if there is no consistent selection, then this ‘‘justification’’ might be problematic.

A: Yes, it is indeed a little problematic. As I just said, the whole selection project is a little problematic, and this *is* a mark against it. But if you want to ‘‘justify’’ equilibrium from a cognitive point of view, then equilibrium selection is not the only way of doing it; you can do it in other ways also. Have you seen the recent paper[49] by Brandenburger and myself? It is a different cognitive approach to Nash equilibrium, not using selection. You get equilibrium when certain informational assumptions are satisfied; the equilibrium does not have to be unique.

So, selection is just one way of ‘‘justifying’’ equilibrium. That it is inconsistent with consistency is unfortunate, or regrettable, but it is not the end of the world, certainly not the end of the world for equilibrium, not even the end of the world for selection. We often have inconsistencies, it is something one learns to live with, both as a physicist and as a game theorist.

[49] ‘‘Epistemic Conditions for Nash Equilibrium,’’ *Econometrica* **63** (1995), 1161–1180.

There is one more point to be made in this connection. Harsanyi and Selten base their selections on the mathematical form (strategic or perhaps extensive) of the game. Thus in two games with the same mathematical form, the same equilibrium must be selected. Now *that* is surely *not* needed to ‘‘justify’’ the idea of equilibrium from the point of view of advising the player. When you are giving advice, you do know about the real-life context of the game, and you *can* base your advice on that. You could easily select different equilibria in different contexts. For example, in my controversy with Roth and Shafer about the NTU Shapley value, I discussed two games with precisely the same mathematical form—one in a trading context, the other in a political context—whose ‘‘natural’’ or ‘‘expected’’ equilibria are quite different.[50] For a simpler example, think of a two-person game where each player must choose ‘‘L’’ or ‘‘R’’; each gets 1 if they choose the same, 0 otherwise. Suppose they are driving toward each other, and the ‘‘L’’ or ‘‘R’’ represents the side of the road on which they drive. If it’s in England, I would select (L,L); in the Netherlands, (R,R). Harsanyi and Selten suggest randomizing, which seems a little crazy.[51]

In brief, there’s no good reason to base the selection on the mathematical form only; the context, the history, also matter.

Q: Werner also had a question about consistency. Could we turn to the last part of that question, which reads: ‘‘Could you comment on the extremely opposite justifications of Nash equilibrium which partly require perfect rationality and partly deny any cognition?’’

A: I have already responded to that at some length in this interview. They are just different approaches. There is no problem with that; on the contrary, it sheds light on them. It helps to fortify, to corroborate the concept, to validate it. Güth thinks of them as one ‘‘versus’’ the other;[52] he thinks if one is ‘‘right,’’ the other must be ‘‘wrong.’’ I think they are both right, they both illuminate the concept, from different angles. And as I have said, I don’t believe in justifications; it’s not religion and it’s not law; we are not being accused of anything, so we don’t have to justify ourselves.

Q: Let’s move on to the next question on Güth’s list: ‘‘In my view, the notion of a strategy with all its partly counterfactual considerations seems

[50] ‘‘On the Non-Transferable Utility Value: A Comment on the Roth−Shafer Examples,’’ *Econometrica* **53** (1985), 667−677; see specifically Section 8, 674−675.

[51] They might answer that I’m making too much of a degenerate, nongeneric example. But that misses the point. Suppose we perturb the payoffs very slightly, so that the Harsanyi−Selten selection is pure. I would still select (L,L) in England and (R,R) in the Netherlands. They, of course, would select the same in both countries. That’s almost as crazy.

[52] W. Güth and H. Kliemt, ‘‘On the Justification of Strategic Equilibrium—Rationality Requirements versus Conceivable Adaptive Processes,’’ DP 46, Economics Series, Humboldt University, Berlin, 1995.

already a too demanding concept for a descriptive theory. Correlated equilibria are of an even more complex nature.''

A: Well, both assertions sound strange. It's true that the notion of a strategy is an abstraction, but it's conceptually a simple object. Anyway, if you're going to challenge that, then you're challenging the whole conceptual basis of the noncooperative theory, including, of course, the notion of a Nash equilibrium.

As for correlated equilibria, conceptually they are fairly simple objects. The set of correlated equilibria is always a convex, compact polyhedron with finitely many extreme points. It is a very simple object, very easy to work with, much easier than the set of Nash equilibria. Nash equilibria are algebraically very complex objects. Harsanyi once wrote an article describing the deep algebraic geometry of Nash equilibria.[53] By contrast, a schoolchild can work out the correlated equilibria. They are much simpler objects.

Q: By now I have understood that assumptions do not count.

A: It's not that assumptions don't count, but that they come *after* the conclusions; they are *justified by* the conclusions. The process goes this way: Suppose you have a set of assumptions, which logically imply certain conclusions. One way to go is to argue about the innate plausibility of the assumptions; then if you decide that the assumptions sound right, then logically you must conclude that the conclusions are right.

That's the way that I reject, that's bad science.

The other way is not to argue about the assumptions at all, but to look at the *conclusions only*. Do our observations jibe with the *conclusions*, do the conclusions sound right? If yes, then that's a good mark for the assumptions. And then we can go and derive other conclusions from the assumptions, and see whether *they're* right. And so on. The more conclusions we have that jibe with our observations, the more faith we can put in the assumptions.

That's the way that I embrace, that's good science. Logically, the conclusions follow from the assumptions. But empirically, scientifically, the assumptions follow from the conclusions!

Q: Let's move to the next question on Güth's list: Like in traditional evolutionary biology, in evolutionary game theory one often assumes a ''genetically'' determined behavior. In modern ethology this is rather debated (apes, for instance, are known to have well developed cognitive

[53] Unpublished. See also S. H. Schanuel, L. K. Simon, and W. R. Zame (1991), ''The Algebraic Geometry of Games and the Tracing Procedure,'' in *Game Equilibrium Models II* (R. Selten, Ed.), Berlin: Springer-Verlag.

systems). Should one not allow for a more continuous transition from no cognition at all (e.g., when studying primitive organs like plants) to more or less rational behavior (when studying the behavior of apes and humans) to resemble the rather gradual evolutionary process as, for instance, measured by DNA differences?

A: We discussed that above. No, one should not allow for that. We don't have to talk about apes, we can talk about humans. Humans undoubtedly have well developed cognitive systems and, in spite of that, much human behavior fits into the evolutionary paradigm. Most human actions are not calculated. I know that I don't calculate. Maybe you think that I don't have a well developed cognitive system: be that as it may, I hardly ever calculate when making decisions. Much of behavior is not calculated; it is either genetically determined or it is determined by experience, by learning—which is very similar to genetic determination, in the sense of Dawkins's memes.[54] A meme and a gene behave mathematically almost the same. Moreover, as I said above, to the extent that we *do* calculate, it may be a result of evolution.

Q: You have been very influential in developing a rigorous definition of common knowledge of rationality (CKR) and elaborating its implications. Recently it has been argued that CKR is self-contradictory where one, of course, relies on more centralized players like in the normal form and denies ''trembles'' in the sense of Selten's perfectness idea of equilibria. When do you think is the assumption of CKR justified?

A: Well, to answer this last question, it is almost never justified. It is a far-reaching assumption. Please refer to my paper in the Hahn Festschrift,[55] ''Irrationality in Game Theory,'' which indicates how very very small departures from CKR can lead to behavior that is very different from that of CKR, for example in the centipede game. The departures from CKR are small both in the sense of being tiny probabilities and also in that the failure is at a high level of CKR. In other words, you get mutual knowledge of rationality to a high level, and after that level CKR fails only by a very small probability; and nevertheless, the results are very different from those under CKR. So, CKR is not ''justified''; it does not happen. But that does not mean that CKR is unimportant. It is still very important to know what CKR says and what it implies, and to understand the connection between it and backward induction. Just like it is important to know how a

[54] *See the Selfish Gene*, op. cit. (Footnote 21).

[55] *Economic Analysis of Markets and Games, Essays in Honor of Frank Hahn* (P. Dasgupta, D. Gale, O. Hart, and E. Maskin, Eds.), Cambridge, MA: MIT Press, 1992, pp. 214–227.

perfect gas behaves, though there are no perfect gases; and it is important to know what perfect competition does, although perfect competition does not exist. It is important to know what happens in the ideal state, although ideal states don’t exist.

No, CKR never happens.

Q: Is it self-contradictory?

A: No, that is simply a mistake. The idea that CKR is self-contradictory was due to an inadequate model. If you build your model carefully and correctly, then CKR is not self-contradictory. Admittedly, I had to think for three years before coming up with the right model. It is a very very confusing business, it is very subtle, it takes a lot of thought. The reader is referred to my article on ‘‘Backward Induction and Common Knowledge of Rationality.’’[56] That article should settle this issue, or go a long way toward settling it.

No, CKR is not self-contradictory in games of perfect information.

Q: Well, I heard you lecture about this paper and I must admit I didn’t understand the argument fully. I haven’t read the paper yet, but I am going to do so.

A: Good. Read the paper; it is not difficult and there is a careful conceptual discussion at the end of the paper. The structure is not very elaborate. It has three building blocks: rationality, knowledge, and commonality of knowledge. You first examine each one separately, define it carefully; then you put the three together, and you get backward induction. So first you have to ask yourself: What is rationality? What does it mean to be rational? And that has a simple definition. Not, what does it mean to have common knowledge of rationality, just what does *rationality* mean? And then, what does it mean to have knowledge? What is the exact model of knowledge? And you ask that independently of what rationality is, and what common knowledge is. Just, what does *knowledge* mean? And you find a satisfactory answer to that. And then you have to ask yourself: What is *common* knowledge? And there you use the accepted definition, you just iterate the knowledge operator. The key is to think about knowledge separately, and about rationality separately, and only then, in the end, to go to common knowledge of rationality; and then you get the result. It is to keep the ideas separate, not to confuse them. That is what mathematics is about, that’s where mathematics helps.

Q: Shall we conclude the interview here or is there something that you would like to add?

A: I would like to express my tremendous admiration for Selten, as a human being, and as a pioneer in a number of fields. First of all, in the

[56] Op. cit. (Footnote 45).

equilibrium refinement business. The first refinements, and among those that are sure to survive, are the subgame perfect equilibrium and the ordinary perfect equilibrium, the trembling hand perfect. This will survive, this is a monumental idea. Another very important contribution is the idea of doing experiments. I don't put quite the faith in experiments that Selten does, but one very important function of experiments, which he realizes himself, is that to run an experiment you really have to see what the rules of the game are. You really have to think about the game very very carefully. Just like when you write a computer program for anything, you have to ask yourself, what am I doing here? Now, when you run an experiment it forces you to say, what is this game? What the subjects actually do is perhaps not all that significant (although it can be suggestive), but when you are designing an experiment it is very important to design it right. So, that is an input that Selten recognizes and emphasizes, and it bears more emphasis.

And, as I said last Sunday evening when we were talking about the Nobel Prize winners, a very important contribution of Selten is his work on the evolutionary approach. He has made some important scientific[57] contributions on this, but the main importance is to bring this into the world of game theory, to bring in the work of Maynard Smith[58] and his associates. Another thing to be mentioned is that he had an important influence also on John Harsanyi in value theory. His thesis[59] was basically about the Harsanyi value,[60] and that is a very important concept.

On a different level, another important idea is Selten's umbrella. I'm not sure that all the people at the dinner on Sunday understood what this is about. Selten lives in a world where rain is unpredictable; it can rain any day. Rather than continually investing mental energy in the question ''Should I take an umbrella or shouldn't I?'', he has made it a rule of life always to take an umbrella. Now, if he starts thinking that when he comes to Israel he really does not need an umbrella because it never rains in the summer, then he is contravening this rule of life which is *not* to put

[57] ''A Note on Evolutionarily Stable Strategies in Asymmetric Animal Conflicts,'' *Journal of Theoretical Biology* **84** (1980), 93–101; and ''Evolutionary Stability in Extensive Two-Person Games,'' *Mathematical Social Sciences* **5** (1983), 269–363 (see also ''Correction and Further Development,'' *Mathematical Social Sciences* **16** (1988), 223–266).

[58] J. Maynard Smith (1982), *Evolution and the Theory of Games*, Cambridge: Cambridge University Press.

[59] R. Selten (1964), ''Valuation of *n*-Person Games,'' in *Advances in Game Theory*, Annals of Mathematics Study 52 (R. D. Luce, L. S. Shapley, and A. W. Tucker, Eds.), Princeton: Princeton University Press, pp. 577–627.

[60] J. Harsanyi (1963), ''A Simplified Bargaining Model for the *n*-Person Cooperative Game,'' *International Economic Review* **4**, 194–220.

thought into it. So he has this meme for always carrying an umbrella. He himself is an embodiment of the evolutionary approach to game theory.

So, I have tremendous admiration for Reinhard and I am very happy that he got the Nobel Prize and I wish him continued success in the future and he seems to be strong and going well.

Q: Thank you very much for this interview!