AN INTERVIEW WITH ROGER MYERSON ON NASH AND GAME THEORY

Questions by Yanis Varoufakis

January 2002

Greek translation to appear in Θεωρια Παιγνιων: Αφιερωμα στον John Nash (Game Theory : An Afieroma to John Nash), edited by Petros Raptis.

Q1. Professor Myerson, in your 1999 JEL article you credited John Nash with the discovery of a concept (i.e. the Nash equilibrium) which empowered the economic approach to unify the social sciences.[1] Would you go as far as to argue that, post Nash, game theory must be a compulsory component in the curriculum not only of economics and political science graduate courses but also in sociology, anthropology etc.?

A1. Yes, I am willing to argue that doctoral students in any social science should be required to learn something about game theory, because it is a basic critical methodology that should be available to them in their scholarly discourse. Whenever a social scientist makes a prediction about how people will behave in some kind of situation, the question may be raised as to whether this predicted behavior can be justified as a Nash equilibrium under some reasonable assumptions about what people's preferences and information in these situations. If the theory cannot be so justified then there is serious reason to doubt the validity of this prediction, because it means that at least some people in this situation would not want to behave according to the prediction when they understood it.

I am not saying that every prediction of every social scientist needs to be formally tested according to the methodology of game theory. But a social scientist who does not understand how to make this test is missing an important tool for evaluating social theories, and would be unable to engage in a dialogue with others who might criticize his (or her) predictions game-theoretically.

This argument is essentially the same as the argument for requiring some basic study of statistics in doctoral programs. Even if a social scientist does not regularly use statistical tests to evaluate every theoretical prediction that he might make, every social scientist needs to at least understand what a statistical test is, to be able to participate in a dialogue with others who might criticize his prediction with statistical analysis. I am not saying that students of social science need the same number of weeks of studying game

theory as they need of statistics, but both subjects need to be in the required curriculum for the same reason.

Q2. In the same article you drew an analogy between Nash's presentation of his equilibrium concept and the discovery of the double helix in biology. But is it not the case that, if we agree with your assessment that Nash offers us a chance to unify the social sciences, a better analogy for the Nash equilibrium would be with the much sought after *Theory of Everything* in Physics (e.g. the theory that unifies electromagnetic, gravitational and nuclear forces as part of a single fundamental force of nature)?[2]

A2. I thought that it might be useful to compare Nash equilibrium with the double helix because both were discovered around the same time, and each discovery has profoundly changed the fundamental perspective of its field. If physicists had discovered the Theory of Everything around 1950, I would have made that comparison too. But I do not want to suggest that Nash equilibrium should be the unique theory of everything in social science. Obviously other perspectives are also needed, and there may be other conceptual frameworks that are as valuable and as broad in scope as game theory.

Q3. Nash's equilibrium is usually commended for its predictive/'positive' power concerning rational play (that is, choices under the assumption of infinite order common knowledge rationality). Is this your view also? Or would you recommend it for its normative features too?

A3. I think of Nash equilibrium as a tool for predicting how people will behave in any given game. Such a tool is most useful for social engineering, when you can redesign the game. If some proposed change in the structure of a game would generate equilibrium outcomes that are better (under whatever welfare criterion you find appropriate) then you have a strong reason to change the game in this way.

Q4. B. Bernheim and D. Pearce, in their 1984 *Econometrica* papers, have argued that the Nash Equilibrium may well fail to acknowledge other (non-Nash) strategies that fully rational players might select with good reason.[3] To illustrate, for our readers, the two

games which follow are identically structured (from a strategic viewpoint) and feature a unique Nash equilibrium according to which Player R chooses R2 and Player C chooses C2.

	C1	C2	C3
R1	100,99	0,0	99,100
R2	0,0	1,1	0,0
R3	99,100	0,0	100,99

Game I

	C1	C2	C3
R1	1,0	0,-1	0,1
R2	-1,0	10,10	-1,0
R3	0,1	0,-1	1,0

Game II

Would you concur that, perhaps, R1 is more readily rationalizable in Game 1 than in Game 2? Equivalently, is the Nash equilibrium equally compelling in both games? Is Nash not missing something important in failing to acknowledge (a) that rational players have good cause to treat the two games differently (even if laboring under common knowledge of rationality), and (b) that rationality, even of an infinite degree, does not suffice in order to rule out non-Nash play?

A4. In both of the games that you describe, there is an equilibrium in which each player randomizes with equal probability between his or her first and third strategies, and so player 1 could rationally choose his first strategy (R1) in either game. Of course the Pareto superiority of this randomized equilibrium in your Game 1 may make it a focal equilibrium (in Schelling's sense) for the players there, in which case R1 would be more likely to be chosen in Game 1 than Game 2.

Q5. In many different contexts (e.g. the two games above, Rubinstein's bargaining solution etc.) commonly known rationality does not suffice to weed out, as irrational, non-Nash strategies. In these cases the Nash equilibrium requires something stronger than rationality (S. Hargreaves-Heap and Y. Varoufakis, in their book, call this 'something' the *axiom of consistently aligned beliefs*). Do you agree? If rationality does not do the 'trick', what does? Which is the psychological or social mechanism that generates this remarkable degree of consistency required to shore up the Nash equilibrium in these cases? Might this be the Achilles' heal of your claim that the Nash equilibrium ought to be the key for understanding all social phenomena?

A5. In my book, I base the game-theoretic approach on two assumptions: rationality and intelligence. By "intelligence" I mean that the people whom we are studying are assumed to understand their situation at least as well as we do, and they should understand our theory of behavior in their game if our theory is correct. This "intelligence" assumption may be the same as the axiom of "consistently aligned beliefs" that you mention, but I formulate the intelligence assumption as a constraint on social theorists, not on social situations. There may be many situations where people do not have consistent beliefs about each other's behavior. But in any situation where a social theorist correctly understands how people will behave, they should have the same consistent beliefs if they are intelligent.

Q6. Risk aversion is allowed to influence rational play only through its impact on the players' utility functions. Which means that, once a game is expressed in terms of utility numbers, players can be modeled a risk-neutral maximizers of those utility numbers. Are you satisfied with this pivotal assumption underpinning Nash's equilibrium concept? To make the question more concrete, and to illustrate its essence for our readers, could you please comment on the following game? Assuming that the payoffs are in utils, is it not true that rational players can justify playing the non-Nash, non-rationalizable (a la Bernheim) strategies (R3,C3) even under common knowledge of rationality? And if so, is it not true that Nash's idea of what constitutes rational play is not compelling in this game?

	C1	C2	C3
R1	1,1	-100,-100	2,0
R2	-100,-100	1,1	0,0
R3	0,2	0,0	1,1

A6. Yes, I am content with the use of utility payoffs to summarize players' attitudes towards risk in a game. As with any mathematical theory of human behavior, utility theory does not always successfully predict the decisions that people make, but it is a compelling and general theory that can account for much observed behavior. The power of utility theory has been essential to the development of game theory since the publication of von Neumann and Morgenstern's book.

In the game that you describe, there is an equilibrium in which each player randomizes with equal probability between his or her second and third strategy, and so I have no difficulty with the prediction that (R3,C3) might be the outcome of this game.

Q7. Do you believe that, assuming common knowledge of rationality as well as perfect/symmetrical information, defection in each round is the uniquely rational strategy in the finitely repeated prisoner's dilemma? Equivalently, do you think that rational co-operation in the early rounds of this game can be explained *only* as part of a sequential equilibrium scenario according to which players invest in a reputation for irrationality (in the presence of some initial uncertainty regarding their rationality/utility)?

A7. When Robert Axelrod invited me to participate in his original study of the repeated prisoners' dilemma game, I assumed that he was running a finitely repeated version of the game, in which defecting at every stage is the unique equilibrium strategy, and I thought that this solution was so obvious that I did not bother entering his contest. In fact, he was not running a simple finitely repeated game, and the fact that other players did not share my view of the "obvious" solution made other strategies much better. His experimental findings challenged game theorists, and the result was that we learned to better appreciate how small perturbations in a game can greatly change its set of

equilibria. Most notably, Kreps, Milgrom, Roberts, and Wilson showed (in the *Journal of Economic Theory*, 1982) that early cooperation would be the unique sequential equilibrium outcome of a natural perturbation of this game, in which there are small-probability doubts about whether each of us might be a tit-for-tat machine. I find this a very useful and realistic explanation of the observed cooperative behavior in these games. But I would not say that any game-theoretic argument was the only way to explain such experimental behavior.

Q8. Are you satisfied with the current 'state-of-the-are' regarding multiple Nash equilibria? Has the so-called refinement project succeeded or would you agree that it has degenerated to the point where it is no longer of much interest?

A8. I have to admit that the literature on refinements of Nash equilibrium concept seemed to get into a maze of difficulties by the late 1980s, in a way that reminded me of the literature on cooperative solution concepts in the late 1960s. We found many examples of games where some Nash and sequential equilibria look irrational, and many refinements were proposed to eliminate the bad equilibria, but no single refinement emerged that could handle all these examples in a satisfactory way. It seemed that an endless proliferation of refinements was possible.

I think that the problem with our approach was that everyone was focusing on the equilibrium concept and not on the game concept. For example, we could define a nonstandard game in strategic form to be a strategic-form game together with a sequence of perturbed games (with the same players and strategies) that have payoffs which converge in the limit to the first game's payoffs. Then our solution concept for a nonstandard game should be the set of its Nash equilibria that are limits of the equilibria for the games in the sequence. This concept would allow a modeler to specify the refinement that seems appropriate in his application.

Another approach that also deserves more investigation is the "dual reduction" approach that I developed in my paper that appeared in *Games and Economic Behavior* in 1997. There I showed that there is a class of elementary games which have no imperfection problem when correlation is allowed, and that any non-elementary game can

be reduced to an elementary game by a process of dual reduction, which generalizes the idea of eliminating dominated strategies.

Q9. Turning now to Nash's other crucial contribution, are you convinced by his solution to the bargaining problem? Do you believe in its predictive/'positive' value? How about its normative worth? (E.g. should an arbitrator split a 'pie' between N persons in accordance with the Nash bargaining solution?)

A9. The Nash bargaining solution is a very beautiful idea, but its interpretation is difficult. I think of it, first and foremost, as a mathematical theory of fair-sharing for guiding the prescription of an impartial arbitrator when there is no obvious way of comparing different players' payoffs. Then, once we recognize that it is a natural recommendation for an impartial arbitrator, the Nash bargaining solution may also become the focal equilibrium of a bargaining game that has multiple equilibria. But as Schelling argued long ago, other focal factors often lead players in real situations to implement bargaining equilibria that may be very different from the Nash bargaining solution. For example, when each player is paid in dollars, but the players cannot make side payments to transfer these dollars among themselves, then criterion of equal monetary gain for both players often determines the focal equilibrium that they implement, even though it is far from the Nash bargaining solution.

Q10. Would you agree with the proposition that costly disagreement can be due either to imperfect rationality or imperfect information (or both)?

A10. As a game theorist, I would never try to explain costly disagreement by imperfect rationality of the players, although I am sure that costly foolishness often happens in real life. I think that imperfect information and informational incentive constraints are a very powerful explanation of costly disagreements in real life. But I would not say that incomplete information is the only explanation of costly disagreement. Bargaining games usually have multiple equilibria, and people may get stuck in a bad equilibrium even when information constraints allow the possibility of

efficient agreement. In such situations, however, an effective arbitrator could help the bargainers by leading them to focus on a better equilibrium of their bargaining game.

Q11. Robert Sugden has argued that Nash's axioms (on which his bargaining solution is founded) are mere conventions that could have been otherwise.[4] If Sugden is right, would it be fair to say that Nash's 'solution' is one out of many rationally 'agreeable' agreements? In that case, can we not interpret conflict/disagreement as the outcome of an historical process which shapes these social conventions?

A11. Schelling argued in his book *Strategy of Conflict* that historical processes can shape expectations that select focal equilibria of bargaining games. This focal-point effect is probably more useful for understanding real bargaining situations than the Nash bargaining solution.

Q12. Are you satisfied with the current theory regarding what happens in dynamic games when players intentionally step off the Nash equilibrium path? Are such deviations irrational *by definition* (as Harsanyi and Aumann have argued)? Or can you imagine circumstances under which experimentation with off-equilibrium strategies is, potentially, rational (even under commonly known rationality and symmetrical information)?

A12. The essence of the refinements problem in extensive form is the question of what you could reasonably believe about me when I surprise you with some unexpected action. Yes, I am not satisfied with our understanding of this question. Sequential equilibrium assigns minimal restrictions on what you could believe in such circumstances, and in some examples it allows beliefs that seem intuitively unreasonable, but we do not have a generally accepted theory of how to more reasonably restrict these beliefs. But the idea that a player might experiment off the equilibrium path cannot happen with positive probability, by definition of equilibrium! However, an infinitesimal probability of such experimental deviation could be admitted by an economic modeler who uses a nonstandard definition of game (game plus perturbed sequence of games), as I suggested above.

Q13. Granted that evolutionary game theory demonstrates the robustness of Nash's equilibrium (in the sense that all evolutionary equilibria are Nash equilibria, though the opposite is untrue), how important a contribution has it made to the study of socio-economic phenomena? Do you think evolutionary game theory will eventually overshadow conventional game theory within the social sciences?

A13. I look to evolutionary game theory mainly to provide another kind of refinement, for culturally familiar games. For example, consider a coordination game like the famous "battle of sexes" example. If this game is played by individuals in a laboratory with no social context, I would consider the symmetric randomized equilibrium to be a natural prediction. But if they play in a social context, knowing a long history of play by other pairs of players in the past, then I would expect the historical process to converge to one of the two nonsymmetric pure-strategy equilibria (which are better for both players than the randomized equilibrium). More generally, if we could find some such evolutionary model that would always (for any game) converge to some nonempty set of equilibria, then this set could become an important refinement of the Nash equilibrium concept for culturally familiar situations.

Q14. Evolutionary models presume irrational agents who either inherit or mimic successful behaviour. By contrast, conventional game theory posits hyper-rational agents. Could the two extremes be bridged after we acknowledge that, on many situations (e.g. in the bargaining game, Games 2 and 2 in Question 4 above etc.), rationality (of whatever degree) fails to dissolve indeterminacy and that, when this happens, even hyper-rational agents are simply drawn to the more successful strategies (since their Reason has failed to point the way to the optimal strategy)?

A14. As I said in my answer to the previous question, I am hoping that someone will use evolutionary game theory to reduce the indeterminacy of Nash equilibrium in many games. I would be very glad if you could do it! But to be game-theoretically useful, this theory must identify a nonempty set of equilibria for any game. That makes the problem difficult.

Q15. In his 1993 AER paper Matthew Rabin argued that game theorists must take into consideration the psychological aspects of gaming; that is, the possibility that one's utility from an outcome does not only depend on what happens in the end but also on one's beliefs regarding the rationale/intentions of one's opponents. If Rabin is right, then the strategic structure of the game (e.g. its normal form representation) is a function of people's beliefs about one another and, evidently, even simple games (e.g. the prisoner's dilemma) become indeterminate. Do you agree with Rabin? If so, do you also agree that the Nash equilibrium loses much of its explanatory power in this context?[5]

A15. I think that you and Rabin are suggesting that, in real life, my preference for some outcome must always depend on how I think it may shape the future, which depends in turn on how others think about this outcome. I agree that this is true in almost any real social situation that I can imagine. The assumption that players can evaluate their preferences for outcomes at some point in time, without any further analysis of these outcomes' impact on the world thereafter, is a simplification that seems essential for game-theoretic analysis. To some extent we can get around this simplifying assumption by going to repeated games, but even in repeated games there is an essential assumption that a player's total payoff can be computed from partial payoffs in each period that are independent of the path of play after the period. When players always care about long-run relationships with others, we should look to sociological or social-psychological theories to complement our game-theoretic analysis.

Q16. Evolutionary and psychological games are popular post-1970 developments. While the former helps weed out some of the multiple equilibria afflicting game theoretical analyses, the latter has the opposite effect. Would you say that the problem with multiple equilibria is, on the whole, intensifying or getting more manageable?

A16. Andrew McLennan has recently shown that randomly-generated games with many players and strategies usually have huge numbers of equilibria. Results like his suggest that the multiple-equilibrium problem is very significant.

Q17. Turning on to more philosophical matters, in your 1996 *Games and Economic Behavior* article you argued that social science ought to model social institutions on the assumption that agents themselves are not flawed (even if the institutions are). However, and this might be of particular interest to our Greek readers, you seem to define a flawed person as someone who fails to be an efficient servant to her/his desires/preferences. Are you convinced that *homo economicus i*s, by definition, not flawed? We ask this because we know that at least one of our predecessors (Socrates) would disagree.

A17. I have argued that game theory has helped return modern economic to the breadth of concern about social institutions that characterized the ancient Greek philosophers who gave economics its name. In particular, like the ancients, we now do not need to put questions about the allocation of goods in the market into a separate methodological category from questions about the allocation of powers in the government. But you are correct in observing that I am still drawing another distinction that ancient scholars like Socrates did not make: between questions about the design of social institutions and questions about the education of individuals. I believe that this is an important and useful distinction, and I believe that much ancient social philosophy was confounded by a failure to make this distinction. In particular, it may be quite dangerous to assume, as Plato and Lenin sometimes did, that an educational process to generate a new and better kind of man can be relied on to make a new political structure perform well. For purposes of evaluating and comparing social institutions, it is better to assume fixed characteristics of the individuals in them. Questions of education are also important. But I believe that, unlike Socrates, we should separate questions about how to educate individuals from questions about how to reform our social institutions, allocating the former questions to psychology and the latter questions to economics. Imperfection of human rationality must loom large in any psychological theory of education, but I would argue that they are less important in an economic theory of social institutions.

Q18. Aristotle famously proclaimed that: “The weaker are always anxious for justice and equality. The strong pay heed to neither” (Politics, s1318b) Do you agree?

A18. I have been recently working on game-theoretic models of aristocratic government, which I believe has been the most common form of government in human history. I would agree that such political systems are generally characterized by a lack of concern for equality between aristocrats and commoners. But principles of equality and justice among aristocrats of a similar rank are often extremely important in the functioning of aristocratic systems.

Q19. The above question regarding Aristotle's cynical outlook on justice and equality seems to resonate with a neo-Humean argument in modern political economy which, in turn, seems to have been influenced by evolutionary game theory. For instance, Bob Sugden argued in his 1986 book that moral beliefs emerge as devices which reinforce spontaneously evolved conventions. This account begins with everyday games which are simultaneously important and indeterminate (e.g. the hawk-dove game which captures contests over property rights). Certain discriminatory conventions evolve in order to minimize conflict and resolve the indeterminacy. However, these conventions assign the 'good' social roles asymmetrically (e.g. they favour tall people over short people) and people living under them find it hard to justify the emerging distribution. Sugden's crucial argument is that, at that stage, people's moral beliefs change and they start believing in the justice of the conventions (e.g. both tall and short people start believing that the tall people deserve more!). Do you see an analogy between this evolutionary account and Aristotle? Do you agree with either of them? What is, in your opinion, the rational basis for justice and equality?

A19. I would agree that the stability and Pareto-superiority of nonsymmetric equilibria in games like "hawk-dove" and "battle of sexes" helps us to understand how and why societies break symmetry among people, with the result that similar people are separated into rich and poor, weak and powerful.

Going the other way, Machiavelli in his *Discourses* has argued that justice and equity for broader classes of citizens were vital for mobilizing larger numbers for military action in the history of ancient Greece and Rome, and since the French revolution we have seen the advantages of military mobilization in modern republics. Such

observations may tell us something about the evolution of our modern concept of a universal right to equal protection under the law.

Q20. Finally Professor Myerson, if game theory had not been invented, and John Nash had stuck to his mathematics, how different would your academic career be today? Would your understanding of the world (e.g. globalised capitalism, world hunger, the feebleness and wonder of human nature, the lure of power) be any different?

A20. How would my career have been different without John Nash? What a wonderful question! My difficulty in answering it is an indication of how influential he has been in everything that I have done. I wanted to be a mathematical social scientist since I was 12 years old, when I read Isaac Asimov's science-fiction novel *Foundation*. But my concept of what kinds of mathematical models should be studied was completely transformed when I read Nash's and Harsanyi's papers in college.

Of course I would like to imagine that, if Nash had not done it, perhaps I might have formulated his equilibrium concept. But Nash equilibrium seems like a deceptively obvious concept, and one of the points in my 1999 paper in the *Journal of Economic Literature* was to explain why this concept was so hard for people to recognize before Nash. My instinct in research has always been to look for the right general mathematical structures for fruitfully modeling social situations. So perhaps in a lifetime of work on mathematical social science I might have eventually stumbled onto something of what he did in three years.

One's mature understanding of the world probably depends more on the historical record that one has studied than on the particular methodological framework that one uses to analyze it. But I do believe that, by using modern economic analysis, we can now achieve a deeper and broader understanding of the world than Thucydides and Machiavelli could achieve without our modern analytical tools. In this regard, my own incremental insights would be halved at least without Nash's foundations of noncooperative game theory.

Footnotes (to Q1, Q2, Q4, Q11, Q15):

[1] That is, to move beyond the analysis of prices/quantities and to encompass every social interaction in which agents have well defined objectives.

[2] By contrast, and notwithstanding the immense importance of DNA, the latter has not had a similar unifying effect among different natural sciences.

[3] Even in simple one-shot games featuring a unique NE in pure strategies.

[4] For example, while in some social setting the independence of irrelevant alternatives is relevant (e.g. ruling out agreements that would have not been reached is deemed by bargainers to be an utterly irrelevant 'move') in another social context some other convention may be present (e.g. the convention that such 'erasure' of potential agreements must tilt the agreement in favour of the bargainer who would have been most upset had the erased agreement been imposed upon them).

[5] I.e. what psychological mechanism or rational faculty would bring about a coincidence of first and second order beliefs?