

Nash Equilibrium and Mechanism Design*

Eric Maskin

Institute for Advanced Study

and

Princeton University

November 2008

* This short paper was prepared for a conference at Princeton University June 13-14, 2008 in celebration of John Nash's 80th birthday. I thank the NSF for research support.

A *Nash equilibrium* (called an “equilibrium point” by John Nash himself; see Nash 1950) of a game occurs when each player chooses a strategy from which unilateral deviations do not pay. The concept of Nash equilibrium is far and away Nash’s most important legacy to economics and the other behavioral sciences. This is because it remains the central solution concept—i.e., prediction of behavior—in applications of game theory to these fields. As I shall review below, Nash equilibrium has some important shortcomings, both theoretical and practical. I will argue, however, that these drawbacks are far less troublesome in problems of *mechanism design* than in many other applications of game theory.

1. *Solution Concepts*

Game-theoretic solution concepts divide into those that are *noncooperative*—where the basic unit of analysis is the individual player—and those that are *cooperative*, where the focus is on coalitions of players. John von Neumann and Oskar Morgenstern themselves viewed the cooperative part of game theory as more important, and their seminal treatise, von Neumann and Morgenstern (1944), devoted fully three quarters of its space to cooperative matters. This pattern was followed by leading game theory textbooks, such as Luce and Raiffa (1957) and Owen (1970) up until the mid-1980s. Today, by contrast, Nash equilibrium—the noncooperative concept *par excellence*—dominates the standard textbooks, and such leading texts as Fudenberg and Tirole (1991), Myerson (1991), and Osborne and Rubinstein (1994) give short shrift (or no shrift at all) to the cooperative side.

I believe that there are three (related) reasons for the historical shift from cooperative to noncooperative theory¹: (a) most cooperative theory ignores *externalities*, the possibility that a coalition can be affected by the actions of those not in the coalition; (b) it assumes that a Pareto efficient outcome will be reached; and (c) it supposes that the grand coalition (the coalition of all players) will form. Point (a) corresponds to the fact that typically games in cooperative theory are represented in characteristic function form, according to which the payoffs that any particular coalition can achieve are independent of which other coalitions form. Point (b) is embodied in the efficiency axioms typically imposed by cooperative solution concepts, requiring outcomes to be Pareto efficient, and (c) is implied by efficiency in superadditive games² (which are the games usually studied in the literature). These features of cooperative theory are problematic because most applications of game theory to economics involve settings in which externalities *are* important, Pareto *inefficiency* arises, and the grand coalition does *not* form. To see this, one need look no further than the classic game-theoretic model, Cournot duopoly.

Nash himself proposed a unification of noncooperative and cooperative theory that has come to be known as the *Nash Program* (Nash 1951). Yet, while there has been important work following this idea up in the theoretical literature (see Serrano 2005 for a recent survey), the Nash Program has not yet had much effect on applications.

Nash equilibrium has been successful as a solution concept first because it is *logically coherent*. Specifically, it is the only concept that is consistent both with (1) expected payoff-maximization by players (rational behavior) and (2) correct forecasts by players about what others will do (rational expectations). Moreover, it has proved to

¹ For a fuller exposition of these points, see Maskin (2003).

² A game is superadditive if the union of two disjoint coalitions can obtain at least the sum of the payoffs of the two separate coalitions.

make good predictions of behavior both in experimental and field settings, at least when subjects have acquired sufficient experience playing the game in question.

2. Drawbacks

Nevertheless, Nash equilibrium has several important shortcomings. First, many games have *multiple* equilibria, and players may not be clear about which one to focus on. If the players can communicate with each other before the game is played, they may be able to select an equilibrium through negotiation (that is why a Nash equilibrium is sometimes referred to as a “self-enforcing agreement”). But negotiation does not always suffice to resolve multiplicity. Consider, for example, the game of Table 1 (borrowed from Aumann 1990), in which (U, L) and (D, R) are both equilibria (there is also a mixed strategy equilibrium). Players may attempt to negotiate the outcome (U, L) , which Pareto dominates the other equilibrium. Thus, player 1 will announce that she plans to play U and player 2 that he plans to play L . Notice, however, that these professions may not be credible. In particular, regardless of what she does herself, player 1 is better off if player 2 takes action L . But player 2 will play L only if he thinks there is a sufficiently high

		player 2	
		<i>L</i>	<i>R</i>
player 1	<i>U</i>	9,9	0,8
	<i>D</i>	8,0	7,7

Table 1

probability that 1 will choose U . Thus, player 1 has the incentive to say that she intends to play U regardless of whether that is actually true. Moreover, because U is a risky strategy for 1 (she could get a payoff as low as 0 with U , whereas with D her lowest possible payoff is 7), she might well play D if she is not very confident that 2 will play L . In other words, her announcement that she will play U is not really believable, and neither is the announcement L by player 2. Negotiation between the two players may, therefore, not accomplish much.

Of course, even without communication, multiple equilibria do not always cause a problem. Consider, for example, the game of Table 2. There are two equilibria:

		player 2	
		<i>L</i>	<i>R</i>
player 1	<i>U</i>	9,9	0,0
	<i>D</i>	0,0	1,1

Table 2

(U, L) and (D, R) (plus a mixed strategy). However, (U, L) stands out as obviously superior; it is *focal* in the sense of Schelling (1960). Unfortunately, not all games have one particular equilibrium toward which players will naturally gravitate. For instance, in the game of Table 3—the classic Battle of the Sexes—the equilibria

		player 2	
		<i>L</i>	<i>R</i>
player 1	<i>U</i>	2,1	0,0
	<i>D</i>	0,0	1,2

Table 3

(U, L) and (D, R) are exactly symmetric, and so there is no obvious criterion that would direct players to one equilibrium rather than the other.

Even when Nash equilibrium is unique, rationality on the part of the players does not by itself guarantee that equilibrium will be reached, as a player's theory about what the others will do may be incorrect. Look at the game in Table 4.

		<i>b₁</i>	<i>b₂</i>	<i>b₃</i>
<i>a₁</i>	0,7	2,5	7,0	
<i>a₂</i>	5,2	3,3	5,2	
<i>a₃</i>	7,0	2,5	0,7	

Table 4

There is a unique Nash equilibrium (a_2, b_2) . However, any other outcome is *rationalizable* (see Bernheim 1984 and Pearce 1984) in the sense of being consistent with the players' common knowledge of their rationality. For example, it is optimal for player 1 to play a_1 if she anticipates that player 2 will play b_3 . And this anticipation is justified if player 1 thinks 2 has reason to believe that 1 will play a_3 , etc..

3. *Mechanism Design*

Although all these problems with Nash equilibrium are important, they are typically much less severe than usual when the game at hand is the outcome of *mechanism design*. The theory of mechanism design is the “engineering” part of economic theory. One starts with a particular goal or objective and then enquires if and how a mechanism—that is, a game—could be designed that attains that goal in equilibrium (in which case the game is said to implement the goal). In other words, the game is *chosen* not given.

There are three major analytical advantages that accrue from a game being chosen by a mechanism designer rather than simply being given by “nature.” First, since the designer can specify the rules in advance, the *players themselves* should presumably know exactly which game is being played. Consider, by comparison, the uncertainty that, Ford, Chrysler, and General Motors face in their “game” with one another: the timing, the possible moves, and the possible payoffs are all quite unclear.

Second, in mechanism design, the *analyst* observing and studying a game's execution also knows the rules of the game. And this feature makes experimental and empirical field work considerably easier than in the usual case, where—as in the automobile industry—we have only highly simplified and very approximate models to go

by. Indeed, one complaint about experimental economics is that lessons learned in the lab are at times difficult to apply to the field because the respective games in the two settings are not guaranteed to be the same. But in mechanism design, the games can be *constructed* to be the same, and so the standard objection no longer applies.

Finally, in mechanism design, the *games themselves* can be chosen to have attractive properties. For example, in some standard settings, implementing games can be constructed so that they have a unique Nash equilibrium (see Maskin and Sjöström 2003). In fact, one can sometimes ensure that there are no rationalizable outcomes other than this unique equilibrium. Specifically, Abreu and Matsushima (1992) show that this is possible quite generally if one doesn't require that goals be achieved exactly but only approximately.

It is even sometimes possible to construct implementing mechanisms with equilibria that are considerably stronger than ordinary Nash equilibrium. For instance, in auction settings with private values in which buyers have quasilinear utility for the good being sold, one can attain the goal of efficiency by means of a Vickrey (second-price auction), in which bidding one's actual valuation is a dominant strategy (i.e., regardless of what others do, bidding one's valuation is optimal). Moreover, in some interdependent values settings, a generalization of a Vickrey auction achieves much the same thing (but with *ex post* equilibrium rather than dominant-strategy equilibrium); see Crémer and McLean 1988, Dasgupta-Maskin 2000, and Bergemann and Morris 2008.

Thus, from several perspectives, Nash equilibrium is a much less problematic solution as used in mechanism design than in many other areas of economics and political

science. Indeed, mechanism design provides the circumstances perhaps most favorable for Nash equilibrium being a good predictor of human behavior in strategic settings.

References

- Abreu, D. and H. Matsushima (1992), "Virtual Implementation in Iteratively Undominated Information: Complete Information," *Econometrica*, 60, 993-1108.
- Aumann, R.J. (1990), "Nash Equilibria are not Self-Enforcing," in J-J Gabsewicz, J-F Richard, and L. Wolsey (eds), *Economic Decision-Making: Games, Econometrics, and Optimisation*, Amsterdam: North-Holland, 201-206.
- Bergemann, D. and S. Morris (2008), "Ex Post Implementation," *Games and Economic Behavior*, 63, 527-566.
- Bernheim, B.D. (1984), "Rationalizable Strategic Behavior," *Econometrica*, 52, 1007-1028.
- Cr mer, J. and R. McLean (1988), "Full Extraction of the Surplus in Bayesian and Dominant Strategy Auctions," *Econometrica*, 56, 1247-1257.
- Dasgupta, P. and E. Maskin (2000), "Efficient Auctions," *Quarterly Journal of Economics*, Vol. CXV, May 2000, 341-388.
- Fudenberg, D. and J. Tirole (1991), *Game Theory*, Cambridge: MIT Press.
- Luce, R.D. and H. Raiffa (1957), *Games and Decisions*, New York: J. Wiley and Sons.
- Maskin, E. (2003), "Bargaining, Coalitions, and Externalities," Presidential address, Econometric Society.
- Maskin, E. and T. Sj str m (2002), "Implementation Theory," in K. Arrow, A. Sen, and K. Suzumura (eds.), *Handbook of Social Choice Theory Vol. I*, Amsterdam: North Holland, pp. 237-288.
- Myerson, R.B. (1991), *Game Theory: Analysis of Conflict*, Harvard University Press.
- Nash, J.F. (1950), "Equilibrium Points in n-Person Games," *Proceedings of the National Academy of Sciences*, 36, 48-49.
- Nash, J.F. (1951), "Noncooperative Games," *Annals of Mathematics*, 54, 289-295.
- Osborne, M. and A. Rubinstein (1994), *A Course in Game Theory*, Cambridge: MIT Press.
- Owen, G. (1970), *Game Theory*, New York: Academic Press.
- Pearce, D. (1984), "Rationalizable Strategic Behavior and the Problem of Perfection," *Econometrica*, 52, 1029-1050.

Serrano, R. (2005), "Fifty Years of the Nash Program, 1953-2003," *Investigaciones Económicas*, 29, 219-258.

von Neumann, J. and O. Morgenstern (1944), *Theory of Games and Economic Behavior*, Princeton University Press.