From *Game Theory 5 Questions,* edited by V.F. Hendricks and P.G. Hansen, Automatic Press, 2007.

Game Theory: Some Personal Reflections

H. Peyton Young

I. Why were you initially drawn to game theory?

It would be more accurate to say that I stumbled into game theory because I needed it to solve other problems I was working on. My first brush with the subject occurred in the early 1970s when I was teaching a new course called *Mathematics in the Social Sciences* at the Graduate School of the City University of New York. The topics included voting rules, Arrow's theorem, the Gale-Shapley matching algorithm, the Banzhaf voting power index, the Shapley value, and legislative apportionment, which I was then developing with my colleague Michel Balinski. The course had an applied flavor and involved studying practical methods used for election, apportionment, and college admissions, as well as theories about what methods might be best in principle.

It occurred to me that the *strategic* aspects of voting rules needed to be considered in addition to the usual axiomatic treatments, which had been the dominant paradigm since Arrow's work (Arrow, 1963). This led me to surmise that virtually all voting schemes on three or more alternatives must be manipulable, in the sense that the truthful reporting of preferences may sometimes fail to be a Nash equilibrium. However, I knew virtually nothing about game theory, so I quickly boned up on the subject using Luce and Raiffa's 1957 classic (which is still well worth reading today). I was on my way to proving the result when it occurred to me that someone else might have thought of the idea already. This led me to phone up Kenneth Arrow, who put me on to the papers of Gibbard and Satterthwaite that were just then starting to circulate (Gibbard, 1973; Satterthwaite, 1975).

While this discovery was rather discouraging, and led me to set aside game theory for awhile, it also suggested that it might be profitable to shift attention from negative results in voting theory to positive ones. Impossibility theorems notwithstanding, societies will go on voting by one means or another; the operational question is how to put practical voting methods on a sound theoretical footing. This led me to characterize the classical methods of Borda and Condorcet from first principles, and to show that they can be interpreted as statistical procedures for estimating the choice or ranking that is most likely to be "correct" when voters' judgments are subject to error (Young, 1974, 1975, 1986, 1988).

My next encounter with game theory arose quite by chance in the context of a consulting project. In the early 1980s I was working with two Japanese civil engineers for the municipal water authority of Malmo, Sweden, which wanted to know how to fairly allocate the costs of expansion among its customers. In the course of working on this project, I realized that some of the fairness criteria that I had developed for legislative apportionment could be adapted to this context. A particular example is monotonicity: when the size of the pie to be allocated increases, nobody should get a smaller portion than before. This led to one of the first applied papers in cooperative game theory, showing how ideas like the Shapley value and nucleolus played out in an actual situation with numbers estimated from data (Young, Okada, and Hashimoto, 1982). It also led to my first foray into theory, in which I showed that a simple *marginal contributions* *principle* could be used to axiomatize both the Shapley value for finite cooperative games and the Aumann-Shapley value for nonatomic games without invoking the additivity axiom (Shapley, 1953; Aumann and Shapley, 1974; Young, 1985a, 1985b).

To sum up, I was initially drawn to game theory not for abstract or philosophical reasons, but because it provided useful tools for solving concrete problems. Moreover, it was *cooperative* game theory and its implications for practical matters of fair division that inspired my earliest theoretical work.

II. What examples from your work or the work of others illustrates the use of game theory for foundational studies and/or applications?

Equilibrium behavior often requires players to use probabilistic strategies in order to create uncertainty in the minds of their opponents. Choice under uncertainty is therefore a central aspect of game theoretic reasoning. But the issue of uncertainty in games goes much deeper than this; indeed, it arises whenever the game has multiple equilibria, for then it is uncertain what equilibrium (if any) will be played. Uncertainty also arises when the players do not know their opponents' utility functions, or even whether their opponents are rational. The issue of what players need to know about a game in order to 'solve' it touches on deep problems in epistemology (Lewis, 1969; Aumann, 1976; Aumann and Brandenburger, 1995). This is one of many instances in which game theory has turned out to have important implications for academic disciplines that at first glance seem far removed from the subject.

In spite of the large demands it places on knowledge and common knowledge, however, many game theorists persist in using Bayesian Nash equilibrium, and refinements thereof, as a central solution concept. I suspect that this is largely because Bayesian rationality offers such a rich playground for clever argumentation. My own feeling is that Bayesian rationality is more of a minefield than a playground. This is not only because it places extraordinary demands on players' reasoning abilities, but also because the conditions under which Bayesian learning actually leads to equilibrium behavior turn out to be very demanding as well.

A number of authors, including Binmore (1987), Jordan (1993), Kalai and Lehrer (1993), and Nachbar (1997, 2005), have examined this issue from different perspectives. Dean Foster and I illustrate the problems with Bayesian learning as follows (Foster and Young, 2001). Suppose that two players are engaged in an infinitely repeated game of matching pennies. Assume that the payoffs are approximately what they would be in a textbook version of matching pennies, but that each of the payoff values is perturbed (once only) by a small random shock. The distribution of the payoff shocks can be common knowledge, but the players do not know their opponents' *realized* payoff values. Suppose that everyone is Bayesian, perfectly rational, and forward-looking. As the game proceeds, can they learn to predict the repeated-game strategy of the opponent, and will play converge to equilibrium play? The answer is: frequently not. It can be shown that, when the shocks are sufficiently small, one or both of the players will *almost surely* fail to learn to predict their opponent's next-period behavior with even approximate accuracy, and period-by-period play will be far from Nash equilibrium play most of the time.

The root of the difficulty is the assumption of exact optimization. An optimizing player who thinks his opponent is playing a mixed strategy can change his behavior very abruptly even though his *beliefs* about the opponent's behavior change only slightly. In a game that has only mixed strategy equilibria, the upshot is that, as each player tries to learn the behavior of his opponent, his own behavior becomes so erratic that *it* is unlearnable. In other words, the interactive nature of the learning process leads to the conclusion that not *both* players can be rational and learn to predict the behavior of the other. This, and related results of Nachbar and Jordan, mean that the use of Bayesian reasoning in games needs to be approached with considerable caution.

III. What is the proper role of game theory in relation to other disciplines?

Over the past fifty years game theory invaded first economics, then the rest of social science, and is now colonizing new territory in biology, computer science, and philosophy. One could say without much exaggeration that this is the imperial age of game theory. As in other empires, however, the act of conquest may change the conqueror in unintended ways. Let me venture a couple of predictions along these lines. Up through the 1980s the principal conquests were in economics, specifically the theory of industrial organization and mechanism design. In these settings it seemed reasonable to assume that players are highly rational, interact over long periods of time, and can employ complex forward-looking strategies. This led theorists to focus on the properties of equilibrium in repeated games, which was a major advance over the preceding literature.

I believe that this era is now coming to a close. The fields where game theory is currently having the greatest current impact are computer science, artificial intelligence, and biology, and this will lead to a change of emphasis in the development of theory. In particular, both biology and computer science call for versions of game theory in which 'rationality' plays a less prominent role than in economics. Computer scientists, for example, are concerned with distributed systems of information-processors and how to design protocols that govern their interaction. Game theory is highly relevant to this problem even though the "agents" may not be rational in the sense customarily assumed by economists.

Biology is another subject where game theory is having an important impact, and here too the demands of the application are rebounding on the questions that theorists are asking. The focus of analysis is on the long-run dynamics of large populations of players, whose relative fitness depends on how they are programmed to play. This approach originated in the work of the biologist John Maynard Smith (1982), and has had a substantial impact on the directions that game theory is taking and the ways in which it is being applied in the social sciences.[1] There is no reason why rationality cannot be accommodated in these models; indeed, an interesting open problem is to identify situations where the most highly rational agents in a heterogeneous population actually do have a long-run selective advantage.

My larger point is that game theory's colonization of biology, computer science, and other subjects is having a profound impact on the way that rationality is

[1] See among others Axelrod, 1984; Foster and Young, 1990; Young, 1993a; Kandori, Mailath, and Rob, 1993; Weibull, 1995; Epstein and Axtell, 1996; Vega-Redondo, 1996; Samuelson, 1997; Fudenberg and Levine, 1998; Young, 1998; Epstein, 2006.

treated in game theory, on the modeling of out-of-equilibrium dynamics, and a host of other issues in game theory itself.

4 & 5. What do you consider the most neglected topics and/or contributions in late 20th century game theory? What are the most important open problems in game theory and what are the prospects for progress?

I will address these last two questions in tandem. As I mentioned earlier, cooperative game theory is an unjustly neglected topic of research . This was not always the case: von Neumann and Morgenstern put a great deal of emphasis on the cooperative form, and many of the pioneers in game theory made major contributions to the topic (Shapley, 1953; Aumann and Maschler, 1964; Schmeidler, 1969; Aumann and Shapley, 1974). In recent decades, however, the noncooperative approach has increasingly gained the upper hand. Indeed, this trend has gone so far that many textbooks on game theory scarcely give cooperative theory a mention. One reason for this development, as I have already suggested, is that the topics in economics where game theory made its earliest inroads -- mechanism design and industrial organization – seem particularly well-suited to the noncooperative approach.

Another reason why cooperative game theory has languished is that its practical applications have not been widely recognized. Earlier I mentioned the problem of sharing costs among the beneficiaries of a public facility. Similar problems arise in setting rates for public utilities (Zajac, 1978). More generally, cooperative game theory is relevant to any situation where scarce resources are to be allocated fairly among a group of claimants. How, for example, should slots at busy airports be allocated among airlines? Which transplant patient should be first in line for the next kidney? How should political representation in a national legislature be fairly divided among parties and geographical regions? Some economists insist that such problems would be solved if they were simply left to the workings of the market. Unfortunately, this overlooks the point that markets are moot unless property rights have been defined and vested in individuals, which is precisely what methods of fair allocation are about.

In my book, *Equity In Theory and Practice* (1994), I examined various fairness concepts from both a foundational and practical standpoint. Cooperative solution concepts like the core and the Shapley value, as well as semi-cooperative notions like the Nash bargaining solution and the Kalai-Smorodinsky solution, provide the entry point for thinking about the meaning of allocative fairness. A close examination of practice, however, suggests that one must go substantially beyond these approaches to formulate a theory that has descriptive validity.

Three central points emerge from the analysis. First, fairness must be judged in the context of the problem at hand. Criteria for allotting transplant organs may be quite different from criteria that pertain to the allocation of legislative seats, and neither may be relevant to the allocation of offices in the workplace or dormitory rooms at college. In other words, notions of justice tend to be compartmentalized and context-specific, a view that has its roots in Aristotelian philosophy, and has been advanced by political philosophers such as Walzer (1983) and Elster (1992).

A second key point is that, in practice, solutions to fairness problems tend to be *decentralized* in the following sense: an allocation is deemed to be fair for a group of claimants only when every subgroup deems that they fairly divide the resources allotted to them. This *subgroup consistency principle* is very ancient. It is implicit, for example, in certain Talmudic doctrines concerning the division of inheritances (Aumann and Maschler, 1985). It also features in many modern solution concepts, such as the core, the nucleolus, and the Nash bargaining solution (Sobolev, 1975; Lensberg, 1988), and in real-world allocation methods such as rules for apportioning seats in legislatures (Balinski and Young, 1982; Young, 1994).

The cooperative game approach to fair division proceeds from an axiomatic standpoint. There is, however, another way of thinking about fairness norms that builds on *noncooperative* game theory. Norms of fair division – indeed norms in general – are often the unpremeditated outcome of historical chance and precedent. What is fair in one society may not be deemed fair in another, because people's expectations are conditioned by precedent, and precedents accumulate through the vagaries of history.

Such processes can be modeled noncooperatively using the framework of evolutionary game theory. As I mentioned earlier, this approach was originally inspired by biological applications, and typically has three key features: i) there is a large population of interacting players; ii) the players have heterogeneous characteristics, including different payoffs, information, and behavioral repertoires, iii) they adapt their behavior based on local conditions and experience, and are purposeful but not always perfectly rational. The focus is on the *dynamics* of such a process, not merely on its equilibrium states. One of the main contributions of the theory is to show that some equilibria have a much higher probability of arising than do others (Foster and Young, 1990; Kandori, Mailath, and Rob, 1993; Young, 1993a). It therefore delivers a theory of equilibrium selection that is based on evolutionary principles rather than on *a priori* principles of 'reasonableness', as in the earlier theory developed by Harsanyi and Selten (1988).

To illustrate how the evolutionary approach can be applied to the study of fairness norms, consider the classical problem of how two individuals would divide a pie. The simplest noncooperative formulation is due to John Nash (1950): each player names a fraction of the pie, and they get their demands provided that both can be satisfied; otherwise they get nothing. Any pair of demands that sums to unity constitutes a noncooperative equilibrium of the one-shot game. If the players are allowed to bargain over time, much tighter predictions are possible. In the standard model, players alternate in making demands, which are either accepted or rejected (Stahl, 1972; Rubinstein, 1982). When the players are perfectly rational and discount future payoffs at the same rate, the outcome of the unique subgame perfect equilibrium is the Nash bargaining solution.

Neither the one-shot demand game nor the alternating offers game is evolutionary in spirit, because they are concerned with what *two particular bargainers* would do in equilibrium, not what a *population of bargainers* would do. To recast the problem in an evolutionary framework, consider a large population of agents who engage in pairwise bargains from time to time. Suppose that the outcomes of previous bargains affect how people bargain in the future, due to the salience of precedent. Once a particular way of dividing the pie becomes entrenched due to custom, people start to think that this is the only fair and proper way to divide the pie, and it therefore continues in force.

To allow for asymmetric interactions, suppose that there are two distinct populations of potential bargainers who are randomly matched each period (e.g., employers and employees). Each matched pair plays the Nash demand game described earlier. Assume for simplicity that all agents in a given population have the same utility function, but that the utility functions differ between populations. To capture the idea that current expectations are shaped by precedent, suppose that each current player looks at a random sample of earlier demands by the opposing side, and chooses a trembled best reply given the sample frequency distribution. (The 'tremble' captures the idea that the process is jostled by small unobserved utility shocks, so that players usually choose a best reply but not always.) It can be shown that, starting from arbitrary initial conditions, players' expectations eventually coalesce around a specific division of the pie, and this endogenously generated norm of division is, with high probability, the Nash bargaining solution. Furthermore, when players are heterogeneous with respect to their degree of risk aversion, a natural generalization of the Nash bargaining solution results (Young, 1993b).

This example shows that there is no need to make extreme assumptions about players' rationality in order for game theory to yield interesting results. Unlike the alternating offers model, where perfect rationality and common knowledge of perfect rationality are assumed, neither is needed in the evolutionary model. Players choose myopic best replies based on fragmentary information, they occasionally make mistakes, and they have no *a priori* knowledge of their opponents' payoffs, behaviors, or degree of rationality. Nevertheless the two models yield essentially the same outcome.

More generally, the evolutionary model of bargaining illustrates how game theory can be used to study the emergence of norms. Over time, interactions among people build up a stock of precedents that may cause their expectations to gravitate toward a particular equilibrium, which then becomes entrenched as a social norm: everyone adheres to it because everyone expects everyone else to adhere to it. When the underlying game is concerned with the division of scarce resources, the resulting equilibrium can be interpreted as a *fairness* norm (Hume, 1739; Binmore, 1994; Young , 1998).

I conclude by hazarding several predictions about the future development of game theory. The first is that rationality, and arguments over how rational the players "really" are, will fade in importance. As I have already argued, game theory can be applied to systems of interacting agents whether or not they are rational in the conventional sense. This insight was initially provided by applications of game theory to biology, and is being buttressed by current applications to computer science, artificial intelligence, and distributed learning.

My second prediction is that game theory will continue to evolve in response to real problems that arise in economics, politics, computing, philosophy, biology and other subjects, a development that von Neumann and Morgenstern would surely have welcomed. While its major successes to date have largely been in economics, game theory is not a sub-discipline of economics; it is more like statistics, a subject in its own right with applications across the academic spectrum.

My third prediction is more of an admonition: game theory will continue to thrive if it remains receptive to new ideas suggested by applications, but risks degenerating if it does not. John von Neumann cautioned about this tendency in mathematics more generally, and game theorists would do well to heed his warning (von Neumann, 1956):

> "I think that it is a relatively good approximation to truth – which is much too complicated to allow anything but approximations – that mathematical ideas originate in empirics… As a mathematical discipline travels far from its empirical source, or still more, if it is a second and third generation only indirectly inspired by ideas coming from "reality," it is beset with very grave dangers. It becomes more and more purely aestheticizing, more and more purely *l'art pour l'art*. … [W]henever this stage is reached, the only remedy seems to me to be the rejuvenating return to the source: the reinjection of more or less directly empirical ideas. I am convinced that this was a necessary condition to conserve the freshness and the vitality of the subject and that this will remain equally true in the future."

References

Arrow, Kenneth J. (1963). *Social Choice and Individual Values*, 2nd edition. New Haven, CT: Yale University Press.

Aumann, Robert J. (1976). "Agreeing to disagree," *The Annals of Statistics*, 4: 1236-1239.

Aumann, Robert J., and Adam Brandenburger (1995). "Epistemic conditions for Nash equilibrium," *Econometrica* 63: 1161-1180.

Aumann, Robert J, and Lloyd S. Shapley (1974). *Values of Non-Atomic Games*. Princeton NJ: Princeton University Press.

Aumann, Robert J., and Michael Maschler (1964). " The bargaining set for cooperative games," in Melvin Dresher, Lloyd S. Shapley, and Albert W. Tucker, *Advances in Game Theory, Annals of Mathematics Studies No. 52*. Princeton NJ: Princeton University Press.

Aumann, Robert J. and Michael Maschler (1985). "Game theoretic analysis of a bankruptcy problem from the Talmud," *Journal of Economic Theory*, 36: 195-213.

Axelrod, Robert (1984). *The Evolution of Cooperation*. New York: Basic Books.

Balinski, Michel L., and H. Peyton Young (1982). *Fair Representation: Meeting the Ideal of One Man One Vote.* New Haven: Yale University Press. 2nd edition published in 2001 by The Brookings Institution Press, Washington DC.

Binmore, Ken (1987). "Modeling rational players I," *Economics and Philosophy*, 3: 9-55 and "Modeling rational players II", *Economics and Philosophy*, 4: 179-214.

Binmore, Ken (1994). *Game Theory and the Social Contract I: Playing Fair.* Cambridge MA: MIT Press.

Elster, Jon (1992). *Local Justice: How Institutions Allocate Scarce Goods and Necessary Burdens.* New York: Russell Sage Foundation.

Epstein, Joshua (2006). *Generative Social Science.* Princeton NJ: Princeton University Press.

Epstein, Joshua, and Robert Axtell (1996). *Growing Artificial Societies: Social Science from the Bottom U*p. Cambridge MA: MIT Press.

Foster, Dean, and Peyton Young (1990). "Stochastic evolutionary game dynamics," *Theoretical Population Biology*, 38: 219-232.

Foster, Dean P., and H. Peyton Young (2001). "On the impossibility of predicting the behavior of rational players," *Proceedings of the National Academy of Sciences of the USA*, 98: 12848-12853.

Fudenberg, Drew, and David Levine (1998). *Theory of Learning in Games.* Cambridge MA: MIT Press.

Gibbard, Alan (1973). "Manipulation of voting schemes: a general result," *Econometrica* 41: 587-601.

Harsanyi, John, and Reinhard Selten (1988). *A General Theory of Equilibrium Selection in Games*. Cambridge MA; MIT Press.

Hume, David (1739). *A Treatise of Human Nature*. Clarendon Press, Oxford, 1978.

Jordan, James S. (1993). "Three problems in learning mixed-strategy equilibria," *Games and Economic Behavior*, 5: 368-386.

Kalai, Ehud, and Ehud Lehrer (1993). "Rational learning leads to Nash equilibrium," *Econometrica*, 61: 1019-1045.

Kandori, Michihiro, George Mailath, and Rafael Rob (1993). "Learning, mutation, and long-run equilibria in games," *Econometrica*, 61: 29-56.

Lensberg, Terje (1988). "Stability and the Nash solution," *Journal of Economic Theory*, 45: 330-341.

Lewis, David (1969). *Convention: A Philosophical Study*. Cambridge MA: Harvard University Press.

Luce, R. Duncan and Howard Raiffa (1957). *Games and Decisions*. New York: John Wiley.

Maynard Smith, John (1982), *Evolution and the Theory of Games*. Cambridge: Cambridge University Press.

Nachbar, John H. (1997). "Prediction, optimization, and learning in games," *Econometrica*, 65: 275-309.

Nachbar, John H. (2005). "Beliefs in repeated games," *Econometrica*, 73, 459-480.

Nash, John (1950), "The bargaining problem," *Econometrica*, 18: 155-162.

Neumann, John von (1956). "The mathematician," in James R. Newman, *The World of Mathematics*, vol. 4. New York: Simon and Schuster.

Rubinstein, Ariel (1982)." Perfect equilibrium in a bargaining model," *Econometrica*, 50: 97-110.

Samuelson, Larry (1997). *Evolutionary Games and Equilibrium Selection*. Cambridge MA: MIT Press.

Satterthwaite, Mark A. (1975). "Strategy-proofness and Arrow's conditions: existence and correspondence theorems for voting procedures and social welfare functions," *Journal of Economic Theory*, 10: 198-217.

Shapley, Lloyd S. (1953). "A value for n-person games," in Harold W. Kuhn and Albert W. Tucker, eds., *Contributions to the Theory of Games, II, Annals of Mathematics Studies No. 28*. Princeton NJ: Princeton University Press.

Schmeidler, David (1969), "The nucleolus of a characteristic function game," *SIAM Journal on Applied Mathematics*, 17, 1163-70.

Sobolev, A. I. (1975). "Characterization of the principle of optimality for cooperative games through functional equations," in N. N. Voroby'ev, ed., *Mathematical Methods in the Social Sciences*. Vilnius: Vipusk 6, 92-151.

Stahl, Ingolf (1972). *Bargaining Theory*. Stockholm: Stockholm School of Economics.

Vega-Redondo, Fernando (1996). *Evolution, Games, and Economic Behaviour*. Oxford UK: Oxford University Press.

Walzer, Michael (1983). *Spheres of Justice*. New York: Basic Books.

Weibull, Jorgen W. (1995). *Evolutionary Game Theory*. Cambridge MA: MIT Press.

Young, H. Peyton (1974). "An axiomatization of Borda's rule," *Journal of Economic Theory*, 9, 43-52.

Young, H. Peyton (1975), "Social choice scoring functions," *SIAM Journal on Applied Mathematics*, 28: 824-838.

Young, H. Peyton (1985a). "Monotonic solutions of cooperative games," *International Journal of Game Theory*, 14: 65-72.

Young, H. Peyton (1985b). "Producer incentives in cost allocation," *Econometrica,* 53: 757-765.

Young, H. Peyton (1986). "Optimal ranking and choice from pairwise comparisons," in Bernard Grofman and Guilllermo Owen, eds., *Information Pooling and Group Decision Making*. Greenwich CT: JAI Press.

Young, H. Peyton (1988). "Condorcet's theory of voting," *American Political Science Review*, 82: 1231-1244.

Young, H. Peyton (1993a). "The evolution of conventions," *Econometric*a, 61: 57-84.

Young, H. Peyton (1993b). "An evolutionary model of bargaining," *Journal of Economic Theory*, 59: 145-168.

Young, H. Peyton (1994). *Equity: In Theory and Practice*. Princeton NJ: Princeton University Press.

Young, H. Peyton (1998). *Individual Strategy and Social Structure*. Princeton NJ: Princeton University Press.

Young, H. Peyton, Norio Okada and T. Hashimoto (1982). "Cost allocation in water resources development," *Water Resources Research*, 18: 361-373.

Zajac, Edward E. (1978). *Fairness or Efficiency: An Introduction to Public Utility Economics*. Cambridge MA: Ballinger.