The importance of being nice, retaliatory, forgiving and clear

Messrs Gorbachev, Reagan, Kasparov and Karpov have one thing in common. They are playing games. The strategies they ought to adopt are the stuff of game theory, a 30-year-old branch of mathematics that enthusiasts thought would have transformed economics and a bagful of other social sciences by now. Why is the world still waiting?

Game theory is not just about card games. Nor is it merely a way of generating clever puzzles or paradoxes. To understand the theory's significance, it is none the less worth exploring one such paradox: the dollar-auction game devised by Dr Martin Shubik of Yale University. The game goes as follows. Auction a dollar bill, but make it a quiet auction in which the highest bidder buys the dollar at the amount he bid, but the second-highest bidder has to hand over the amount he bid as well—with no dollar in return. Dr Shubik guessed that inexperienced players might end up bidding more than a dollar for the dollar.

Experiments by psychologists and countless parlour games have proved him right. Once two players have carried the bidding, in dollars, to one dollar against 90 cents, the second-highest bidder has to choose between giving $1.90 for the dollar, or paying 90 cents for nothing. Typically, he chooses to raise the bid. The first psychologist who formally tested the theory found that the students in his experiments often bid more than $3 for the dollar. But they also learnt from experience; as the game was replayed, the auction gradually began to peter out at lower bids.

At first sight this seems such an absurd situation that the game can have no real-life counterpart—least of all in a proper auction where only the highest bidder pays up. But Dr Shubik noted that the dollar-auction game had much in common with an arms race: his bidders, once ensnared, quickly reach a point where higher bids add to their potential losses without increasing their chance of winning. Similarly, in an arms race, both sides spend ever-increasing amounts of money without improving their strength relative to each other. Dr Shubik's game is a simple model of futile escalation.

The "prisoner's dilemma" is probably the best-known of these puzzles. Two hypothetical prisoners, suspects in a crime, are separately offered the same deal. If both confess, both will go to jail for five years. If only one confesses, implementing the other, he is set free and his partner gets 20 years. If neither confesses, both go to jail for one year on a lesser charge.

Figure 1 shows these options in what game theorists call a payoff matrix. The rows show the choices facing prisoner A, and the columns the choices facing prisoner B. Each cell in the matrix shows the payoff for prisoner A followed by the payoff for prisoner B. So, for example, the top right-hand cell says that if A confesses and B does not confess, A goes free and B goes to jail for 20 years.

First, consider the problem from A's point of view. If B confesses, A should confess too, or else he will go to jail for 20 years. Suppose, on the other hand, that B does not confess; A will go free if he confesses, or go to jail for one year if he too decides not to confess. So whatever B decides to do, A is better off if he confesses. B looks at the problem in the same way, and reaches the same conclusion. Whatever A decides to do, he is better off if he confesses. The outcome is that both confess, and both therefore go to jail for five years. The paradox is that if the prisoners had been less careful about choosing the right strategy, they might both have refused to confess, and got off with one year each.

Like the dollar-auction game, the prisoner's dilemma seems to be just an entertaining and abstract paradox. It might be more practical than that. First, it is about
making a decision in the face of uncertainty about somebody else’s behaviour. Second, an opportunity to co-operate is missed for want of the appropriate machinery (in this case, an enforceable contract between the prisoners). Both notions clash with conventional economic theory. This is built on a model of competition in which individuals—people and firms—have no effect on the outcome in free markets; in particular, they take prices as given by market conditions. All the marvellous talents of the invisible hand first described by Adam Smith—the process by which scarce resources are matched with individual preferences through the price system—depend on this lack of power. Once a firm can move the price in its favour, or a worker his wage, the invisible hand is frozen. The prisoner’s dilemma shows, in a nutshell, how “rational” self-interest can then make everybody worse off.

Economists knew this long before John von Neumann and Oskar Morgenstern published their classic, “The Theory of Games and Economic Behavior”, in 1953. They already had their theories of monopoly, oligopoly, and wage bargaining. These, though, were treated as special cases. The basic competitive theory was reckoned to be appropriate most of the time; and that theory still underlies the prejudices of most professional economists. But recent economic research indicates that the competitive model is really the special case, in a more general theory that gives bargaining power its proper place at the centre of the analysis.

The shock of the new
Many orthodox economists still resist this revolution in economic thinking. As in all sciences, defenders of the conventional methods accept change. And the mathematics of game theory is tricky—at least by the puny standards of the social sciences. This obstacle to game theory is crumbling: today’s graduate students regard mathematical forms of argument as a test of their intellectual viritity.

For many economists, though, the real trouble with game theory is that its methods make it hard to reach firm conclusions. Game theorists regard this indeterminacy as an advantage: it reflects the complexity of decision-making in the real world. Unfortunately, whatever claim economics has to be taken seriously rests heavily on its role as a guide for policymakers; if its advice can be simple and snappy, so much the better. Nothing in game theory can match the certainties of the pure theory of microeconomics—or the pretence of certainty in some branches of macroeconomics.

Even the simplest games can carry an ambiguous message. Take the simplest of all: two-person zero-sum games with “equilibrium” points. (A game is zero-sum if a winner can emerge only at the expense of other players.) Figure 2, from “Game Theory” by Dr Morton Davis, professor of mathematics at New York University, shows the payoff matrix facing two politicians. On an issue in a political campaign, each has three options: favour course X, favour course Y, or dodge the issue. Each cell shows the vote that A will get for each combination of his strategy with B’s possible strategies. (B’s vote is not shown. It is simply 100% minus A’s vote.)

With these vote forecasts, neither player has a strategy that is superior regardless of what the other does. For instance, if B dodges the issue, A should too. But if B dodges the issue, B should favour Y. The way to break such circles of reasoning is to see that B should never dodge the issue; whatever A does, B always does better if he favours X over dodging. This collapses the game to the two left-hand columns. Now whatever B chooses from his remaining options, A does best to favour Y. And if A favours Y, B should favour X. This leads the game to the circled cell; it is called an equilibrium because neither player can improve his position by changing strategy unilaterally.

There is a quick way for A to find the equilibrium payoff. If it exists, it is the largest in its column and the smallest in its row. Why? This is a zero-sum game, so A knows that B will always force him to the minimum payoff on whichever row A picks. A’s best bet is therefore to pick the row with the highest minimum value—game theorists call this the maximin. B faces a similar choice. He chooses the column that gives him his maximin too; this is equivalent to choosing the column that contains the lowest maximum payoff for A—the minimax, as A would see it. When the minimax and maximin are one and the same, as in figure 2, there must be an equilibrium.

Even such a simple game cannot give a watertight conclusion. A might think, for example, that B is such a stupid opponent that if A dodges the issue B will do the same. That would give A his best possible result, in the lower right-hand cell; it would be a gamble on B’s incompetence. The players will settle on the equilibrium point only if they both do as much as possible to promote their own interests.

Not all two-person zero-sum games have equilibrium points, though some of the ones that do not can be solved in a slightly more complicated way. Players can best pursue their interests in such cases by adopting what game theorists call mixed strategies—choosing one strategy for some of the time, then another, and another, in rotation. At this point—still the relatively simple realm of two players in a zero-sum setting—the solutions begin to get more complicated. Game theory only starts to get into its stride when it turns to many-person games. Then the scope for complication multiplies because of a new possibility—that groups of players might form coalitions.

Consider another of Dr Shubik’s games. Three contestants, A, B and C, each have a pistol and a balloon: From fixed positions, they fire at each other’s balloon; when one is hit, its owner is out. When only one balloon remains, its owner is the winner. At the outset, and after each shot, the players decide by lot who is to fire next; each can choose any remaining balloon as his target. Suppose also that the contestants rank A, B and C in descending order of skill. Can game theory work out a rule for rational play?

Underlying the game is a payoff matrix as before, except that it is now three dimensional. There seems no need to draw it; the right strategy is obvious. Suppose it is A’s turn to shoot. His chances of a hit are the same whether he aims at B’s balloon or C’s. But it would be better to hit B’s, because that would leave him vulnerable only to C in the next round, and C is not as good a shot as B. Each contestant arrives at the same conclusion, and therefore all agree on the rule of thumb first at the balloon of your stronger opponent.

So far, so trivial. But then who is most likely to win? The contestants’ chances can be precisely worked out if we know how accurately they can shoot. Suppose A hits a balloon with 80% of his shots, B hits with 60%, and C with 40%. It turns out that C, the worst shot, is most likely to win! (The respective probabilities of winning are 0.30 for A, 0.33 for B, and 0.37 for C.) The reason is that if A and B decide to follow the “rational” strategy, they will concentrate on shooting at each other, leaving C to face less than his fair share of attacks.

A better bet would be for A and B to collaborate to shoot down C’s balloon, and then fight it out between themselves. (In that case, their respective chances of winning would improve to 0.44 and 0.47, C’s falls to 0.09.) But there are many other possible strategies. For example, A might promise not to shoot at C unless C
shoots at him first; that would lead C to postpone A's retaliation by concentrating his erratic fire on B. (Under this strategy, the probabilities switch to 0.44 for A, 0.20 for B, and 0.36 for C.)

These and many other possible solutions cannot be judged right or wrong without information on the prospects for collusion—and in particular on the scope for trust or enforceable contracts between the players. It is the same kind of indeterminacy that crops up in non-zero-sum games like the prisoner's dilemma. Conventional economics seems less equivocal than game theory only because it ignores these aspects of the environment in which people make decisions.

This is changing, though slowly. Economists are using game theory more than before. They are applying the methods to decision-making by oligopolies, bargaining in labour markets, government policy on international trade, and—a particularly active area for recent research—trading in financial markets. Even where the formal methods of mathematical game theory have yet to catch on, jargon is blazing a trail. Zero-sum games seem to be sprouting everywhere, not least in the titles of best-selling books. Game theorists hope that, as before, where vocabulary leads, minds will eventually follow.

Economics has not been the only discipline to resist the advantages that game theory might bring. One application, in management science, was all but ignored until recently. Suppose a corporation has three divisions sharing the same factory: each division, for the sake of argument, occupies exactly a third of the premises, and generally seems to account for one-third of the factory's fixed overhead costs. These add up to $300. How much should head office charge each of its divisions—

which it wants to treat as separate profit centres—for these expenses?

The obvious answer is $100 each. But consider figure 3 on the next page. Stage one shows that charging each division $100 in overheads means that the one with the smallest net profit makes a loss; since it must stand or fall as a separate profit centre it is shut down. In stage two, a second division, which now has to pay half instead of one-third of the overheads, also starts to make a loss. It too is closed. In the last stage, the division saddled with all the overheads now makes a loss as well, and has to close. Yet the original three-division group had made total after-overheads profits of $220. (When Dr Shubik wrote a paper—in 1962—explaining how simple sharing of overheads could mislead managers in this way, he struggled in vain to get it published.)

Another way in which game theory can

---

**Why butterflies are bourgeois**

Game theory assumes that people act out of rational self-interest. Replace rational self-interest with Darwin's notion of "fitness" and game theory can also be applied to evolution. Evolution is the survival of the fittest. The fittest are those with the best strategies of behaviour. The best strategies depend on what other animals are doing. Evolution is therefore a game between strategies.

The main proponent of this school of evolutionary thought is Dr John Maynard Smith, professor of biological sciences at Sussex University. He invented the idea of an "evolutionarily stable strategy" (ESS)—a strategy that outscores all competition and will not therefore allow another behaviour to invade, until conditions changed.

Suppose, for example, that two animals of the same species are fighting over food. They can adopt one of two strategies, "hawk" and "dove". Hawks do not fight, but share the food. If challenged, they retreat. Hawks fight to the bitter end and, if they lose, limp away badly injured. They are like the people bidding for Dr Shubik's dollar, who end up paying too much. Obviously, when all the other players are doves, hawks do well because they beat the doves and stand little chance of injury. But when hawks are common, doves do better, because, although they rarely get anything to eat, at least they do not get injured. Depending on the numbers you put in, the ESS will turn out to be some mixed strategy of hawk and dove.

This may seem woefully childish. But it can lead to insight into animal fights. Add in some other strategies, like "uncertain" which fights only if its opponent plays hawk or "bully" which fights only if its opponent plays dove or "assessor" which plays hawk if it is bigger than its opponent and dove if it is smaller. Now things are beginning to sound realistic. Red deer stags, for instance, play assessor, going through ritualised roaring contests before, very rarely, fighting.

The value of these games is that they reveal the conditions under which particular strategies work. For example, the maths shows that assessor is an ESS against hawk and dove if the cost of assessment (in deer, the energy wasted in roaring) is less than half the value of what is being fought over (a harem of females).

There is a butterfly called the speckled-wood which defends spots of sunlit woodland floor against rivals. In such fights, the challengers always give way. Dr Nick Davies of Oxford University found that he could reverse the outcome of fights by removing a butterfly, allowing its rival to take over the sunspot and then releasing the original owner in the role of challenger. The new challenger (though he had once been the owner) always gave way. But if Dr Davies unerringly introduced a butterfly to the sunspot so that both thought they owned it, a long battle ensued.

This experiment can be described by a game between hawk, dove and "bourgeois", where the latter adopts the tactics of the butterflies: only the owner fights. Bourgeois is the ESS. After all, there are plenty of sunspots, so there is no sense wasting time and energy on long fights. But if you add in a fourth strategy, "proletarian", which is exactly the opposite of bourgeois (challenger always wins), it turns out to be just as capable of winning as bourgeois. Yet animals do not seem to use this rule.

These are very specific examples. Game theory can make some more general predictions. For instance, biologists have long been intrigued by altruistic acts. Altruism between relatives is easy to explain: animals who help their offspring or other close relatives will leave more descendants behind. But there are animals—including man—that help non-relatives.

One possible explanation is "reciprocal altruism". In theory, evolution could reward those who help others, so long as they get help in return. Remember Dr Rapoport's triumphant tit-for-tat. By Dr Maynard Smith's definition, tit-for-tat is an ESS. And animals do help each other in this way. Vervet monkeys buy assistance in fights by grooming each other. The grommed monkey then vies to the aid of the groomsman when the latter gets involved in a fight. The trouble with reciprocal altruism is that it depends on a praiseworthy contract between the actors. It is the prisoner's dilemma again.

---

**One got its sums wrong**

There is a butterfly called the speckled-wood which defends spots of sunlit woodland floor against rivals. In such fights, the challengers always give way. Dr Nick Davies of Oxford University found that he could reverse the outcome of fights by removing a butterfly, allowing its rival to take over the sunspot, and then releasing the original owner in the role of challenger. The new challenger (though he had once been the owner) always gave way. But if Dr Davies unerringly introduced a butterfly to the sunspot so that both thought they owned it, a long battle ensued.

This experiment can be described by a game between hawk, dove and "bourgeois", where the latter adopts the tactics of the butterflies: only the owner fights. Bourgeois is the ESS. After all, there are plenty of sunspots, so there is no sense wasting time and energy on long fights. But if you add in a fourth strategy, "proletarian", which is exactly the opposite of bourgeois (challenger always wins), it turns out to be just as capable of winning as bourgeois. Yet animals do not seem to use this rule.

These are very specific examples. Game theory can make some more general predictions. For instance, biologists have long been intrigued by altruistic acts. Altruism between relatives is easy to explain: animals who help their offspring or other close relatives will leave more descendants behind. But there are animals—including man—that help non-relatives.

One possible explanation is "reciprocal altruism". In theory, evolution could reward those who help others, so long as they get help in return. Remember Dr Rapoport's triumphant tit-for-tat. By Dr Maynard Smith's definition, tit-for-tat is an ESS. And animals do help each other in this way. Vervet monkeys buy assistance in fights by grooming each other. The grommed monkey then vies to the aid of the groomsman when the latter gets involved in a fight. The trouble with reciprocal altruism is that it depends on a praiseworthy contract between the actors. It is the prisoner's dilemma again.
have a growing impact is through computer studies. By repeating a game under a series of different strategies, it is often possible to discover that a particular strategy—without being a solution in the formal sense—proves more successful than the rest. If the game represents a familiar kind of problem from politics or economics, the result might be important. Dr. Robert Axelrod, a professor of political science at the University of Michigan, has organised computer tournaments based on the prisoner's dilemma; he describes what happened, and draws some adventurous conclusions, in a new book, "The Evolution of Co-operation". Dr. Axelrod invited some economists, political scientists, mathematicians and computer programmers to devise a computer program that would play in repeated episodes of the dilemma. He set each program in turn, against all the other entries; in the first tournament there were 16 different strategies and every pair played games of 200 moves each. The payoff matrix, with points instead of years in jail, is shown below in figure 4. With a reward of three points for co-operation, constant collusion over 200 moves would therefore give the program a score of 1,000. The reward for one-sided defection was five points and, if both defected, one point.

The success of a particular program depends on the form of the other programs. For example, the rule "always cooperate" (i.e., never confess) would do very well if it met another "always cooperate" (both would score 1,000) but very badly if it met an "always defect" (the cooperator would score 0 and the defector would score 1,000). An extra factor in this version of the prisoner's dilemma is repetition; that suggests the possibility of designing a program which spends early moves trying to figure out its opponent's behaviour, and then switching to a strategy designed to exploit it. What kind of rule might produce the highest score on average, in games with many other different rules?

The answer is "tit-for-tat", a simple program devised by Dr. Anatol Rapoport, a psychology professor at Toronto University. It said "co-operate on the first move, and then do whatever the other player did on the previous move". This rule scored 505 points, beating far more complicated rules that tried to probe the opponent. In a second tournament, which attracted 63 entries, it won again. This is even more striking, because Dr. Axelrod circulated detailed notes on the outcome of the first competition, so entrants realised that they had to beat tit-for-tat. They could not.

Tit-for-tat has four properties to thank for its success. First, it is nice. Nice rules—those that are never the first to defect—were much more successful on average than those that start by defecting. Second, it is retaliatory. It punishes defectors immediately, which helped it to resist programs that tried to exploit weakness. Third, it is forgiving. It only remembers the previous move, so it is always ready to re-establish fruitful co-operation with any like-minded rule. Finally, it is clear. Programs that try to fathom their opponents can quickly see that tit-for-tat cannot be pushed around; so it makes sense to co-operate with it.

Then Dr. Axelrod subjected the most successful programs to another test. Suppose that the competitions were repeated indefinitely through computer simulation, so that after each competition (called a "generation") the least successful programs were phased out and the more successful ones increased in number; would tit-for-tat still thrive, or does it depend on the presence of less successful programs to do well?

Figure 5 at the bottom of this column shows that it continued to outscore the opposition. Rules outside the top 15 quickly fell away. Note that after around 200 generations "Harrington", the rule that finished eighth in the original competition, suddenly faltered—and it was the only nasty rule to finish in the top 15 (it was one of the programs that analysed its opponents' behaviour to figure out how to beat them). It could do well only by exploiting other nasty rules; as they petered out, it did likewise.

The odd thing about tit-for-tat is that, for all its high scoring, it can never win a game. The best it can ever do is tie; the rest of the time it loses. Its virtue is that it can combine with other friendly rules so that it tends to tie or lose in high-scoring games. Unfriendly rules win more games, but score fewer points. "Always defect", for example, must always either tie or win, but typically does so in low-scoring games. Are players in the real world interested in high scores achieved through co-operation, or in winning? Winning probably counts for more than Dr. Axelrod—a great believer in people using tit-for-tat at the negotiating table—wants to admit...

The great virtue of game theory is that it does not only accommodate a wide range of assumptions about what motivates people (and butterflies—see the box on the previous page) and about the institutional circumstances in which people have to make choices. It also forces researchers to make their assumptions explicit. In economics, by contrast, the all-powerful assumption that people are selfish maximisers of their own welfare is rarely stated, let alone examined. The challenge that the game theory approach still has to meet is to find recruits who combine a delight in these minutiae—laws, customs, prejudice and the other paraphernalia of economic life—with a knack for hard maths.