nomic concepts. They are essential elements of twenty-first-century capitalism, and they play an important role in reality-based economics. But Bator, writing in the mid-1950s, couldn't address all of the problems that plague modern economies. Some hadn't become manifest; others hadn't been subjected to productive analysis.

One common problem arises when two or more parties to a given transaction have different incentives. The senior executives of a corporation, for example, may be more interested in ramping up the value of their stock options before they retire than in preserving the long-term interests of the stockholders. The problem, in this case, is designing an incentive package that aligns the two sets of interests.

Another class of problems emerges when the parties on either side of a transaction have different amounts of information. An important issue in health reform, for example, is that individuals know more about their health than insurance companies do, so the insurers are understandably wary about taking on individual new clients with pre-existing conditions. I use the word "understandably" because, today, almost all health insurers are profit-making concerns with a shareholder base that demands a certain level of earnings. From the perspective of an individual insurer, it is perfectly rational to turn away sick people. From the perspective of society as a whole, it is inhumane and inefficient. Sick people who don't get treated tend to get worse, at which point, often, they end up in the emergency room, where somebody has to pay for their care. This is an example of what I call rational irrationality. In the coming chapters, we will come across many more examples of it.

11. THE PRISONER'S DILEMMA AND RATIONAL IRRATIONALITY

As the problems of pollution and free riding make clear, many types of market failure come down to the fact of human interdependence. What I do affects your welfare; what you do affects mine. The same applies in business. When General Motors cuts its prices or offers interest-free loans, Ford and Chrysler come under pressure to match GM's deals, even if their finances are already stretched. If Merrill Lynch sets up a hedge fund to invest in collateralized debt obligations or some other newfangled securities, Morgan Stanley will feel obliged to launch a similar fund so its wealthy clients don't defect. Now, the chairmen of the Big Three automakers, despite all the criticism they have received recently, are presumably fairly rational, intelligent fellows who would rather coexist peaceably than get into damaging competition. (For now, we will set aside the mental capacities of Wall Street CEOs.) Isn't there some way they can get together and work out solutions that are beneficial to each of their companies and to their customers, too? Often the answer is no.

Consider the following example. Two firms, each of which owns a coal-burning power plant, are competing to supply electricity to a small town. As things stand, they split the market and each of them makes $20 million in annual operating profits. However, the local residents' association is pursuing a long-running lawsuit to try to stop their smoke emissions, which often blow into residential areas. Defending the suit and paying some high-priced publicity consultants costs the firms $10 million a year each, reducing their net profits to $10 million. At an individual cost of $5 million a year, the power plants could each install a filtering system that would capture most of their damaging emissions. If the two firms introduced the filters, the residents' association would drop its lawsuit, and the firms would each make net profits of $15 million.

At first glance, it seems obvious that the two plants should quit fighting the lawsuit, install the filters, and raise their net profits by $5 million. Things aren't so simple, however. The estimated cost of operating the filters assumes that the two firms introduce them simultaneously and raise the price of the electricity they produce, which brings them some extra revenue. Explicit price fixing of this sort is illegal in most countries: in the United States, it constitutes a violation of the Sherman Antitrust Act. This means that each of the firms has to reach its decision independently, taking into account what the other might do. If one of them installs the filters and the other one doesn't, the laggard will be able to supply power at a lower price, and its market share will increase, perhaps dramatically. For sake of argument, let's assume its annual net profits rise to $20 million, and that the profits of the other firm, the one that installed the filters, fall to $5 million. Now what should the firms do?

A convenient way of depicting the situation is to present both of the firms' options and payoffs in the same diagram. (See Table 11.1.) In each entry, the profits in millions of dollars of Firm A appear first and the profits of Firm B appear second. For example, in the upper-right-hand corner, Firm A makes $5 million and Firm B makes $20 million.

The way to think about this situation is that Firm A chooses a row and Firm B chooses a column. For example, if Firm A picks the first row and Firm B picks the first column, it means that both firms decide to install the filters: the outcome is shown in the top-left-hand corner, with both firms making $15 million in profits. If Firm A installs the filters and Firm B doesn't, the outcome is the top-right-hand corner, with profits of $5 million for Firm A and $20 million for Firm B.

TABLE 11.1: THE INSTALLATION GAME

		FIRM B	
		INSTALL	DON'T INSTALL
FIRM A	INSTALL	15, 15	5, 20
	DON'T INSTALL	20, 5	10, 10

There are at least two ways to figure out what is the best strategy for the firms to follow. The first one, which I mistakenly followed when I was presented with a problem of this nature in my college entrance interview, is to fixate on the top-left corner, which is clearly the most desirable outcome, and to try to rationalize why the two firms would select it. Initially, this doesn't seem very hard. The total payout if both firms installed the filters would be $30 million, which is higher than any of the other outcomes. Surely, the two firms would realize this. Being rational, they would gladly agree to install the filter systems, earning $15 million each. Everybody would win, the local town included.

To see what is wrong with this reasoning, place yourself in the shoes of the CEO of Firm A and imagine that you have just convinced yourself that installing the filters is the right thing to do. Since your counterpart at Firm B is almost as smart as you are, it is only safe to assume that he has reached the same conclusion, which means Firm B will also be installing the filters. In that case, however, your firm ought to renege on its conversion plans, thereby increasing its putative profits from $15 million to $20 million—bravo! Before congratulating yourself on your fiendish cunning, take a breath. The process doesn't end there. If you can figure out this logic, so can the head of Firm B, which means he, too, might well end up deciding not to approve the new filter system. And if that is a real possibility, you certainly shouldn't go ahead with your plans for installation, lest you end up seeing your firm's profits reduced to just $5 million!

This type of logic—"If I do this, he'll do that; if he does that, I'll do this"—might seem a little perverse to those uninitiated in game theory, but if you mull it over for a while, you will realize it is inexorable. The only sustainable outcome is the one in which Firm A and Firm B both continue to pump out toxic emissions and make profits of $10 million, just half of what they could have earned by installing the filters. The firms would like to settle on that cooperative outcome, in which they would both earn $15 million, but the temptation of trying to steal extra market share, and the threat of the other firm doing the same thing, are too great.

The situation that Firms A and B find themselves in is known as the prisoner's dilemma. In many cases, it is the basis of what I call rational irrationality, by which I mean a situation in which the application of rational self-interest in the marketplace leads to an inferior and socially irrational outcome. When the prisoner's dilemma was first introduced, back in the early 1950s, many people refused to accept that the two firms wouldn't be able to reach the cooperative solution. Economists and mathematicians were excitedly exploring the new science of game theory that John von Neumann and Oskar Morgenstern had invented in their 1944 treatise *Theory of Games and Economic Behavior*. Many smart people held out great hope for game theory, imagining it could solve many of the outstanding problems in the social sciences. The key to this process was thought to lie in extending the solution methods that von Neumann and Morgenstern had introduced, most of which applied to zero-sum games, such as coin-tossing and poker. In games of that nature, the players compete against one another, and one player's winnings are another player's losses. But many types of economic activity, such as international trade and investing in the stock market, involve the possibility of cooperation and mutual gains: they are positive-sum games. During the late 1940s, some progress was made in tackling this broader category of problems when John Nash, a Princeton mathematician, introduced a general method for solving non-zero-sum games, but much remained unclear.

Merrill Flood and Melvin Dresher were two mathematicians working at the RAND Corporation, which the Pentagon had founded in the aftermath of World War II to engage in scientific research "for the public welfare and security of the United States of America." Much of the work undertaken at RAND had military implications, but it was also an important center of operations research and other applications of mathematics. Flood got interested in how people actually played the sorts of non-zero-sum economic games that von Neumann and Nash

had theorized about: Did they cooperate and reach a mutually beneficial solution? Or did they lapse into cutthroat competition? One of Flood's first experiments involved his three teenage children. He offered them a babysitting job and arranged an informal reverse auction, asking each of the teenagers to submit the lowest wage they would accept to do the job, with the bidding starting at $4.00. Despite explicit encouragement from Dresher to get together and avoid a bidding war, the young Dreshers ended up submitting competing bids, and the winning bid, which landed the job, was just $0.90. If the three teenagers had agreed to submit a single bid of four dollars, they could have divided the work and money among them, with *each* receiving $1.33.

That was how children behaved; what about supersmart adults? In January 1950, Flood and Dresher recruited a couple of friends, another RAND scholar named John D. Williams and a UCLA economist named Armen Alchian, to play a game similar to the one involving Firm A and Firm B, and they had them play it a hundred times in a row. At this stage, Flood and Dresher didn't have a name for their game; in their research they simply called it "a Non-cooperative Pair." The outcome of the exercise was inconclusive. Alchian chose the cooperative strategy (the equivalent to "Install") sixty-eight times out of a hundred, and he selected the self-seeking strategy (the equivalent to "Not Install") thirty-two times. Williams picked the cooperative strategy seventy-eight times and the self-seeking strategy twenty-two times. The results could be read as providing evidence that cooperation was feasible, after all; but they could also be interpreted as demonstrating that even experts in game theory sometimes couldn't resist the self-seeking strategy that would lead to an inferior outcome.

The prisoner's dilemma received its name a few months later. Albert Tucker, a Princeton mathematician who had an affiliation with RAND, was asked to give a talk on game theory to a group of psychologists at Stanford. To make the concepts underlying Flood and Dresher's experiment easier to understand, he invented a story about two men who, having been arrested and charged with jointly carrying out a crime, are held and questioned separately, with no means of communicating. Both men have reason to believe that if they both deny the charge they will be cleared, or charged with a lesser crime. If they both confess they will face a heavy fine. But if one confesses to the crime and the other doesn't, the one who confesses will be given a reward and the one who sticks to his denial will be punished severely. The question is: Should the men confess or deny?

The setup is logically identical to the power plant game, and it can be represented with the same type of diagram. (See Table 11.2.) The only difference I will make to Tucker's original story is to substitute specific prison sentences for fines as the punishments. Each number in the table refers to the number of years in jail that the players will serve. If both men confess to the crime, they each get five years. (This outcome is shown in the lower-right-hand corner.) If they both stick to their denials, they get a year each on a lesser charge. If one confesses and the other denies, the one who confesses gets off free and the other gets fifteen years.

TABLE 11.2: THE PRISONER'S DILEMMA

		PRISONER 2	
		DENY	CONFESS
PRISONER 1	DENY	1, 1	15, 0
	CONFESS	0, 15	5, 5

As in the power plant game, the most desirable outcome appears to be the top-left-hand entry (Deny, Deny), which involves total time in jail of just two years. But what are the chances of the two prisoners achieving that solution? This time, let's tackle the problem in a slightly different manner—the way somebody schooled in game theory would attack it. Assume you are Prisoner 1. The key to analyzing any game is to figure out what your opponent is likely to do and then select your best option in response. Prisoner 2 has two options. If he sticks to his denial, the game is reduced to the elements of the first column. Your choices are to deny and face a year in jail, or confess and get discharged: clearly, confessing is the better option. Now, let's examine what happens if Prisoner 2 selects his other option and confesses. The game is reduced to the elements of the second column. If you stick to your denial, you will get fifteen years in prison; if you confess, you will get five years. Once again, confessing is the better option. The peculiar twisted logic of the game should now be clear: whatever Prisoner 2 does, Prisoner 1 is better off confessing, and the same reasoning applies to Prisoner 2's choice. In the language of game theory, confessing

is a "dominant strategy," even though it leads to a bad outcome for both players.

Many people find the prisoner's dilemma infuriating; their instinctive reaction is that there must be some way out of it. According to William Poundstone, the author of an illuminating account of early game theory, from which I have taken some historical details, the founders of the problem felt the same way. "Both Flood and Dresher say they initially hoped that someone at RAND would 'resolve' the prisoner's dilemma," Poundstone writes. "They expected Nash, Von Neumann, or someone to mull over the problem and come up with a new and better theory of non-zero-sum games. The theory would address the conflict between individual and collective rationality typified by the prisoner's dilemma. Possibly it would show, somehow, that cooperation is rational after all. The solution never came."

One criticism that is often made of the prisoner's dilemma is that the noncooperative solution (Confess, Confess) can't possibly be the rational outcome, because rational decision-makers would never choose an inferior outcome. This would be true if the two prisoners were deciding jointly what to do. However, collusion is ruled out by assumption: the essence of the game is that the two players have to choose independently. Another argument sometimes made is that since either Confess or Deny must be the best strategy to follow, both players, being rational, will end up choosing the same strategy. That narrows the possible list of solutions to "Confess, Confess" and "Deny, Deny." Since "Deny, Deny" is clearly the better choice for both players, it must be the logical solution. The problem with this reasoning is that it ignores the element of conflict that is key to the game. If Prisoner 1 believes that Prisoner 2 will deny, it doesn't matter how he reaches that conclusion: it always makes sense for him to confess.

Another objection to the prisoner's dilemma is that it is a one-shot game, whereas many types of economic interactions occur repeatedly: you go to work for the same company every day; Morgan Stanley and Merrill Lynch compete against each other every quarter. If economic decision-makers take into account the long-run gains from cooperation, isn't it sensible to cooperate in the short run, too? This turns out to be a deep question. On a theoretical level, it can be shown that as long as the game is repeated many times and the players don't know when it will end, the cooperative outcome ("Deny, Deny"/"Install, Install") is a rational solution, but so is the noncooperative solution ("Confess, Confess"/"Don't Install, Don't Install")!

Unlike in the case of the one-shot game, where "Confess, Confess" is the single dominant strategy, in the repeated game there are many rational strategies to follow. Another one, for instance, is to flip a coin before each round, confessing if it comes up heads and denying if it comes up tails. In 1980, Robert Axelrod, of the University of Michigan, organized a prisoner's dilemma tournament in which game theorists from around the country submitted fifteen different strategies for playing the repeated game. Axelrod then matched the strategies against one another in a computer tournament. Each match consisted of two hundred rounds, with points awarded according to a payoff box that looked very much like the one in Table 11.2. The winner was a simple version of tit for tat. The exact strategy was: "Cooperate on the first round and then in each consecutive round copy what your opponent did in the previous round." As long as the opponent cooperates, such a strategy can sustain cooperation indefinitely.

Axelrod's findings have received a lot of attention, and they may well help to explain how cooperation is sustained in many areas of human society and even in the animal kingdom. Playing tit for tat signals to other players that you are willing to cooperate, but you aren't a sap. If your opponent plays nasty, he gets punished, which gives him an incentive to revert to cooperating. Applying Axelrod's results to economics is dangerous, though. In many business situations, players don't have the luxury of taking a longer view, which makes the environment more akin to the one-shot game. If you are CEO of a company and you follow a cooperative strategy, only to have a rival CEO double-cross you, the mistake could well cost you your job. On Wall Street, traders are judged on a quarterly basis; if they have one bad quarter, the consequences can be disastrous. Even in journalism, the logic of cutthroat competition can be difficult to overcome. At the offices of one British tabloid, there used to be a sign hanging in the newsroom that read, "Do it to them before they do it to you!" Sadly, that is what most of us learn to do. Experiments with ordinary people playing one-shot prisoner's dilemmas show that many of them start out by denying their guilt and hoping that the other player does the same thing. However,

they rapidly discover that cooperating with somebody whose actions they can't control doesn't pay off. By the tenth trial, about 90 percent of the players choose the self-preserving strategy of confessing.

The prisoner's dilemma can crop up in any situation that involves elements of conflict and cooperation. In many industries, firms would like to restrict supply and raise prices, but such a strategy of tacit collusion is difficult to sustain, because each of the players has an incentive to cheat and raise its output. This is especially true in industries that produce a homogenous good, such as oil, where cheating is difficult to detect.

A typical production game is shown in Table 11.3. Here, again, the numbers represent millions of dollars in profits. Each player has a choice of two strategies. "Low" means low output and "High" means high output. The mutually preferred outcome is "Low, Low," with both firms restricting their output and making profits of $100 million. But if either firm plays "Low," the other one has an incentive to play "High." If this happens, the firm that plays "Low" makes only $25 million in profit. Therefore, the dominant strategy for both firms is to play "High," which results in them each making profits of $50 million.

TABLE 11.3: A PRODUCTION GAME

		FIRM 2	
		LOW	HIGH
FIRM 1	LOW	100, 100	25, 200
	HIGH	200, 25	50, 50

Prisoner's dilemmas aren't necessarily a bad thing. In some industries, such as oil and airlines, they help prevent cartels from operating effectively and keep prices down. OPEC, the association of oil-producing countries, is notoriously unstable. Every few months its members get together and make noise about cutting their output quotas to raise prices, but it rarely happens. The governments of many oil-rich nations are so dependent on oil revenues that they can't resist opening the spigots, and the result is widespread cheating on the production quotas. For Western oil consumers, this is a good thing. Much of the time, it ensures a ready supply of cheap gasoline and heating oil. (Whether it is in the long-term interest of Western consumers to run down energy supplies rapidly is another question.)

With some products, however, an increase in output isn't necessarily socially desirable, and the prisoner's dilemma can lead firms to over-supply the market. The residential mortgage may well be such a product. At any given time, a limited number of households have the income and employment security necessary to service a mortgage to maturity. But when house prices are rising and interest rates are low, banks and other financial companies lower their credit standards in search of quick profits, lending to borrowers who are one piece of bad news away from defaulting. Initially, some of the more responsible financial companies hold back—in terms of Table 11.3, they play "Low"—but as they see their opponents gaining market share, the pressure to switch strategies becomes irresistible. Eventually, almost all the firms in the industry adopt a high-output strategy, with deleterious consequences for one and all.

To capture the essential logic of the prisoner's dilemma, I have described the two-person game. In reality, banks and other lenders have to take into account the actions of many rivals: they are facing an "n-person" prisoner's dilemma. With just two people playing the game, it is difficult to sustain a cooperative outcome. When there are ten, or twenty, or a hundred players, it is virtually impossible—a fact highly germane to the overexploitation of natural resources, such as tropical rain forests, the fish in the sea, and the sub-Saharan plains. In 1968, Garrett Hardin, a Texan ecologist who died in 2003, tackled this problem in a famous article, "The Tragedy of the Commons." The example Hardin used was that of a pasture shared by local herders. The pasture is limited in size, and all the herders know that overgrazing will render it useless for everybody. At the same time, though, the herders' incomes are determined by the size of their herds, which gives them an incentive to add more animals to the pasture.

As Hardin pointed out, each herdsman is interested primarily in his own welfare. In deciding whether to add another animal to his herd, he carefully considers the likely consequences of such a step. The upside is that once the animal is grown, he can sell it and keep all of the proceeds for himself. The downside is that adding to his herd increases the dangers of overgrazing, which damages the pasture. However, since

all the farmers together will share the costs of overgrazing, the prospect of selling another animal will loom larger in the mind of the individual herdsman. "Adding together the component partial utilities, the rational herdsman concludes that the only sensible course for him to pursue is to add another animal to his herd," Hardin wrote. "And another; and another . . . But this is the conclusion reached by each and every rational herdsman sharing a commons. Therein is the tragedy. Each man is locked into a system that compels him to increase his herd without limit—in a world that is limited. Ruin is the destination toward which all men rush, each pursuing his own best interest in a society that believes in the freedom of the commons."

The story of the herdsman is just a parable, but it illustrates some vital issues relating to the sustainability of economic development. In each country, rivers, lakes, aquifers, mountains, and wilderness regions are part of the commons. The earth's atmosphere is part of the global commons, as are the oceans and oil and gas reserves. Biodiversity in the plant and animal kingdoms is another element of the global commons. Every time a species becomes extinct, the stock of natural resources is irreversibly depleted. Inevitably, the question becomes one of figuring out how to preserve the commons for future generations, or, where that is not possible, how rapidly to deplete it.

Free market economists often argue that privatizing common resources would ensure that they were used more responsibly. In some cases, this may be true—many historians believe the enclosure of common lands in fifteenth- and sixteenth-century England helped to raise agricultural productivity and stimulate economic growth. But privatization doesn't remove the conflict between private benefits and social benefits that defines the commons problem. In the United States, logging companies have decimated parts of the California redwood region, while in New York, Florida, and many other states, rampant real estate development threatens the supply of natural drinking water. On a global level, pollution, congestion, overfishing, and deforestation remain chronic problems, which are only likely to get worse as the global population continues to expand, globalization intensifies, and the demand for natural resources increases. Specifying property rights may well be a necessary part of tackling these enormously complex issues, but blind reliance on self-interest and the market is a recipe for further environmental catastrophes.

The first step in preventing such outcomes is recognizing the pervasive nature of rational irrationality and how difficult it is to overcome. "Game theorists get a lot of stick for denying that the individual behavior that leads to such disasters is irrational," Ken Binmore, one of the world's leading game theorists, notes in a recent book.

> Our critics ask how can it possibly be rational for a society to engineer its own ruin. Can't we see that everybody would be better off if everybody were to grab less of the common resource? The error in such reasoning is elementary. A player in the human game of life isn't some abstract entity called "everybody." We are all separate individuals, each with our own aims and purposes. Even when our capacity for love moves us to make sacrifices for others, we each do so in our own way and for our own reasons. If we pretend otherwise, we have no hope of ever getting to grips with the Tragedy of the Commons.

12. HIDDEN INFORMATION AND THE MARKET FOR LEMONS

In the late summer of 1966, significant things were happening in California's Bay Area. In Haight-Ashbury, a run-down neighborhood of cheap apartments and vacant buildings just east of Golden Gate State Park, a vibrant subculture was developing around marijuana, LSD, and the psychedelic music of Jefferson Airplane and the Grateful Dead, two local bands; in Candlestick Park, out near the airport, which was then home to the San Francisco Giants football team, the Beatles played what turned out to be their final concert before paying fans; across the water in Oakland, Bobby Seale and Huey P. Newton were founding the Black Panther Party.

In nearby Berkeley, George Akerlof, a twenty-six-year-old graduate of MIT's prestigious Ph.D. program in economics, was starting his teaching career. From early childhood, Akerlof had appeared destined for academia. His father, who emigrated from Sweden to the United States in the 1920s, was a chemistry professor. His mother, who met his father while she was in graduate school, came from a bookish family of German Jews. Akerlof was a bright and studious kid. In high school, he later recalled, "I belonged to a small group of students, who in today's terminology would be called nerds . . . Socially, I was a misfit. I failed