Prisoners' Dilemmas and How to Resolve Them

MANY CONTEXTS, ONE CONCEPT

What do the following situations have in common?

- Two gas stations at the same corner, or two supermarkets in the same neighborhood, sometimes get into fierce price wars with each other.
- In general election campaigns, both the Democratic and the Republican parties in the United States often adopt centrist policies to attract the swing voters in the middle of the political spectrum, ignoring their core supporters who hold more extreme views to the left and the right, respectively.
- "The diversity and productivity of New England fisheries was once unequalled. A continuing trend over the past century has been the overexploitation and eventual collapse of species after species. Atlantic halibut, ocean perch, Haddock and Yellowtail Flounder… [have joined] the ranks of species written-off as commercially extinct."[1]
- Near the end of Joseph Heller's celebrated novel *Catch-22*, the Second World War is almost won. Yossarian does not want to be among the

> last to die; it won't make any difference to the outcome. He explains this to Major Danby, his superior officer. When Danby asks, "But, Yossarian, suppose everyone felt that way?" Yossarian replies, "Then I'd certainly be a damned fool to feel any other way, wouldn't I?"[2]

Answer: They are all instances of the prisoners' dilemma.[*] As in the interrogation of Dick Hickock and Perry Smith from *In Cold Blood* recounted in chapter 1, each has a personal incentive to do something that ultimately leads to a result that is bad for everyone when everyone similarly does what his or her personal interest dictates. If one confesses, the other had better confess to avoid the really harsh sentence reserved for recalcitrants; if one holds out, the other can cut himself a much better deal by confessing. Indeed, the force is so strong that each prisoner's temptation to confess exists regardless of whether the two are guilty (as was the case in *In Cold Blood*) or innocent and being framed by the police (as in the movie *L.A. Confidential).*

Price wars are no different. If the Nexon gas station charges a low price, the Lunaco station had better set its own price low to avoid losing too many customers; if Nexon prices its gas high, Lunaco can divert many customers its way by pricing low. But when both stations price low, neither makes money (though customers are better off).

If the Democrats adopt a platform that appeals to the middle, the Republicans may stand to lose all these voters and therefore the election if they cater only to their core supporters in the economic and social right wings; if the Democrats cater to their core supporters in the minorities and the unions, then the Republicans can capture the middle and therefore win a large majority by being more centrist. If all others fish conservatively, one fisherman going for a bigger catch is not going to deplete the fishery to any significant extent; if all others are fishing aggressively, then any single fisherman would be a fool to try single-handed conservation.[3] The result is overfishing and extinction. Yossarian's logic is what makes it so difficult to continue to support a failed war.

A LITTLE HISTORY

How did theorists devise and name this game that captures so many economic, political, and social interactions? It happened very early in the history of the subject. Harold Kuhn, himself one of the pioneers of game theory,

recounted the story in a symposium held in conjunction with the 1994 Nobel Prize award ceremonies:

> Al Tucker was on leave at Stanford in the Spring of 1950 and, because of the shortage of offices, he was housed in the Psychology Department. One day a psychologist knocked on his door and asked what he was doing. Tucker replied: "I'm working on game theory," and the psychologist asked if he would give a seminar on his work. For that seminar, Al Tucker invented prisoner's dilemma as an example of game theory, Nash equilibria, and the attendant paradoxes of non-socially-desirable equilibria. A truly seminal example, it inspired dozens of research papers and several entire books.[4]

Others tell a slightly different story. According to them, the mathematical structure of the game predates Tucker and can be attributed to two mathematicians, Merrill Flood and Melvin Dresher, at the Rand Corporation (a cold war think tank).[5] Tucker's genius was to invent the story illustrating the mathematics. And genius it was, because presentation can make or break an idea; a memorable presentation spreads and is assimilated in the community of thinkers far better and faster, whereas a dull and dry presentation may be overlooked or forgotten.

A Visual Representation

We will develop the method for displaying and solving the game using a business example. Rainbow's End and B. B. Lean are rival mail-order firms that sell clothes. Every fall they print and mail their winter catalogs. Each firm must honor the prices printed in its catalog for the whole winter season. The preparation time for the catalogs is much longer than the mailing window, so the two firms must make their pricing decisions simultaneously and without knowing the other firm's choices. They know that the catalogs go to a common pool of potential customers, who are smart shoppers and are looking for low prices.

Both catalogs usually feature an almost identical item, say a chambray deluxe shirt. The cost of each shirt to each firm is $20.* The firms have estimated that if they each charge $80 for this item, each will sell 1,200 shirts, so

each will make a profit of (80–20) × 1,200 = 72,000 dollars. Moreover, it turns out that this price serves their joint interests best: if the firms can collude and charge a common price, $80 is the price that will maximize their combined profits.

The firms have estimated that if one of them cuts its price by $1 while the other holds its price unchanged, then the price cutter gains 100 customers, 80 of whom shift to it from the other firm, and 20 who are new—for example, they might decide to buy the shirt when they would not have at the higher price or might switch from a store in their local mall. Therefore each firm has the temptation to undercut the other to gain more customers; the whole purpose of this story is to figure out how these temptations play out.

We begin by supposing that each firm chooses between just two prices, $80 and $70.† If one firm cuts its price to $70 while the other is still charging $80, the price cutter gains 1,000 customers and the other loses 800. So the price cutter sells 2,200 shirts while the other's sales drop to 400; the profits are (70–20) × 2,200 = $110,000 for the price cutter, and (80–20) × 400 = $24,000 for the other firm.

What happens if both firms cut their price to $70 at the same time? If both firms reduce their price by $1, existing customers stay put, but each gains the 20 new customers. So when both reduce their price by $10, each gains 10 × 20 = 200 net sales above the previous 1,200. Each sells 1,400 and makes a profit of (70–20) × 1,400 = $70,000.

We want to display the profit consequences (the firms' payoffs in their game) visually. However, we cannot do this using a game tree like the ones in chapter 2. Here the two players act simultaneously. Neither can make his move knowing what the other has done or anticipating how the other will respond. Instead, each must think about what the other is thinking at the same time. A starting point for this thinking about thinking is to lay out all the consequences of all the combinations of the simultaneous choices the two could make. Since each has two alternatives, $80 or $70, there are four such combinations. We can display them most easily in a spreadsheet-like format of rows and columns, which we will generally refer to as a game table or payoff table. The choices of Rainbow's End (RE for short) are arrayed along the rows, and those of B. B. Lean (BB) along the columns. In each of the four cells corresponding to each choice of a row by RE and of a column by BB, we show two numbers—the profit, in thousands of dollars, from selling this shirt. In each cell, the number in the southwest corner belongs to the row player, and the number in the northeast corner belongs to the column player.* In the jargon of game theory, these numbers are called payoffs.* To make it abundantly clear which payoffs belong

to which player, we have also put the numbers in two different shades of gray for this example.

		B. B. Lean (BB)	
		80	70
Rainbow's End (RE)	80	72,000 72,000	110,000 24,000
	70	24,000 110,000	70,000 70,000

Before we "solve" the game, let us observe and emphasize one feature of it. Compare the payoff pairs across the four cells. A better outcome for RE does not always imply a worse outcome for BB, or vice versa. Specifically, both of them are better off in the top left cell than in the bottom right cell. This game need not end with a winner and a loser; it is not zero-sum. We similarly pointed out in chapter 2 that the Charlie Brown investment game was not zero-sum, and neither are most games we meet in reality. In many games, as in the prisoners' dilemma, the issue will be how to avoid a lose-lose outcome or to achieve a win-win outcome.

The Dilemma

Now consider the reasoning of RE's manager. "If BB chooses $80, I can get $110,000 instead of $72,000 by cutting my price to $70. If BB chooses $70, then my payoff is $70,000 if I also charge $70, but only $24,000 if I charge $80. So, in both cases, choosing $70 is better than choosing $80. My better choice (in fact my best choice, since I have only two alternatives) is the same no matter what BB chooses. I don't need to think through their thinking at all; I should just go ahead and set my price at $70."

When a simultaneous-move game has this special feature, namely that for a player the best choice is the same regardless of what the other player or players choose, it greatly simplifies the players' thinking and the game theorists' analysis. Therefore it is worth making a big deal of it, and looking for it to simplify the solution of the game. The name given by game theorists for this property is *dominant strategy*. A player is said to have a dominant strategy if that same strategy is better for him than all of his other available strategies no matter what strategy or strategy combination the other player or players choose. And

we have a simple rule for behavior in simultaneous-move games:*

RULE 2: If you have a dominant strategy, use it.

The prisoners' dilemma is an even more special game—not just one player but both (or all) players have dominant strategies. The reasoning of BB's manager is exactly analogous to that of RE's manager, and you should fix the idea by going through it on your own. You will see that $70 is also the dominant strategy for BB.

The result is the outcome shown in the bottom right cell of the game table; both charge $70 and make a profit of $70,000 each. And here is the feature that makes the prisoners' dilemma such an important game. When both players use their dominant strategies, both do worse than they would have if somehow they could have jointly and credibly agreed that each would choose the other, dominated strategy. In this game, that would have meant charging $80 each to obtain the outcome in the top left cell of the game table, namely $72,000 each.*

It would not be enough for just one of them to price at $80; then that firm would do very badly. Somehow they must both be induced to price high, and this is hard to achieve given the temptation each of them has to try to undercut the other. Each firm pursuing its own self-interest does not lead to an outcome that is best for them all, in stark contrast to what conventional theories of economics from Adam Smith onward have taught us.†

This opens up a host of questions, some of which pertain to more general aspects of game theory. What happens if only one player has a dominant strategy? What if none of the players has a dominant strategy? When the best choice for each varies depending on what the other is choosing simultaneously, can they see through each other's choices and arrive at a solution to the game? We will take up these questions in the next chapter, where we develop a more general concept of solution for simultaneous-move games, namely Nash equilibrium. In this chapter we focus on questions about the prisoners' dilemma game per se.

In the general context, the two strategies available to each player are labeled "Cooperate" and "Defect" (or sometimes "Cheat"), and we will follow this usage. Defect is the dominant strategy for each, and the combination where both choose Defect yields a worse outcome for both than if both choose Cooperate.

Some Preliminary Ideas for Resolving the Dilemma

The players caught on the horns of this dilemma have strong incentives to make joint agreements to avoid it. For example, the fishermen in New England might agree to limit their catch to preserve the fish stocks for the future. The difficulty is to make such agreements stick, when each faces the temptation to cheat, for example, to take more than one's allotted quota of fish. What does game theory have to say on this issue? And what happens in the actual play of such games?

In the fifty years since the prisoners' dilemma game was invented, its theory has advanced a great deal, and much evidence has accumulated, both from observations about the real world and from controlled experiments in laboratory settings. Let us look at all this material and see what we can learn from it.

The flip side of achieving cooperation is avoiding defection. A player can be given the incentive to choose cooperation rather than the originally dominant strategy of defection by giving him a suitable reward, or deterred from defecting by creating the prospect of a suitable punishment.

The reward approach is problematic for several reasons. Rewards can be internal—one player pays the other for taking the cooperative action. Sometimes they can be external; some third party that also benefits from the two players' cooperation pays them for cooperating. In either case, the reward cannot be given before the choice is made; otherwise the player will simply pocket the reward and then defect. If the reward is merely promised, the promise may not be credible: after the promisee has chosen cooperation, the promisor may renege.

These difficulties notwithstanding, rewards are sometimes feasible and useful. At an extreme of creativity and imagination, the players could make simultaneous and mutual promises and make these credible by depositing the promised rewards in an escrow account controlled by a third party.[6] More realistically, sometimes the players interact in several dimensions, and cooperation in one can be rewarded with reciprocation in another. For example, among groups of female chimpanzees, help with grooming is reciprocated by sharing food or help with child minding. Sometimes third parties may have sufficiently strong interests in bringing about cooperation in a game. For example, in the interest of bringing to an end various conflicts around the world, the United States and the European Union have from time to time promised economic assistance to combatants as a reward for peaceful resolutions of their disputes. The United States rewarded Israel and Egypt in this way for cooperating to strike the Camp David Accords in 1978.

Punishment is the more usual method of resolving prisoners' dilemmas. This

could be immediate. In a scene from the movie *L.A. Confidential*, Sergeant Ed Exley promises Leroy Fontaine, one of the suspects he is interrogating, that if he turns state's witness, he will get a shorter sentence than the other two, Sugar Ray Coates and Tyrone Jones. But Leroy knows that, when he emerges from jail, he may find friends of the other two waiting for him!

But the punishment that comes to mind most naturally in this context arises from the fact that most such games are parts of an ongoing relationship. Cheating may gain one player a short-term advantage, but this can harm the relationship and create a longer-run cost. If this cost is sufficiently large, that can act as a deterrent against cheating in the first place.*

A striking example comes from baseball. Batters in the American League are hit by pitches 11 to 17 percent more often than their colleagues in the National League. According to Sewanee professors Doug Drinen and John-Charles Bradbury, most of this difference is explained by the designated hitter rule.[7] In the American League, the pitchers don't bat. Thus an American League pitcher who plunks a batter doesn't have to fear direct retaliation from the opposing team's pitcher. Although pitchers are unlikely to get hit, the chance goes up by a factor of four if they have just plunked someone in the previous half inning. The fear of retaliation is clear. As ace pitcher Curt Schilling explained: "Are you seriously going to throw at someone when you are facing Randy Johnson?"[8]

When most people think about one player punishing the other for past cheating, they think of some version of tit for tat. And that was indeed the finding of what is perhaps the most famous experiment on the prisoners' dilemma. Let us recount what happened and what it teaches.

TIT FOR TAT

In the early 1980s, University of Michigan political scientist Robert Axelrod invited game theorists from around the world to submit their strategies for playing the prisoners' dilemma in the form of computer programs. The programs were matched against each other in pairs to play a prisoners' dilemma game repeated 150 times. Contestants were then ranked by the sum of their scores.

The winner was Anatol Rapoport, a mathematics professor at the University of Toronto. His winning strategy was among the simplest: tit for tat. Axelrod was surprised by this. He repeated the tournament with an enlarged set of contestants. Once again Rapoport submitted tit for tat and beat the competition.

Tit for tat is a variation of the eye for an eye rule of behavior: Do unto others as they have done onto you.* More precisely, the strategy cooperates in the first

period and from then on mimics the rival's action from the previous period.

Axelrod argues that tit for tat embodies four principles that should be present in any effective strategy for the repeated prisoners' dilemma: clarity, niceness, provocability, and forgivingness. Tit for tat is as *clear* and simple as you can get; the opponent does not have to do much thinking or calculation about what you are up to. It is *nice* in that it never initiates cheating. It is *provocable*—that is, it never lets cheating go unpunished. And it is *forgiving*, because it does not hold a grudge for too long and is willing to restore cooperation.

One of the impressive features about tit for tat is that it did so well overall even though it did not (nor could it) beat any one of its rivals in a head-on competition. At best, tit for tat ties its rival. Hence if Axelrod had scored each competition as a winner-take-all contest, tit for tat would have only losses and ties and therefore could not have had the best track record.*

But Axelrod did not score the pairwise plays as winner-take-all: close counted. The big advantage of tit for tat is that it always comes close. At worst, tit for tat ends up getting beaten by one defection—that is, it gets taken advantage of once and then ties from then on.

The reason tit for tat won the tournament is that it usually managed to encourage cooperation whenever possible while avoiding exploitation. The other entries either were too trusting and open to exploitation or were too aggressive and knocked one another out.

In spite of all this, we believe that tit for tat is a flawed strategy. The slightest possibility of a mistake or a misperception results in a complete breakdown in the success of tit for tat. This flaw was not apparent in the artificial setting of a computer tournament, because mistakes and misperceptions did not arise. But when tit for tat is applied to real-world problems, mistakes and misperceptions cannot be avoided, and the result can be disastrous.

The problem with tit for tat is that any mistake "echoes" back and forth. One side punishes the other for a defection, and this sets off a chain reaction. The rival responds to the punishment by hitting back. This response calls for a second punishment. At no point does the strategy accept a punishment without hitting back.

Suppose, for example, that both Flood and Dresher start out playing tit for tat. No one initiates a defection, and all goes well for a while. Then, in round 11, say, suppose Flood chooses Defect by mistake, or Flood chooses Cooperate but Dresher mistakenly thinks Flood chose Defect. In either case, Dresher will play Defect in round 12, but Flood will play Cooperate because Dresher played Cooperate in round 11. In round 13 the roles will be switched. The pattern of one playing Cooperate and the other playing Defect will continue back and forth,

until another mistake or misperception restores cooperation or leads both to defect.

Such cycles or reprisals are often observed in real-life feuds between Israelis and Arabs in the Middle East, or Catholics and Protestants in Northern Ireland, or Hindus and Muslims in India. Along the West Virginia–Kentucky border, we had the memorable feud between the Hatfields and the McCoys. And in fiction, Mark Twain's Grangerfords and Shepherdsons offer another vivid example of how tit for tat behavior can end in a cycle of reprisals. When Huck Finn tries to understand the origins of the Grangerford-Shepherdson feud, he runs into the chicken-or-egg problem:

> "What was the trouble about, Buck?—land?"
> "I reckon maybe—I don't know."
> "Well, who done the shooting? Was it a Grangerford or a Shepherdson?"
> "Laws, how do I know? It was so long ago."
> "Don't anybody know?"
> "Oh, yes, pa knows, I reckon, and some of the other old people; but they don't know now what the row was about in the first place."

What tit for tat lacks is a way of saying "Enough is enough." It is too provocable, and not forgiving enough. And indeed, subsequent versions of Axelrod's tournament, which allowed possibilities of mistakes and misperceptions, showed other, more generous strategies to be superior to tit for tat.*

Here we might even learn something from monkeys. Cotton-top tamarin monkeys were placed in a game where each had the opportunity to pull a lever that would give the other food. But pulling the lever required effort. The ideal for each monkey would be to shirk while his partner pulled the lever. But the monkeys learned to cooperate in order to avoid retaliation. Tamarin cooperation remained stable as long as there were no more than two consecutive defections by one player, a strategy that resembles tit for two tats.[9]

MORE RECENT EXPERIMENTS

Thousands of experiments on the prisoners' dilemma have been performed in classrooms and laboratories, involving different numbers of players, repetitions, and other treatments. Here are some important findings.[10]

First and foremost is that cooperation occurs significantly often, even when each pair of players meets only once. On average, almost half of the players choose the cooperative action. Indeed, the most striking demonstration of this was on the Game Show Network's production of *Friend or Foe*. In this show, two-person teams were asked trivia questions. The money earned from correct answers went into a "trust fund," which over the 105 episodes ranged from $200 to $16,400. To divide the trust fund, the two contestants played a one-shot dilemma.

Each privately wrote down "friend" or "foe." When both wrote down friend, the pot was split evenly. If one wrote down foe while the other wrote friend, the person writing foe would get the whole pot. But if both wrote foe, then neither would get anything. Whatever the other side does, you get at least as much, and possibly more, by writing down foe than if you wrote friend. Yet almost half the contestants wrote down friend. Even as the pot grew larger there was no change in the likelihood of cooperation. People were as likely to cooperate when the fund was below $3,000 as they were when it was above $5,000. These were some of the findings in a pair of studies by Professors Felix Oberholzer-Gee, Joel Waldfogel, Matthew White, and John List.[11]

If you are wondering how watching television counts as academic research, it turns out that more than $700,000 was paid out to contestants. This was the best-funded experiment on the prisoners' dilemma, ever. There was much to learn. It turns out that women were much more likely to cooperate than men, 53.7 percent versus 47.5 percent in season 1. The contestants in season 1 didn't have the advantage of seeing the results from the other matches before making their decision. But in season 2, the results of the first 40 episodes had been aired and this pattern became apparent. The contestants had learned from the experience of others. When the team consisted of two women, the cooperation rate rose to 55 percent. But when a woman was paired with a guy, her cooperation rate fell to 34.2 percent. And the guy's rate fell, too, down to 42.3 percent. Overall, cooperation dropped by ten points.

When a group of subjects is assembled and matched pairwise a number of times, with different pairings at different times, the proportion choosing cooperation generally declines over time. However, it does not go to zero, settling instead on a small set of persistent cooperators.

If the same pair plays the basic dilemma game repeatedly, they often build up to a significant sequence of mutual cooperation, until one player defects near the end of the sequence of repetitions. This happened in the very first experiment conducted on the dilemma. Almost immediately after they had thought up the game, Flood and Dresher recruited two of their colleagues to play the dilemma

game a hundred times.[12] On 60 of these rounds, both players chose Cooperate. A long stretch of mutual cooperation lasted from round 83 to round 98, until one player sneaked in a defection in round 99.

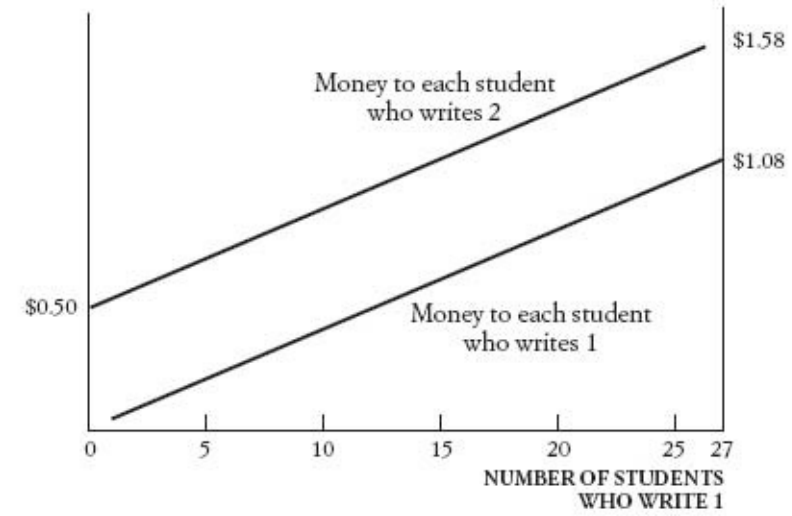
Actually, according to the strict logic of game theory, this should not have happened. When the game is repeated exactly 100 times, it is a sequence of simultaneous-move games, and we can apply the logic of backward reasoning to it. Look ahead to what will happen on the 100th play. There are no more games to come, so defection cannot be punished in any future rounds. Dominant strategy calculations dictate that both players should choose Defect on the last round. But once that is a given, the 99th round becomes effectively the last round. Although there is one more round to come, defection on the 99th round is not going to be selectively punished by the other player in the 100th round because his choice in that round is foreordained. Therefore the logic of dominant strategies applies to the 99th round. And one can work back this sequential logic all the way to round 1. But in actual play, both in the laboratory and the real world, players seem to ignore this logic and achieve the benefits of mutual cooperation. What may seem at first sight to be irrational behavior—departing from one's dominant strategy—turns out to be a good choice, so long as everyone else is similarly "irrational."

Game theorists suggest an explanation for this phenomenon. The world contains some "reciprocators," people who will cooperate so long as the other does likewise. Suppose you are not one of these relatively nice people. If you behaved true to your type in a finitely repeated game of prisoners' dilemma, you would start cheating right away. That would reveal your nature to the other player. To hide the truth (at least for a while), you have to behave nicely. Why would you want to do that? Suppose you started by acting nicely. Then the other player, even if he is not a reciprocator, would think it possible that you are one of the few nice people around. There are real gains to be had by cooperating for a while, and the other player would plan to reciprocate your niceness to achieve these gains. That helps you, too. Of course you are planning to sneak in a defection near the end of the game, just as the other player is. But you two can still have an initial phase of mutually beneficial cooperation. While each side is waiting to take advantage of the other, both are benefiting from this mutual deception.

In some experiments, instead of pairing each subject in the group with another person and playing several two-person dilemmas, the whole group is engaged in one large multiperson dilemma. We mention a particularly entertaining and instructive instance from the classroom. Professor Raymond

Battalio of Texas A&M University had his class of 27 students play the following game.[13] Each student owned a hypothetical firm and had to decide (simultaneously and independently, by writing on a slip of paper) whether to produce 1 and help keep the total supply low and the price high or produce 2 and gain at the expense of others. Depending on the total number of students producing 1, money would be paid to students according to the following table:

Number of students who write 1	Payoff to each student who writes 1	Payoff to each student who writes 2
0		$0.50
1	$0.04	$0.54
2	$0.08	$0.58
3	$0.12	$0.62
. . .	. . .	. . .
25	$1.00	$1.50
26	$1.04	$1.54
27	$1.08	



This is easier to see and more striking in a chart: The game is "rigged" so that students who write 2 (Defect) always get 50 cents more than those who write 1 (Cooperate), but the more of them that write 2, the less their collective gain. Suppose all 27 start planning to write 1, so each would get $1.08. Now one thinks of sneaking a switch to 2. There would be 26 1s, and each would get $1.04 (4 cents less than in the original plan), but the switcher would get $1.54

(46 cents more). The same is true irrespective of the initial number of students thinking of writing 1 versus 2. Writing 2 is a dominant strategy. Each student who switches from writing 1 to writing 2 increases his own payout by 46 cents but decreases that of each of his 26 colleagues by 4 cents—the group as a whole loses 58 cents. By the time everyone acts selfishly, each maximizing his own payoff, they each get 50 cents. If they could have successfully conspired and acted so as to minimize their individual payoff, they would each receive $1.08. How would you play?

In some practice plays of this game, first without classroom discussion and then with some discussion to achieve a "conspiracy," the number of cooperative students writing 1 ranged from 3 to a maximum of 14. In a final binding play, the number was 4. The total payout was $15.82, which is $13.34 less than that from totally successful collusion. "I'll never trust anyone again as long as I live," muttered the conspiracy leader. And what was his choice? "Oh, I wrote 2," he replied. Yossarian would have understood.

More recent laboratory experiments of multiperson dilemmas use a format called the contribution game. Each player is given an initial stake, say $10. Each can choose to keep part of this and contribute a part to a common pool. The experimenter then doubles the accumulated common pool and divides this equally among all the players, contributors and noncontributors alike.

Suppose there are four players, say A, B, C, and D, in the group. Regardless of what the others are doing, if person A contributes a dollar to the common pool, this increases the common pool by $2 after the doubling. But $1.50 of the increment goes to B, C, and D; A gets only 50 cents. Therefore A loses out by raising his contribution; conversely he would gain by lowering it. And that is true no matter how much, if anything, the others are contributing. In other words, contributing nothing is the dominant strategy for A. The same goes for B, C, and D. This logic says that each should hope to become a "free rider" on the efforts of the others. If all four play their dominant strategy, the common pool is empty and each simply keeps the initial stake of $10. When everyone tries to be a free rider, the bus stays in the garage. If everyone had put all of their initial stakes in the common pool, the pool after doubling would be $80 and the share of each would be $20. But each has the personal incentive to cheat on such an arrangement. This is their dilemma.

The contribution game is not a mere curiosity of the laboratory or theory; it is played in the real world in social interactions where some communal benefit can be achieved by voluntary contributions from members of the group, but the benefit cannot be withheld from those who did not contribute. Flood control in a village, or conservation of natural resources, are cases in point: it is not possible

to build levees or dams so that flood waters will selectively go to the fields of those who did not help in the construction, and it is not practicable to withhold gas or fish in the future from someone who consumed too much in the past. This creates a multiperson dilemma: each player has the temptation to shirk or withhold his contribution, hoping to enjoy the benefits of the others' contributions. When they all think this way, the total of contributions is meager or even zero, and they all suffer. These situations are ubiquitous, and of such magnitude that all of social theory and policy needs a good understanding of how the dilemmas might be resolved.

In what is perhaps the most interesting variant of the game, players are given an opportunity to punish those who cheat on an implicit social contract of cooperation. However, they must bear a personal cost to do so. After the contribution game is played, the players are informed about the individual contributions of other players. Then a second phase is played, where each player can take an action to lower the payoffs of other players at a cost to himself of so many cents (typically 33) per dollar reduction chosen. In other words, if player A chooses to reduce B's payoff by three dollars, then A's payoff is reduced by one dollar. These reductions are not reallocated to anyone else; they simply return to the general funds of the experimenter.

The results of the experiment show that people engage in a significant amount of punishment of "social cheaters," and that the prospect of the punishment increases the contributions in the first phase of the game dramatically. Such punishments seem to be an effective mechanism for achieving cooperation that benefits the whole group. But the fact that individuals carry them out is surprising at first. The act of punishing others at a personal cost is itself a contribution for the general benefit, and it is a dominated strategy; if it succeeds in eliciting better behavior from the cheater in the future, its benefits will be for the group as a whole, and the punisher will get only his small share of this benefit. Therefore the punishment has to be the result of something other than a selfish calculation. That is indeed the case. Experiments on this game have been conducted while the players' brains were being imaged by PET scan.[14] These revealed that the act of imposing the penalty activated a brain region called the dorsal striatum, which is involved in experiencing pleasure or satisfaction. In other words, people actually derive a psychological benefit or pleasure from punishing social cheaters. Such an instinct must have deep biological roots and may have been selected for an evolutionary advantage.[15]

HOW TO ACHIEVE COOPERATION

These examples and experiments have suggested several preconditions and strategies for successful cooperation. Let us develop the concepts more systematically and apply them to some more examples from the real world.

Successful punishment regimes must satisfy several requirements. Let us examine these one by one.

Detection of cheating: Before cheating can be punished, it must be detected. If detection is fast and accurate, the punishment can be immediate and accurate. That reduces the gain from cheating while increasing its cost, and thus increases the prospects for successful cooperation. For example, airlines constantly monitor each other's fares; if American were to lower its fare from New York to Chicago, United can respond in under five minutes. But in other contexts, firms that want to cut their prices can do so in secret deals with the customers, or hide their price cuts in a complicated deal involving many dimensions of delivery time, quality, warranties, and so on. In extreme situations, each firm can only observe its own sales and profits, which can depend on some chance elements as well as on other firms' actions. For example, how much one firm sells can depend on the vagaries of demand, not just on other firms' secret price cuts. Then detection and punishment become not only slow but also inaccurate, raising the temptation to cheat.

Finally, when three or more firms are simultaneously in the market, they must find out not only whether cheating has occurred but who has cheated. Otherwise any punishments cannot be targeted to hurt the miscreant but have to be blunt, perhaps unleashing a price war that hurts all.

Nature of punishment: Next, there is the choice of punishment. Sometimes the players have available to them actions that hurt others, and these can be invoked after an instance of cheating even in a one-time interaction. As we pointed out in the dilemma in *L.A. Confidential*, the friends of Sugar and Tyrone will punish Leroy when he emerges from jail after his light sentence for turning state's witness. In the Texas A&M classroom experiment, if the students could detect who had reneged on the conspiracy for all of them to write 1, they could inflict social sanctions such as ostracism on the cheaters. Few students would risk that for the sake of an extra 50 cents.

Other kinds of punishments arise within the structure of the game. Usually

this happens because the game is repeated, and the gain from cheating in one play leads to a loss in future plays. Whether this is enough to deter a player who is contemplating cheating depends on the sizes of the gains and losses and on the importance of the future relative to the present. We will return to this aspect soon.

Clarity: The boundaries of acceptable behavior, and the consequences of cheating, should be clear to a prospective cheater. If these things are complex or confusing, the player may cheat by mistake or fail to make a rational calculation and play by some hunch. For example, suppose Rainbow's End and B. B. Lean are playing their price-setting game repeatedly, and RE decides that it will infer that BB has cheated if RE's discounted mean of profits from the last seventeen months is 10 percent less than the average real rate of return to industrial capital over the same period. BB does not know this rule directly; it must infer what rule RE is using by observing RE's actions. But the rule stated here is too complicated for BB to figure out. Therefore it is not a good deterrent against BB's cheating. Something like tit for tat is abundantly clear: if BB cheats, it will see RE cutting its price the very next time.

Certainty: Players should have confidence that defection will be punished and cooperation rewarded. This is a major problem in some international agreements like trade liberalization in the World Trade Organization (WTO). When one country complains that another has cheated on the trade agreement, the WTO initiates an administrative process that drags on for months or years. The facts of the case have little bearing on the judgment, which usually depends more on dictates of international politics and diplomacy. Such enforcement procedures are unlikely to be effective.

Size: How harsh should such punishments be? It might seem that there is no limit. If the punishment is strong enough to deter cheating, it need never actually be inflicted. Therefore it may as well be set at a sufficiently high level to ensure deterrence. For example, the WTO could have a provision to nuke any nation that breaks its undertakings to keep its protective tariffs at the agreed low levels. Of course you recoil in horror at the suggestion, but that is at least in part because you think it possible that some error may cause this to happen. When

errors are possible, as they always are in practice, the size of the punishment should be kept as low as is compatible with successful deterrence in most circumstances. It may even be optimal to forgive occasional defection in extreme situations—for example, a firm that is evidently fighting for its survival may be allowed some price cuts without triggering reactions from rivals.

Repetition: Look at the pricing game between Rainbow's End and B. B. Lean. Suppose they are going merrily along from one year to the next, holding prices at their joint best, $80. One year the management of RE considers the possibility of cutting the price to $70. They reckon that this will yield them an extra profit of $110,000–$72,000 = $38,000. But that can lead to a collapse of trust. RE should expect that in future years BB will also choose $70, and each will make only $70,000 each year. If RE had kept to the original arrangement, each would have kept on making $72,000. Thus RE's price cutting will cost it $72,000–$70,000 = $2,000 every year in the future. Is a one-time gain of $38,000 worth the loss of $2,000 every year thereafter?

One key variable that determines the balance of present and future considerations is the interest rate. Suppose the interest rate is 10% per year. Then RE can stash away its extra $38,000 and earn $3,800 every year. That comfortably exceeds the loss of $2,000 in each of those years. Therefore cheating is in RE's interest. But if the interest rate is only 5% per year, then the $38,000 earns only $1,900 in each subsequent year, less than the loss of $2,000 due to the collapse of the arrangement; so RE does not cheat. The interest rate at which the two magnitudes just balance is 2/38 = 0.0526, or 5.26% per year.

The key idea here is that when interest rates are low, the future is relatively more valuable. For example, if the interest rate is 100%, then the future has low value relative to the present—a dollar in a year's time is worth only 50 cents right now because you can turn the 50 cents into a dollar in a year by earning another 50 cents as interest during the year. But if the interest rate is zero, then a dollar in a year's time is worth the same as a dollar right away.*

In our example, for realistic interest rates a little above 5%, the temptation for each firm to cut the price by $10 below their joint best price of $80 is quite finely balanced, and collusion in a repeated game may or may not be possible. In chapter 4 we will see how low the price can fall if there is no shadow of the future and the temptation to cheat is irresistible.

Another relevant consideration is the likelihood of continuation of the relationship. If the shirt is a transient fashion item that may not sell at all next

year, then the temptation to cheat this year is not offset by any prospect of future losses.

But Rainbow's End and B. B. Lean sell many items besides this shirt. Won't cheating on the shirt price bring about retaliation on all those other items in the future? And isn't the prospect of this huge retaliation enough to deter the defection? Alas, the usefulness of multiproduct interactions for sustaining cooperation is not so simple. The prospect of multiproduct retaliation goes hand in hand with that of immediate gains from simultaneous cheating in all of those dimensions, not just one. If all the products had identical payoff tables, then the gains and losses would both increase by a factor equal to the number of products, and so whether the balance is positive or negative would not change. Therefore successful punishments in multiproduct dilemmas must depend in a more subtle way on differences among the products.

A third relevant consideration is the expected variation in the size of the business over time. This has two aspects—steady growth or decline, and fluctuations. If the business is expected to grow, then a firm considering defection now will recognize that it stands to lose more in the future due to the collapse of the cooperation and will be more hesitant to defect. Conversely, if the business is on a path of decline, then firms will be more tempted to defect and take what they can now, knowing that there is less at stake in the future. As for fluctuations, firms will be more tempted to cheat when a temporary boom arrives; cheating will bring them larger immediate profits, whereas the downside from the collapse of the cooperation will hit them in the future, when the volume of business will be only the average, by definition of the average. Therefore we should expect that price wars will break out during times of high demand. But this is not always the case. If a period of low demand is caused by a general economic downturn, then the customers will have lower incomes and may become sharper shoppers as a result—their loyalties to one firm or the other may break down, and they may respond more quickly to price differences. In that case, a firm cutting its price can expect to attract more customers away from its rival, and thereby reap a larger immediate gain from defection.

Finally, the composition of the group of players is important. If this is stable and expected to remain so, that is conducive to the maintenance of cooperation. New players who do not have a stake or a history of participation in the cooperative arrangement are less likely to abide by it. And if the current group of players expects new ones to enter and shake up the tacit cooperation in the future, that increases their own incentive to cheat and take some extra benefit right now.

SOLUTION BY KANTIAN CATEGORICAL IMPERATIVE?

It is sometimes said that the reason some people cooperate in the prisoners' dilemma is that they are making the decision not only for themselves but for the other player. That is wrong in point of fact, but the person is acting as if this is the case.

The person truly wants the other side to cooperate and reasons to himself that the other side is going through the same logical decision process that he is. Thus the other side must come to the same logical conclusion that he has. Hence if the player cooperates, he reasons that the other side will do so as well, and if he defects, he reasons that it will cause the other side to defect. This is similar to the categorical imperative of the German philosopher Immanuel Kant: "Take only such actions as you would like to see become a universal law."

Of course, nothing could be further from the truth. The actions of one player have no effect whatsoever on the other player in the game. Still people think that somehow their actions can influence the choice of others, even when their actions are invisible.

The power of this thinking was revealed in an experiment done with Princeton undergraduates by Eldar Shafir and Amos Tversky.[16] In their experiment, they put students in a prisoners' dilemma game. But unlike the usual dilemma, in some treatments they told one side what the other had done. When students were told that the other side had defected on them, only 3 percent responded with cooperation. When told that the other side had cooperated, this increased cooperation levels up to 16 percent. It was still the case that the large majority of students were willing to act selfishly. But many were willing to reciprocate the cooperative behavior exhibited by the other side, even at their own expense.

What do you think would happen when the students were not told anything about the other player's choice at all? Would the percentage of cooperators be between 3 and 16 percent? No; it rose to 37 percent. At one level, this makes no sense. If you wouldn't cooperate when you learned that the other side had defected and you wouldn't cooperate when you learned that the other side had cooperated, why would you then cooperate when you don't know what the other side had done?

Shafir and Tversky call this "quasi-magical" thinking—the idea that by taking some action, you can influence what the other side will do. People realize they can't change what the other side has done once they've been told what the other side has done. But if it remains open or undisclosed, then they imagine that

their actions might have some influence—or that the other side will somehow be employing the same reasoning chain and reach the same outcome they do. Since Cooperate, Cooperate is preferred to Defect, Defect, the person chooses Cooperate.

We want to be clear that such logic is completely illogical. What you do and how you get there has no impact at all on what the other side thinks and acts. They have to make up their mind without reading your mind or seeing your move. However, the fact remains that if the people in a society engage in such quasi-magical thinking, they will not fall victim to many prisoners' dilemmas and all will reap higher payoffs from their mutual interactions. Could it be that human societies deliberately instill such thinking into their members for just such an ultimate purpose?

DILEMMAS IN BUSINESS

Armed with the tool kit of experimental findings and theoretical ideas in the previous sections, let us step outside the laboratory and look at some instances of prisoners' dilemmas in the real world and attempts at resolving them.

Let us begin with the dilemma of rival firms in an industry. Their joint interests are best served by monopolizing or cartelizing the industry and keeping prices high. But each firm can do better for itself by cheating on such an agreement and sneaking in price cuts to steal business from its rivals. What can the firms do? Some factors conducive to successful collusion, such as growing demand or lack of disruptive entry, may be at least partially outside their control. But they can try to facilitate the detection of cheating and devise effective punishment strategies.

Collusion is easier to achieve if the firms meet regularly and communicate. Then they can negotiate and compromise on what are acceptable practices and what constitutes cheating. The process of negotiation and its memory contributes to clarity. If something occurs that looks prima facie like cheating, another meeting can help clarify whether it is something extraneous, an innocent error by a participant, or deliberate cheating. Therefore unnecessary punishments can be avoided. And the meeting can also help the group implement the appropriate punishment actions.

The problem is that the group's success in resolving their dilemma harms the general public's interest. Consumers must pay higher prices, and the firms withhold some supply from the market to keep the price high. As Adam Smith said, "People of the same trade seldom meet together, even for merriment and

diversion, but the conversation ends in a conspiracy against the public, or in some contrivance to raise prices."[17] Governments that want to protect the general public interest get into the game and enact antitrust laws that make it illegal for firms to collude in this way.* In the United States, the Sherman Antitrust Act prohibits conspiracies "in restraint of trade or commerce," of which price fixing or market-share fixing conspiracies are the prime instance and the ones most frequently attempted. In fact the Supreme Court has ruled that not only are explicit agreements of this kind forbidden, but also any explicit or tacit arrangement among firms that has the effect of price fixing is a violation of the Sherman Act, regardless of its primary intent. Violation of these laws can lead to jail terms for the firms' executives, not just fines for the corporations that are impersonal entities.

Not that firms don't try to get away with the illegal practices. In 1996 Archer Daniels Midland (ADM), a leading American processor of agricultural products, and their Japanese counterpart, Ajinomoto were caught in just such a conspiracy. They had arranged market sharing and pricing agreements for various products such as lysine (which is produced from corn and used for fattening up chickens and pigs). The aim was to keep the prices high at the expense of their customers. Their philosophy was: "The competitors are our friends, and the customers are our enemies." The companies' misdeeds came to light because one of the ADM negotiators became an informant for the FBI and arranged for many of the meetings to be recorded for audio and sometimes also video.[18]

An instance famous in antitrust history and business school case studies concerns the large turbines that generate electricity. In the 1950s, the U.S. market for these turbines consisted of three firms: GE was the largest, with a market share of around 60 percent, Westinghouse was the next, with approximately 30 percent, and Allied-Chalmers had about 10 percent. They kept these shares, and obtained high prices, using a clever coordination device. Here's how it worked. Electric utilities invited bids for the turbines they intended to buy. If the bid was issued during days 1–17 of a lunar month, Westinghouse and Allied-Chalmers had to put in very high bids that would be sure losers, and GE was the conspiracy's chosen winner by making the lowest bid (but still at a monopolist's price allowing big profits). Similarly, Westinghouse was the designated winner in the conspiracy if the bid was issued during days 18–25, and Allied-Chalmers for days 26–28. Since the utilities did not issue their solicitations for bids according to the lunar calendar, over time each of the three producers got the agreed market share. Any cheating on the agreement would have been immediately visible to the rivals. But, so long as the Department of

Justice did not think of linking the winners to the lunar cycles, it was safe from detection by the law. Eventually the authorities did figure it out, some executives of the three firms went to jail, and the profitable conspiracy collapsed. Different schemes were tried later.[19]

A variant of the turbine scheme later appeared in the bidding at the airwave spectrum auctions in 1996–1997. A firm that wanted the right for the licenses in a particular location would signal to the other firms its determination to fight for that right by using the telephone area code for that location as the last three digits of its bid. Then the other firms would let it win. So long as the same set of firms interacts in a large number of such auctions over time and so long as the antitrust authorities do not figure it out, the scheme may be sustainable.[20]

More commonly, the firms in an industry try to attain and sustain implicit or tacit agreements without explicit communication. This eliminates the risk of criminal antitrust action, although the antitrust authorities can take other measures to break up even implicit collusion. The downside is that the arrangement is less clear and cheating is harder to detect, but firms can devise methods to improve both.

Instead of agreeing on the prices to be charged, the firms can agree on a division of the market, by geography, product line, or some similar measure. Cheating is then more visible—your salespeople will quickly come to know if another company has stolen some of your assigned market.

Detection of price cuts, especially in the case of retail sales, can be simplified, and retaliation made quick and automatic, by the use of devices like “matching or beating competition” policies and most-favored-customer clauses. Many companies selling household and electronic goods loudly proclaim that they will beat any competitor’s price. Some even guarantee that if you find a better price for the same product within a month after your purchase, they will refund the difference, or in some cases even double the difference. At first sight, these strategies seem to promote competition by guaranteeing low prices. But a little gametheoretic thinking shows that in reality they can have exactly the opposite effect. Suppose Rainbow’s End and B. B. Lean had such policies, and their tacit agreement was to price the shirt at $80. Now each firm knows that if it sneaks a cut to $70, the rival will find out about it quickly; in fact the strategy is especially clever in that it puts the customers, who have the best natural incentive to locate low prices, in charge of detecting cheating. And the prospective defector also knows that the rival can retaliate instantaneously by cutting its own price; it does not have to wait until next year’s catalog is printed. Therefore the cheater is more effectively deterred.

Promises to meet or beat the competition can be clever and indirect. In the competition between Pratt & Whitney (P&W) and Rolls-Royce (RR) for jet aircraft engines to power Boeing 757 and 767 planes, P&W promised all prospective purchasers that its engines would be 8 percent more fuel-efficient than those of RR, otherwise P&W would pay the difference in fuel costs.[21]

A most-favored-customer clause says that the seller will offer to all customers the best price they offer to the most favored ones. Taken at face value, it seems that the manufacturers are guaranteeing low prices. But let's look deeper. The clause means that the manufacturer cannot compete by offering selective discounts to attract new customers away from its rival, while charging the old higher price to its established clientele. They must make general price cuts, which are more costly, because they reduce the profit margin on all sales. You can see the advantage of this clause to a cartel: the gain from cheating is less, and the cartel is more likely to hold.

A branch of the U.S. antitrust enforcement system, the Federal Trade Commission, considered such a clause that was being used by DuPont, Ethyl, and other manufacturers of antiknock additive compounds in gasoline. The commission ruled that there was an anticompetitive effect and forbade the companies from using such clauses in their contracts with customers.[*]

TRAGEDIES OF THE COMMONS

Among the examples at the start of this chapter, we mentioned problems like overfishing that arise because each person stands to benefit by taking more, while the costs of his action are visited upon numerous others or on future generations. University of California biologist Garrett Harding called this the "tragedy of the commons," using among his examples the overgrazing of commonly owned land in fifteenth-and sixteenth-century England.[22] The problem has become well known under this name. Today the problem of global warming is an even more serious example; no one gets enough private benefit from reducing carbon emissions, but all stand to suffer serious consequences when each follows his self-interest.

This is just a multiperson prisoners' dilemma, like the one Yossarian faced in *Catch-22* about risking his life in wartime. Of course societies recognize the costs of letting such dilemmas go unresolved and make attempts to achieve better outcomes. What determines whether these attempts succeed?

Indiana University political scientist Elinor Ostrom and her collaborators and students have conducted an impressive array of case studies of attempts to

resolve dilemmas of the tragedy of the commons—that is, to use and conserve common property resources in their general interest and avoid overexploitation and rapid depletion. They studied some successful and some unsuccessful attempts of this kind and derived some prerequisites for cooperation.[23]

First, there must be clear rules that identify who is a member of the group of players in the game—those who have the right to use the resource. The criterion is often geography or residence but can also be based on ethnicity or skills, or membership may be sold by auction or for an entry fee.*

Second, there must be clear rules defining permissible and forbidden actions. These include restrictions on time of use (open or closed seasons for hunting or fishing, or what kinds of crops can be planted and any requirements to keep the land fallow in certain years), location (a fixed position or a specified rotation for inshore fishing), the technology (size of fishing nets), and, finally, the quantity or fraction of the resource (amount of wood from a forest that each person is allowed to gather and take away).

Third, a system of penalties for violation of the above rules must be clear and understood by all parties. This need not be an elaborate written code; shared norms in stable communities can be just as clear and effective. The sanctions used against rule breakers range from verbal chastisement or social ostracism to fines, the loss of future rights, and, in some extreme cases, incarceration. The severity of each type of sanction can also be adjusted. An important principle is graduation. The first instance of suspected cheating is most commonly met simply by a direct approach to the violator and a request to resolve the problem. The fines for a first or second offense are low and are ratcheted up only if the infractions persist or get more blatant and serious.

Fourth, a good system to detect cheating must be in place. The best method is to make detection automatic in the course of the players' normal routine. For example, a fishery that has good and bad areas may arrange a rotation of the rights to the good areas. Anyone assigned to a good spot will automatically notice if a violator is using it and has the best incentive to report the violator to others and get the group to invoke the appropriate sanctions. Another example is the requirement that harvesting from forests or similar common areas must be done in teams; this facilitates mutual monitoring and eliminates the need to hire guards.

Sometimes the rules on what is permissible must be designed in the light of feasible methods of detection. For example, the size of a fisherman's catch is often difficult to monitor exactly and difficult even for a well-intentioned fisherman to control exactly. Therefore rules based on fish quantity quotas are rarely used. Quantity quotas perform better when quantities are more easily and

accurately observable, as in the case of water supplied from storage and harvesting of forest products.

Fifth, when the above categories of rules and enforcement systems are being designed, information that is easily available to the prospective users proves particularly valuable. Although each may have the temptation after the fact to cheat, they all have a common prior interest to design a good system. They can make the best use of their knowledge of the resource and of the technologies for exploiting it, the feasibility of detecting various infractions, and the credibility of various kinds of sanctions in their group. Centralized or top-down management has been demonstrated to get many of these things wrong and therefore perform poorly.

While Ostrom and her collaborators are generally optimistic about finding good solutions to many problems of collective action using local information and systems of norms, she gives a salutary warning against perfection: “The dilemma never fully disappears, even in the best operating systems…. No amount of monitoring or sanctioning reduces the temptation to zero. Instead of thinking of overcoming or conquering tragedies of the commons, effective governance systems cope better than others.”

NATURE RED IN TOOTH AND CLAW

As you might expect, prisoners’ dilemmas arise in species other than humans. In matters like building shelter, gathering food, and avoiding predators, an animal can act either selfishly in the interest of itself or its immediate kin, or in the interest of a larger group. What circumstances favor good collective outcomes? Evolutionary biologists have studied this question and found some fascinating examples and ideas. Here is a brief sample.[24]

The British biologist J. B. S. Haldane was once asked whether he would risk his life to save a fellow human being and replied: “For more than two brothers, or more than eight cousins, yes.” You share half of your genes with a brother (other than an identical twin), and one-eighth of your genes with a cousin; therefore such action increases the expected number of copies of your genes that propagate to the next generation. Such behavior makes excellent biological sense; the process of evolution would favor it. This purely genetic basis for cooperative behavior among close kin explains the amazing and complex cooperative behavior observed in ant colonies and beehives.

Among animals, altruism without such genetic ties is rare. But reciprocal altruism can arise and persist among members of a group of animals with much

less genetic identity, if their interaction is sufficiently stable and long-lasting. Hunting packs of wolves and other animals are examples of this. Here is an instance that is a bit gruesome but fascinating: Vampire bats in Costa Rica live in colonies of a dozen or so but hunt individually. On any day, some may be lucky and others unlucky. The lucky ones return to the hollow trees where the whole group lives and can share their luck by regurgitating the blood they have brought from their hunt. A bat that does not get a blood meal for three days is at risk of death. The colonies develop effective practices of mutual "insurance" against this risk by such sharing.[25]

University of Maryland biologist Gerald Wilkinson explored the basis of this behavior by collecting bats from different locations and putting them together. Then he systematically withheld blood from some of them and saw whether others shared with them. He found that sharing occurred only when the bat was on the verge of death, and not earlier. Bats seem to be able to distinguish real need from mere temporary bad luck. More interesting, he found that sharing occurred only among bats that already knew each other from their previous group, and that a bat was much more likely to share with another bat that had come to its aid in the past. In other words, the bats are able to recognize other individual bats and keep score of their past behavior in order to develop an effective system of reciprocal altruism.

CASE STUDY: THE EARLY BIRD KILLS THE GOLDEN GOOSE

The Galápagos Islands are the home of Darwin's finches. Life on these volcanic islands is difficult and so evolutionary pressures are high. Even a millimeter change in the beak of a finch can make all the difference in the competition for survival.[*]

Each island differs in its food sources, and finches' beaks reflect those differences. On Daphne Major, the primary food source is a cactus. Here the aptly named cactus finch has evolved so that its beak is ideally suited to gather the pollen and nectar of the cactus blossom.

The birds are not consciously playing a game against each other. Yet each adaptation of a bird's beak can be seen as its strategy in life. Strategies that provide an advantage in gathering food will lead to survival, a choice of mating partners, and more offspring. The beak of the finch is a result of this combination of natural and sexual selection.

Even when things seem to be working, genetics throws a few curveballs into the mix. There is the old saying that the early bird gets the worm. On Daphne

Major, it was the early finch that got the nectar. Rather than wait until nine in the morning when the cactus blossoms naturally open for business, a dozen finches were trying something new. They were prying open the cactus blossom to get a head start.

At first glance, this would seem to give these birds an edge over their latecoming rivals. The only problem is that in the process of prying open the blossom, the birds would often snip the stigma. As Weiner explains:

> [The stigma] is the top of the hollow tube that pokes out like a tall straight straw from the center of each blossom. When the stigma is cut, the flower is sterilized. The male sex cells in the pollen cannot reach the female sex cells in the flower. The cactus flower withers without bearing fruit.[26]

When the cactus flowers wither, the main source of food disappears for the cactus finch. You can predict the end result of this strategy: no nectar, no pollen, no seeds, no fruit, and then no more cactus finch. Does that mean that evolution has led the finches into a prisoners' dilemma where the eventual outcome is extinction?

Case Discussion

Not quite, on two counts. Finches are territorial and so the finches (and their offspring) whose local cactus shut down may end up as losers. Killing next year's neighborhood food supply is not worth today's extra sip of pollen. Therefore these deviant finches would not appear to have a fitness advantage over the others. But that conclusion changes if this strategy ever becomes pervasive. The deviant finches will expand their search for food and even those finches that wait will not save their cactus's stigma. Given the famine that is sure to follow, the birds most likely to survive are those who started in the strongest position. The extra sip of nectar could make the difference.

What we have here is a cancerous adaptation. If it stays small, it can die out. But if it ever grows too large, it will become the fittest strategy on a sinking ship. Once it ever becomes advantageous even on a relative scale, the only way to get rid of it is to eliminate the entire population and start again. With no finches left on Daphne Major, there will be no one left to snip the stigmas and the cacti will bloom again. When two lucky finches alight on this island, they will have an

opportunity to start the process from scratch.

The game we have here is a cousin to the prisoners' dilemma, a life and death case of the "stag hunt" game analyzed by the philosopher Jean-Jacques Rousseau.* In the stag hunt, if everyone works together to capture the stag, they succeed and all eat well. A problem arises if some hunters come across a hare along the way. If too many hunters are sidetracked chasing after hares, there won't be enough hunters left to capture the stag. In that case, everyone will do better chasing after rabbits. The best strategy is to go after the stag *if and only if* you can be confident that most everyone is doing the same thing. You have no reason not to chase after the stag, except if you lack confidence in what others will do.

The result is a confidence game. There are two ways it can be played. Everyone works together and life is good. Or everyone looks out for themselves and life is nasty, brutish, and short. This is not the classic prisoners' dilemma in which each person has an incentive to cheat no matter what others do. Here, there is no incentive to cheat, so long as you can trust others to do the same. But can you trust them? And even if you do, can you trust them to trust you? Or can you trust them to trust you to trust them? As FDR famously said (in a different context), we have nothing to fear but fear itself.

For more practice with prisoners' dilemmas, have a look at the following case studies in chapter 14: "What Price a Dollar?" and "The King Lear Problem."