Купить книгу "Good Thinking"

Good Thinking

Seven Powerful Ideas
That Influence the Way
We Think

Denise D. Cummins

TWO

Game Theory

WHEN YOU ARE NOT THE ONLY ONE CHOOSING

John and Mary are trying to decide how to spend their Friday evening. John prefers to stay in and play videogames. Mary prefers to go to a movie. But they both prefer to be together rather than apart. You can see the problem. Any way they choose, one or both will be unhappy. If they play videogames, John will be happy, but Mary will be bored. If they go to a movie, Mary will be happy, but John will be settling for his second choice. If they go their separate ways, both will be unhappy.

This is a much harder decision to make than it seems at first blush because each decision maker is not the only one choosing, and the outcome for each depends on what the other does, and they both know that. Let's follow Mary and John a bit more.

It's now Monday afternoon, and Mary is trying to avoid an annoying co-worker who keeps asking her out on a date even though he knows she's married. There are only two places to eat near her workplace, Subway Sandwich Shop and Starbucks. If she goes to Subway, and the co-worker goes there as well, she won't be able to avoid him. She will be miserable, but he will be delighted. The same thing will happen if they both end up at Starbucks. But if she goes to Subway, and he goes to Starbucks, she will be relieved, and he will be frustrated. Same thing if she goes to Starbucks, and he goes to Subway. So, once again, the outcome for each party depends on what the other person does, and they both know that.

Meanwhile, John is facing a dilemma of his own at work. He and a co-worker jointly botched a report in a major way, and it ended up costing the company they work for $100,000. Their boss is in a rage and plans to make the person responsible repay the company out of his own pocket. He meets with each man separately and demands to know who botched the report. If they blame each other, he will fine each of them $50,000. If only one blames the other, the person blamed will be fined $100,000, and the other will get off scot-free. If they both refuse to blame the other, then the boss will fine them each $25,000 and write off the remaining $50,000. John has to decide whether to blame his co-worker or to keep mum. His co-worker is facing the same dilemma, and they both know it. So this is a matter of trust, and what happens to both depends on what the other does.

These are the kind of choices we face frequently in life. To a mathematician, these kinds of problems are called *games*, and the optimal choices associated with them can be determined by *game theory*.

The Basics of Game Theory

Oskar Morgenstern and John von Neumann formulated the basic concepts behind game theory in their 1944 book *Theory of Games and Economic Behavior.* First, certain assumptions have to be made that we've already encountered when we learned about Bayesian decision making: Agents have preferences that can be ordered in terms of utility (satisfaction), and they act logically according to those preferences.

A game is a decision-making situation involving more than one player. Each player is trying to maximize his or her payoffs, but each player's actual payoff depends on what the other players do. Games are defined in terms of the set of participants playing, the possible courses of action available to each agent, and the set of all possible payoffs. In *constant-sum games*, the total payoff (sum of what everyone can get) is the same for all possible outcomes. Think of TV networks competing for viewership. If there are ten million viewers, and three million of them are watching NBC, that means the other networks are down three million viewers. If two million of them switch to ABC, ABC gains two million viewers, and NBC loses two million viewers. One player's gain is another's loss, and the sum of the payoffs is the same regardless of who wins viewership and who loses viewership. In a *zero-sum game* (a special type of constant-sum game), payoffs sum to zero. If I win \$1, you lose \$1. So the payoffs are plus one for me and minus one for you, and the sum of the payoffs is zero. In a *non-zero-sum game*, the sum of all payoffs could be negative or positive: Everyone could suffer, or everyone could benefit, but the sum of the suffering or benefit across all players is the same for all possible outcomes. For example, it could be that no matter how this game is played, the sum of all payoffs will be \$50, and everyone will win something. That means that if it's just you and me, and I win \$30, then you will win \$20. Or it could be that no matter how this game is played, the sum of all payoffs will be minus \$50, meaning that if it's just you and me, and I lose \$30, then you will lose \$20.

Games can be cooperative or non-cooperative. In *cooperative games,* players can form coalitions or alliances in order to maximize expected utility. Think of the difference between singles and doubles in tennis. Singles tennis is a non-cooperative game – the players play as individuals and vie to win the match. In doubles, the players play as teams each consisting of two players. The players on one team cooperate to beat the other team to win the match. Basketball, football, and soccer are all examples of cooperative games (which a friend of mine calls "coalitional ball-moving games"). Singles tennis, chess tournaments, and most videogames are *non-cooperative games*; a single individual vies to win the game against a human or computer opponent.

At each stage of the game, the players do something – they choose an action. There can be many outcomes to the game depending on the actions the players take. We can think of these actions as strategic. In a basketball game, players can play offensively or defensively. They can choose to execute a series of passes aimed at positioning the ball strategically. Some strategies lead to better outcomes for a given player than other actions. A player's *best response* is any strategy that yields the highest possible payoff. If you are a player or a coach, your best response is the strategy that is most likely to allow you to win the game.

When the game has reached a state of play in which no player can unilaterally improve the outcome of the game, the game is at *equilibrium*. Each player has adopted a strategy that cannot improve his outcome given the other players' strategies. For example, when one person or one team wins a tennis match, we say that the game has reached equilibrium. The winners can't do any better because they have won the match. The losers can't do any better because there are no more points to win or no more games to play in the match. There could also be a draw, as in a chess stalemate, when neither party can make a move that will improve his or her position. The game is over, but neither wins.

Contrast this with the situation described in the movie *A Beautiful Mind*: A group of guys enter a bar. They all see the sexiest woman in the bar, and they all want to go home with her. If they all compete for her, only one can win, all the other women will be offended and leave, and the rest of the men will go home lonely. But if the men switch strategies from pursuing the sexiest woman to pursuing other women, they increase their chances that they will all go home happy. In other words, the men can do better by switching strategies, and everyone knows that.

In 1950, John Nash formalized this idea for cooperative games. In *Nash equilibrium*, each player plays a best response and correctly anticipates that her partner will do the same. If each player has chosen a strategy, and no player can benefit by changing his or her strategy while the other players keep theirs unchanged, then the current set of strategy choices and the corresponding payoffs constitute a *pure-strategy Nash equilibrium*. To check whether there is a pure-strategy Nash equilibrium, all you have to do is check whether either player can do better by switching strategies.

Game Theory and the Battle of the Sexes

Let's return to the dilemmas faced by John and Mary. In the first one, John preferred videogames to movies, Mary preferred the opposite, and both preferred to be together rather than apart. This game is called the Battle of the Sexes, and it has a very interesting property: It has two pure-strategy Nash equilibria.

As we've described it, the Battle of the Sexes is a *simultaneous game* – that is, the players choose at the same time without knowing

Table 2.1. *Battle of the Sexes Game in Normal Form*

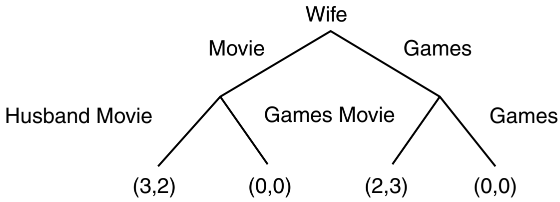
	Mary	
John	Movies	Videogames
Movies	(3,2)	(0,0)
Videogames	(0,0)	(2,3)

what the others have chosen. Simultaneous games are represented using matrices that describe each player's move and payoff (and information). This is called *normal form*, and it constitutes a description of the strategies available to each player along with their payoffs. Table 2.1 presents the Battle of the Sexes game John and Mary face, represented in normal form.

Mary's best choice is movies, and John's is videogames – and they both know this. What if they both adopt their best choices? If Mary adopts her best choice (movies), then John knows he should switch strategies and choose to go to the movies as well. If John adopts his best choice (videogames), then Mary knows she should switch strategies and choose to stay home and play videogames. So there are two pure-strategy Nash equilibria here: movie-movie, and videogames-videogames. How do you break this deadlock?

One way to do this is for John and Mary to take turns – that is, let John have his top choice this time, and then let Mary have her top choice next time, and so on. The Battle of the Sexes then becomes a sequential game. In a sequential game, players alternate moves, knowing what choices have already been made. Suppose John and Mary write down whether they went to a movie or played videogames each time so that they both know where they stand in the game. If every player observes the moves of every other player who has gone before her, the game is one of *perfect information*. Suppose instead that they don't write it down, and Mary's memory is much better for this sort of thing than John's. If some (but not all) players have information about prior moves, the game is one of *imperfect information*. Sequential games are represented using *game trees* showing each move and each possible response along with payoffs (and information). This kind

Extensive form for Sequential Battle of Sexes Game



The Battle of the Sexes as a sequential game, where each player alternates moves and knows which preceding moves were played.

FIGURE 2.1. The Battle of the Sexes in extensive form.

of representation is called *extensive form,* and it includes a complete description of the game, including the order of possible moves, payoffs, and information available to each player at each move. Figure 2.1 presents the Battle of the Sexes for John and Mary in extensive form.

What if John and Mary decide instead to break the deadlock by flipping a coin? If you introduce an element of chance into the game, it is called a *mixed-strategy equilibrium* game rather than a pure-strategy equilibrium game, and the best choice reduces to the probabilities associated with the element of chance introduced. Since Mary and John decided to flip a coin, for any given game, they both have a 50% chance that their preferred option will be chosen. Or they could play Rock, Paper, Scissors, adopting the preferred choice of the person who wins 2 out of 3 rounds. On each round, the chance of winning is 1 out of 3. If they decide to draw straws instead, with whoever draws the shortest straw winning, then the probability of getting one's choice is 1 out of the total number of straws.

Now here is the flash of brilliance of a beautiful mind: Nash proved that if there are a finite number of players and a finite number of strategies in a game, then there has to exist at least one Nash equilibrium, either pure strategy (choose a strategy and stick to it) or mixed strategy (introduce an element of chance). In 1994, the Nobel Prize in Economics was awarded to Nash, John Harsanyi, and Reinhard Selten for their work in game theory.

Table 2.2. *Matching Pennies Game in Normal Form*

Player 2	Player 1	
	Heads	Tails
Heads	(+1, −1)	(−1, +1)
Tails	(−1, +1)	(+1, −1)

Game Theory and Avoiding Mary's Inappropriate Co-Worker

Let's turn to the second problem described in the introduction where Mary was trying to avoid her inappropriate co-worker. This game is called *Matching Pennies*, and it also has a very interesting property: There are no pure-strategy Nash equilibria.

This game is called Matching Pennies because it has the same structure as the game you used to play as kids where you each have a penny in one hand and choose to put it on a table heads up or tails up. One of you wins if the coins match (two heads or two tails), and the other wins if they don't (one head and one tail). Let's say you win if they match, and your friend wins if they don't. If both of you play heads, then you will always win, and your friend will always lose. So your friend has an incentive to switch to playing tails. But that means that now you will always lose, so you have an incentive to switch to playing tails. That means your friend will now always lose, so your friend has an incentive to switch, and around and around you go. There is no pure-strategy Nash equilibrium for this game. Table 2.2 presents the game in normal form.

So what do you do? Probably what every kid who has ever played this game does: randomly switch between playing heads and tails. This gives each of you a 50% chance of winning.

As Nash proved, when you introduce an element of chance in a game that has no pure-strategy equilibrium, there is a mixed-strategy equilibrium. Now both players have a 50% chance of winning, so they might as well stick with what they're doing. So Mary might as well flip a coin to choose whether to go to Subway or Starbucks.

Game Theory and Deciding Whether or Not to Trust

Let's turn to the third situation: Should John cooperate with his co-worker and keep mum, or should he defect and rat him out? This game is called Prisoner's Dilemma because it has the kind of structure used to settle a case when there is not enough evidence to convict either party in a jointly committed crime. If you can get one of them to give evidence against the other, you can convict. The district attorney will try to make a deal, such as immunity, in exchange for evidence against the other guy. This game also has a very interesting property: *dominant-strategy equilibrium.*

In Prisoner's Dilemma, each player has a dominant strategy – that is, each party's best response does not depend on the strategies of the other players. No matter what the other player does, there is one plan that works best for each. If both rivals have dominant strategies that coincide, then the equilibrium is called a *dominant-strategy equilibrium*, a special case of a Nash equilibrium. In one-shot Prisoner's Dilemma, there are only two possible responses (cooperate or defect), and hence only two strategies (cooperate or defect). We can represent John and his co-worker's dilemma in normal form, as shown in Table 2.3.

Look at the top line. If John and his co-worker both rat each other out, they each have to pay $50,000. If John blames his co-worker, but his co-worker keeps mum, John pays nothing, and the co-worker pays the whole $100,000 himself. Now look at the bottom line. If instead John keeps mum, but his co-worker rats him out, then John will pay the whole $100,000 himself. If they both keep mum, then each will be responsible for only $25,000.

Both John and his co-worker are aware of the situation, and they have no control over what the other does. They each assume the other will do what gives him the best payoff. So the dominant strategy for each player in one-shot Prisoner's Dilemma is the same: *defect*.

Perhaps you think if they made a pact beforehand to keep mum, then the best choice (or the fairest choice) for John would be to just keep mum. But when it comes down to the wire, John has no idea whether his co-worker will honor that agreement or give into temptation. His best choice still remains the one that allows him to protect

Table 2.3. *Prisoner's Dilemma Game in Normal Form*

John	Co-worker	
	Defect	Cooperate
Defect	P = (–$50K, –$50K) Punishment for mutual defection	T= (0, –$100K) Temptation to defect
Cooperate	S = (–$100K, 0) Sucker's payoff for cooperating with a defector	R = (–$25K, –$25K) Reward for mutual cooperation

himself. The key point of game theory is that everyone knows the best strategy for each of the players, and everyone assumes the other players will play their best strategy.

What if John and his co-worker were good friends with a prior history of cooperation? Or if John and his co-worker are rivals with a prior history of intense competition? Does this change the nature of the game? In particular, does it change the strategy John should adopt – cooperate rather than defect – or does it not make any difference?

As it turns out, it matters a lot. If John has a history of Prisoner's Dilemma experiences with his co-worker, how he should behave depends on how his co-worker has behaved in the past. When Prisoner's Dilemma is played repeatedly, and everyone can keep track of previous decisions, it can be represented in extensive form shown in Figure 2.2.

In 1980, Robert Axelrod, a professor of political science at the University of Michigan, held a Prisoner's Dilemma tournament in which computer programs played games against one another and themselves repeatedly. The program that yielded the best outcome would be the winner of the tournament. Each program defined a strategy that specified whether to cooperate or defect based on the previous moves of both the strategy and the opponent. Some of the strategies entered in the tournament were:

- *Always defect*: This strategy defects on every turn. This is what game theory advocates. It is the safest strategy since it cannot be taken advantage of. However, it misses the chance to gain larger payoffs by cooperating with an opponent who is ready to cooperate.

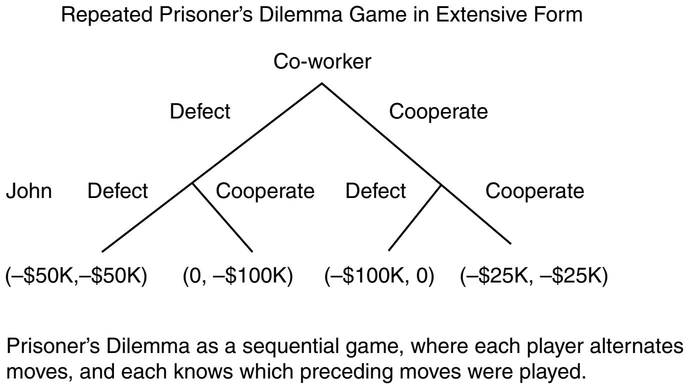


FIGURE 2.2. Prisoner's Dilemma game in extensive form.

- *Always cooperate*: This strategy does very well when matched against itself. However, if the opponent chooses to defect, then this strategy will do poorly.
- *Random*: The strategy cooperates 50% of the time.
- *Tit-for-tat*: This strategy cooperates on the first move, and then does whatever its opponent has done on the previous move.

The first three strategies were prescribed in advance so they could not take advantage of knowing the opponent's previous moves and figuring out its strategy. That is, they didn't learn anything about their opponents as the tournament progressed. But the fourth, tit-for-tat, modified its behavior by looking back at the previous game – and only the previous game.

The results of the tournament were published in 1981 (Alexrod & Hamilton, 1981), and the winner was tit-for-tat. This strategy captures the full benefits of cooperation when matched against a friendly opponent, but does not risk being taken advantage of when matched against an opponent who defects. Note that tit-for-tat is a smart heuristic like those studied by Gigerenzer and his colleagues; it ignores part of the available information, involves practically no computation, and yet is highly successful. (See Chapter 3 for a discussion of Gigerenzer's smart heuristics.)

There were some other interesting results. When matched against itself, the tit-for-tat strategy always cooperates. That is because it always cooperates in the first game, so the outcome of the first game is cooperate-cooperate. When paired against a chronic defector, tit-for-tat cooperates on the first move and then defects forever after. When paired with a mindless strategy like random, tit-for-tat sinks to its opponent's level. For that reason, tit-for-tat can't be called a "best" strategy, but it was powerful enough to win the tournament.

So what should John do? If his prior history indicates that his co-worker is a chronic cooperator, he should cooperate. If instead, his co-worker is a chronic defector, he should defect. If his co-worker is unpredictable, he should defect. So the take-home message here is a player should try to figure out (or guess) his opponent's strategy and then pick a strategy that is best suited for the situation.

Experimental Economics: What Do People Actually Do?

So what happens when you make people play Prisoner's Dilemma? In experimental-economics studies, people play against other humans, and real money changes hands. When they leave the experiment, they go home with the payoffs they accrued while playing the game.

Remember that, according to standard game-theoretic analysis, a rational agent always (a) acts according to self-interest, (b) defines self-interest in terms of maximum payoff to oneself, (c) plays dominant strategy in games that have one, and (d) seeks Nash equilibrium in games that have one.

As we just saw, the rational choice in one-shot Prisoner's Dilemma is to defect. If you do this, an economist would consider you rational because you played the strategy that is most likely to bring you maximum gains. So you may be surprised to find that people cooperate in one-shot Prisoner's Dilemma games about 50% of the time (Camerer, 2003). When asked why, they don't list altruism as a reason. Instead, they typically say that they expect others to cooperate as well. They might say something like, "She'd do the same for me." This is consistent with the principle of reciprocity ("I'll help you if you'll help me in return") rather than unilateral altruism ("Go ahead, you take all the money, I don't mind"). We can say, then, that people approach these games with a bias toward cooperation, an approach that economists find irrational (Kelley & Stahelski, 1970). In fact, people seem to approach these kinds of transactions with an implicit *norm* of cooperation.

Social norms are standards of behavior that are based on widely shared beliefs about how individual group members ought to behave in a given situation. They are usually expressed as conditional rules that lay out how a person is *permitted*, *obligated*, or *forbidden* to behave in a certain way in a specific situation. Usually, demand for a social norm arises when actions cause positive or negative side effects for other people. People comply with social norms voluntarily when the norm matches their goals (self-interest). They comply with social norms involuntarily when group interest conflicts with their self-interest, but others have coercive authority to enforce the norms or the opportunity to punish non-cooperators.

Consider how people behave in Public Goods experiments. Contributions are made to a common good (benefit) that all members can consume, even those who do not contribute anything. All well and good, except for the problem of free riding – taking benefits while contributing nothing yourself. In Public Goods situations, each member has an incentive to free ride on the contributions of others. This can be modeled as Prisoner's Dilemma where cooperation is conditional (cooperate if everyone else does), and defection is free riding. Just as

Box 2.1. How Public-Good Experiments Work

Game Components

- Groups consist of n individuals, where n is greater than 2.
- Each individual is given a monetary endowment E.
- Individuals decide how much of E they keep for themselves and how much they spend on a group project.
- The experimenter multiplies the total amount spent on the group project by a number, b, that is greater than 1 but smaller than n.

- The multiplied sum of the member's contribution constitutes the proceeds from the group project.
- These proceeds are then distributed equally among the n members.

Outcomes

If all members keep their endowments, they each earn E. If all contribute their endowments, the sum of contributions is nE,

- yielding an income of (b/n)nE = bE,
- which is greater than E for each group member.

Example

E = 20, b = 2, n = 4
Income: (b/n)nE = bE

If nobody contributes, each earns 20. If everybody contributes everything, each earns (2/4)4(20) = 2(20) = 40. If everybody contributes 5, each earns (2/4)4(5) = 10, plus 15 they kept for themselves = 25.

Prisoners' Dilemma is a special case of the Public Goods game with n = 2 and two available actions: contributing nothing (defect) or contributing everything (cooperate). Both players in Prisoner's Dilemma are better off if they defect (because b/n < 1) regardless of what the opponent does.

in experiments with Prisoner's Dilemma, the incidence of defection depends on whether or not defectors (free riders) are punished.

One such study was reported by Fehr and Gächter (2000). People were given tokens that could be exchanged for real money. In each trial of the experiment, they were given the opportunity to keep their money or contribute some or all of it to a group project. For each token they kept, they earned another token. For each token contributed to the project, everyone – even those who did not contribute – earned one-quarter of a token. At the end of the experiment, these tokens were converted into real money according to a publicly known exchange rate. A purely self-interested person would never contribute anything in this experiment. Contributions started at about 50% on the first trial. But free riding began to occur – people began earning tokens without contributing anything to the project. As play progressed, people contributed less and less, and, by the tenth trial, contributions had dried up entirely.

But then at the eleventh trial, people were given the opportunity to punish the free riders by imposing a penalty that had to be paid in tokens. Just as in two-person Prisoner's Dilemma experiments, the results were dramatic: Cooperation jumped back up to 60%, then steadily climbed to 100% by the twentieth trial. These results showed that people contribute less if free riding is tolerated, and, over time, contributions can cease entirely.

Just as in two-person Prisoner's Dilemma games, people in Public Goods experiments show a strong willingness to punish others in order to foster cooperation – even if it costs themselves something to do so. In a study by Fehr and Fischbacher (2004a), every time a free rider was punished, both the punisher and the punishee had to pay one token. It turned out that people were willing to pay up to two tokens to punish free riders. This really makes economists scratch their heads because it means the desire to punish norm violations is strong enough to overcome self-interest.

The importance of the threat of punishment also has a great impact on people's behavior in Trust Games. This type of game allows people to act like investment bankers, particularly those managing trust accounts. In one-shot Trust Games, the investor is given money and is invited to invest it with the trustee. They can transfer as much or as little as they want to the trustee, but the experimenter triples the amount transferred. This tripling simulates investment earning. So if you choose to give $1 to the trustee, the trustee actually gets $3. The trustee is free to return as much as he or she chooses, including nothing at all.

If trustees acted according to the principle of pure self-interest, they would keep all of the money. But they don't. Typically, investors transfer about half of their money to the trustee, and trustees return a little less than that to the investors. But what if investors were allowed to impose a penalty if the rate of return was less than a specified rate? Fehr and Rockenbach (2003) tested these conditions and found that trustees returned 50% more money if the rate allowed them to earn some money themselves, but they returned 67% less money if the rate cost them money. This is sometimes referred to as the *Norm of Just Punishment* (as in, "OK, I guess I deserved a little punishment for my greed, but that much punishment is unfair!") Fehr and Rockenbach interpreted their results to mean that when punishment is harsh or unfair, altruism declines.

So what we've learned so far is this: People's behavior in repeated games puzzles economists greatly because we reward cooperation more generously and punish defection more severely than is predicted by standard game-theoretic analyses (Weg & Smith, 1993). In fact, the opportunity to detect and punish cheaters in studies of multiple-trial Prisoner's Dilemma, Public Goods, and Trust Games was largely responsible for producing outcomes that deviate from standard game-theoretic predictions. So not only do we have a bias toward cooperating, we expect others to do so as well, and we will retaliate mightily if they don't.

Think about what that means for an economic theory based on self-interest: People will punish someone who defects in Prisoner's Dilemma, even though such a person is doing what game theory shows is the optimal strategy. Bystanders will even go so far as to pay a penalty so that they can have the opportunity to punish someone who defected in an observed Prisoner's Dilemma game, particularly if the person defected and their partner did not (Fehr & Fischbacher, 2004b). Plainly, people are motivated not only by economic self-interest in Prisoner's Dilemma games but also by norms of fairness and anticipated reciprocity.

Differences in Power and Status Influence How Fairly We Treat Others

Economists are even more puzzled by people's behavior in two other games, Dictator and Ultimatum. In one-shot Dictator, two strangers are given the opportunity to divide a sum of money. The catch is that the experimenter gives one of the parties (the dictator) total decision-making authority to divide the money any way he or she sees fit. The other party has no say in the matter. According to game theory, a rational dictator should keep all the money. Yet dictators typically offer the other party anywhere from 15% to 35% of the stake (Camerer, 2003). This is pure altruism; the game is one-shot, and these are strangers whose identities might be concealed from each other and even from the experimenter. Yet people will give up money they could just as easily take home without fear of repercussion.

In one-shot Ultimatum, the experimenter also assigns roles to two parties in order to divide a sum of money. But there is an important twist: The other party has a say in the matter. In Ultimatum, the person dividing the sum is the proposer who simply proposes how to split the money. The other party, the responder, can either accept or reject the offer. If the offer is accepted, the money is divided as proposed, and they both go home with money in their pockets. But if the responder rejects the money, the entire sum is forfeited; they both go home empty-handed.

If we define rationality as self-interest, then a rational proposer should offer slightly more than nothing, and a rational responder should take whatever is offered. After all, even one penny is better than no money at all, and both parties know that. But that's not what people do. Average offers in Ultimatum are a good deal higher than in Dictator – between 30% and 50% (Camerer, 2003). Offers less than 20% are typically rejected, meaning that people would rather no one get any money than accept an offer they believe to be too low. Again, people seem to approach these games with a *norm of self-interest* and *a norm of fairness* (Eckel & Grossman, 1995; Rabin, 1993).

Do people always operate according the principle of fairness meaning a fifty-fifty split? Not so much. Van Dijk and Vermunt (2000) had people play Dictator and Ultimatum games under conditions of symmetric (we both know everything we need to know) and asymmetric (one party knows something the other doesn't) information. The information manipulated was highly relevant, namely, that the dictators and proposers would be given double the value of each playing token, whereas their partners would receive only the stated value of the token. So if a token was marked "$1," dictators and proposers would receive $2 when they cashed in the token, whereas their partners would get $1. Dictators were unaffected by the information, making the same kinds of offers in both conditions. But proposers were very much affected; they made offers of nearly equal monetary value distribution when their partners knew about the true values of the tokens for each party, but they exploited their partner's ignorance in the asymmetric-information condition by making a seemingly fair offer to split the tokens in half. In reality, those seemingly equal distributions meant the proposer would end up with one-third more money. So dictators had more power and more information than their partners, and they behaved more fairly toward their partners. It was as though having that much of an advantage over their partners triggered the norm of fairness. But when proposers in the Ultimatum game had an informational advantage over their partners, they behaved selfishly. Or, as an economist would put it, when proposers had the advantage of asymmetrical information, they behaved strategically, like brokers who engage in insider trading or Enron executives who encouraged their employees to hold onto their stock options while they sold their own, knowing the company was in trouble and the stocks would be worthless very soon.

Using a different methodological approach, Fiddick and Cummins (2007) found even more intriguing results. People were asked to evaluate a carpooling arrangement in which one party agrees to pay for gasoline if the other party does all the driving. They were shown hypothetical ledgers showing gas payments that indicated varying degrees of compliance on the part of the gas-paying partner (from 100% compliance to as little as 25%). They were asked how willing they would be to continue the arrangement at each level of compliance and how fairly they thought the other person was treating them. The twist was that, in some scenarios, the two parties had equal status and power (both were employees), and in some they were of unequal status and power (one party was the other one's boss). People were far more tolerant of the employee cheating the boss than they were of the boss cheating the employee. This was true even when the employee was described as making a lot more money than the boss because the employee had a home-based computer business on the side. But one crucial factor had to be present for these asymmetries in tolerance and perceived fairness to occur: The employee had to work for that boss. If the parties were described as a boss and an employee from different companies who met through a classified ad, the effect disappeared; equivalent levels of intolerance for cheating were found regardless of employment status. So it seems it is asymmetries in the social relationship, and not asymmetries in costs and benefits, that underlie the effect.

Fiddick and Cummins referred to these results as the *noblesse oblige* effect. *Noblesse oblige* is a French term that can be roughly interpreted as "status and power entail obligation" or "with wealth, power, and prestige come responsibilities toward those less fortunate." In ethics, the term is sometimes used to describe a moral economy wherein privilege is balanced by duty toward those who lack such privilege. The generous behavior of the dictators in van Dijk and Vermunt's (2000) study may also be described this way. They were in total control of both assets and vital information, yet they did not take advantage of their partners. They behaved as though their advantages imbued them with pastoral responsibility toward their partners.

Contrast this with what happens when people size each other up purely on the basis of competitive performance. In a series of studies, Hoffman and colleagues (Hoffman, McCabe, Shachat, & Smith, 1994; Hoffman, McCabe, & Smith, 1996; Hoffman & Spitzer, 1985) manipulated relative standings in a competitive pre-game to see if these relative standings would affect the way people behaved in Dictator and Ultimatum games. Participants were required to complete a current-events test and were then ranked according to the number of correct answers they had. These rankings were posted where everyone could see the results. Then the experimenters paired people so that the pairs consisted of one high-ranking and one lower-ranking individual. The top-ranking person was paired with the person who ranked tenth, the second-highest ranked person was paired with the person who ranked eleventh, the third-highest was paired with the person ranked twelfth, and so on. The higher-ranking person in each pair was assigned the role of dictator in the Dictator game or the role of proposer in the Ultimatum game.

The results were striking. Dictators made significantly greater uneven distributions (favoring themselves) compared to a control condition where no pre-game was played. Proposers made significantly lower offers without raising the rejection rate (again, compared to a control condition). In other words, higher-ranking individuals thought they were entitled to more, and lower-ranking individuals thought they were entitled to less.

Neuroscience Makes It Clearer Why We Behave the Way We Do

So let's take stock: The results of experimental-economics studies show that decision makers are generally less selfish and less strategic than game theory predicts, and they value social factors such as reciprocity, fairness, and relative social status more than the theory predicts. Is this just because we keep making mistakes? Are we trying to behave in ways that an economist defines as rational, but we keep falling short of the mark because of human fallibility? Or are we wired this way? So far, the results of neuroscience studies suggests that the last of these is actually the case: We are wired this way.

As we will see in more detail in the next chapter the front part of the brain tracks probability information, whereas activity in deeper parts of the brain tracks the magnitude of reward or punishment. The reward or punishment could be monetary, or it could be social in nature. So what happens when we have people play Prisoner's Dilemma while undergoing fMRI imaging of their brains? Bottom line: Even when the same amount of money is gained or lost, reciprocated cooperation with another human leads to increased activation in the striatum (reward area) whereas unreciprocated cooperation shows a corresponding decrease in activation in this area (Sanfey, 2007). These results indicate that people find cooperation rewarding and lack of cooperation distressing in Prisoner's Dilemma games. Rilling and colleagues (2002) summarized it this way: Activation of the brain's reward circuitry positively reinforces cooperation, thereby motivating subjects to resist the temptation to selfishly defect.

In related work, neural reward circuitry was found to become active when people donated money to charity and when they observed money being donated to charity. But this was true only if the donations were voluntary; if the donations were made because they were required by someone's job or other involuntary means, these reward circuits did not light up. This means that we find it rewarding to behave altruistically, not just cooperatively.

We know that people will go out of their way to punish defectors in Prisoner's Dilemma, even when they are only observing the game rather than playing it themselves, and even when it costs them money to do so. It turns out that when players are given the option to punish defectors, activation in the brain's reward circuitry occurs, even when the person is losing money to punish the other party. This means that people find it rewarding to punish defectors. Given that Public Goods games have the same kind of mathematical and social structure as Prisoner's Dilemma, it should come as no surprise that these games yield the same neural-imaging results: Reward-related brain areas were activated when free riders were punished (de Quervain et al., 2004).

In the Trustee game, activity in reward pathways of the brain was greatest when the investor repaid generosity with generosity and most subdued when the investor repaid generosity with stinginess (King-Casas and colleagues, 2005). Even more surprisingly, the amount of money investors were willing to fork over could be manipulated chemically using a substance called oxytocin. Oxytocin is sometimes referred to as the "bonding" hormone; it seems to facilitate social bonding and trust. It is a hormone secreted from the posterior lobe of the pituitary gland. It initiates labor in pregnant women and facilitates production of breast milk. This means that both baby and mother are flooded with a feel-good social-bonding hormone during birth and nursing. Oxytocin levels also rise in men and women during sex, again facilitating emotional bonding. In a study by Kosfeld and colleagues (2005), investors and trustees were given oxytocin or a placebo before playing the Trustee game. Oxytocin increased the willingness of the investors to trust – they forked over more money. But it had no impact on the trustees. They behaved the same as trustees always do in these studies, returning slightly less than was invested. So oxytocin made people more willing to trust in this study but not more generous. Similar results were found in the Ultimatum game: Intranasal oxytocin increased generosity by 80% but had no effect in the Dictator game (Zak, Stanton & Ahmadi, 2007), again indicating that oxytocin may make us more trusting but not necessarily more generous. And how do we feel about stingy and generous offers? Stingy offers in Ultimatum games activate brain areas (anterior insula) associated with feelings of disgust (Sanfey and colleagues, 2003). But this is true only if a person is playing with another person; if the other party is a computer, no changes occur in those brain areas.

The results of decision-neuroscience studies like these plainly show that the impact of social aspects of these games cannot be overestimated. People behave as though they are wired to expect long-term reciprocal relationships. Reward circuits become active when we behave cooperatively and generously, and disgust circuits become active when we behave selfishly. The outcome of all this wiring seems to be an attempt to achieve the following social goals: Increase the likelihood that inequity is avoided, foster mutual reciprocity, and encourage punishment of those seeking to take advantage of others. In repeated games – repeated transactions among individuals who will remember one another and their transaction history – reputation becomes exceedingly important. No one wants to engage in transactions with individuals who have a reputation for stinginess and selfishness – not even those who may be planning to behave that way themselves. Whether you're selfish or generous, transacting with a cooperator is always the better bet.

In fact, this point is obvious even to infants. In a set of studies (Hamlin and colleagues, 2007), six-month-old infants watched as a red disc struggled to move up a steep incline. In one condition, a yellow triangle came racing along and pushed the red disc to the top of the incline. In another, a blue square came racing along and pushed the red disc down to the bottom of the incline. In other conditions, a third object either did nothing or was inanimate and could not move. After viewing the show, the infants were shown the three objects and allowed to select which one they wanted to play with. The infants overwhelmingly preferred helpers (cooperators) over neutral parties and neutral parties over those who hindered. The authors concluded that even preverbal infants assess individuals on the basis of their behavior toward others and, moreover, that this kind of social evaluation is a biological adaptation.