ONE

# *Introduction*

After two decades of teaching bright and curious university students, I came to a disturbing conclusion: Despite our best efforts to expose students to the ideas and insights that profoundly shape the way we think and live, most students were still pretty insulated within their particular disciplines. The science majors knew all about hypothesis testing but didn't know the first thing about moral theory. The philosophy and pre-law majors knew all about argumentation but didn't know the first thing about scientific investigation. Outside of the business school, precious few students knew anything about decision theories that drive the equity market and underlie economic policies that impact their lives – right down to whether or not they can get student loans. And outside of the psychology majors, virtually none knew that the way the brain is wired shapes the way we think, act, and feel. And then these bright and well-educated people take jobs as policy-makers, writers, scientists, lawyers, and teachers – bumping about in life with holes where some crucial bits of knowledge ought to be.

Does this really matter? Well, consider a Colorado DUI case that ended in acquittal in spite of overwhelming evidence. "It made no sense," the prosecutor complained. "It was an open-and-shut case. The guy's blood alcohol level was over the limit, he couldn't walk a straight line, and there were open beer cans in the car with his fingerprints on them." So why was the defendant acquitted? "I talked to one of the jury members after the verdict," the attorney reported, "and he said there was an astrologer among them. She cast a chart and argued that, according to the chart, the defendant couldn't have been driving drunk that day. So they couldn't get a majority past reasonable doubt," he sighed.

Do astrological charts constitute reasonable doubt? For most of us, the answer would be "no." But explaining exactly *why* we believe this – and why this jury's decision is troubling – is a challenge. We simply know that most of us would never get on a plane, drive across a bridge, obey a law, or concern ourselves with presidential elections if they were based on this type of reasoning and evidence. Instead, we readily do these things because we assume that planes, bridges, laws, and our system of government are the outcomes of a painstaking process of reasoning, evidence evaluation, and learning from past mistakes. We believe that reason is the steel thread that makes the fabric of our jurisprudence fair, our science accurate, and our social institutions able to withstand change. In short, we hold a core belief that action should be governed by reason.

Coextensive is the core belief that decisions we make while in the throes of heated emotion are likely to be bad ones, and those we make in the cold and clear light of reason will be better. This just seems self-evident. This Wikipedia entry succinctly captures our folk wisdom: "Reason is a way of thinking characterized by logic, analysis, and synthesis. It is often contrasted with emotionalism, which is thinking driven by desire, passion or prejudice. Reason attempts to discover what is true or what is best."

So firmly entrenched are these beliefs that it often comes as a surprise to us to learn that not everyone thinks this way. For example, in *The Suicide of Reason: Radical Islam's Threat to the West* (2007), Lee Harris argues that

> [T]he West has cultivated an ethos of individualism, reason and tolerance, and an elaborate system in which every actor, from the individual to the nation-state, seeks to resolve conflict through words. The entire system is built on the idea of self-interest … Our worship of reason is making us easy prey for a ruthless, unscrupulous and extremely aggressive predator and may be contributing to a slow cultural "suicide."

To thinkers like Harris, reason is what makes us weak, indecisive, and vulnerable. Reason is what ensnarls us in words and makes us slow to act. And there is sufficient evidence that human reasoning is frail and fallible. Scientists who study human reasoning and decision making have documented the alarming frequency with which we are prone to error.

The fallibility of human reasoning was not lost on our founding fathers, nor is it lost on scientists and policy-makers who still depend upon it to make decisions that impact millions of lives. As Ayaan Hirsi Ali puts it

> Enlightenment thinkers, preoccupied with both individual freedom and secular and limited government, argued that human reason is fallible. They understood that reason is more than just rational thought; it is also a process of trial and error, the ability to learn from past mistakes. The Enlightenment cannot be fully appreciated without a strong awareness of just how frail human reason is. That is why concepts like doubt and reflection are central to any form of decision-making based on reason. ("Blind Faiths," *New York Times* 1/6/08)

So how have these all-important modes of "doubt and reflection" been incorporated in our decision making? There are models of reason that dominate western thought. These are the "jewels in the crown" of our method of inquiry. Or, to borrow a term from philosopher Robert Cummins, they constitute our knowledge bridges – reasoning that takes us from what we already know to what we want to know.

The purpose of this book is to lay out each of these "knowledge bridges" in plain English so that educated readers can decide for themselves just how much or how little confidence we should have in our "worship of reason." These methods are

1. Rational choice: Choose what is most likely to give you what you want
2. Game theory: What to do when you're not the only one making choices
3. Moral judgment: How we tell the difference between right and wrong
4. Scientific reasoning, which consists of
   - Hypothesis testing: The search for truth by evaluating evidence
   - Causal reasoning: Explaining, predicting, and preventing events
5. Logic: The search for truth through argumentation

These are the main methods of inquiry and decision making that underlie the decisions we make in our everyday lives, in jurisprudence, in politics, in economics, and in science. Before we can decide whether reason can indeed be trusted, we need to understand the tools of the trade – how the "game of reasoned thought" – is played by the best.

In addition, there are two other modes of reasoning that merit discussion. Although not as well formalized as the previous four, they are ubiquitous in human and non-human cognition.

6. Problem solving: The search for solutions to unwanted situations
7. Analogical reasoning: The heart and soul of insight, discovery, and genius

One last thought must be kept in mind: However rational and flawless these methods may seem, they are not implemented on infallible hardware. Instead, these models are implemented by flesh-and-blood human reasoners, or more specifically, by their neural circuitry. To fully appreciate the whole package of reason, we must be conversant with the way such circuits operate in different circumstances to yield decisions. For this reason, this book will detail important findings from the new fields of decision neuroscience that are pertinent to each of these models of thought.

After reading this book, readers should be empowered to decide for themselves whether human reasoning is as frail or as strong, as dangerous or as benign, or as superfluous or as crucial as it has been made out to be.

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

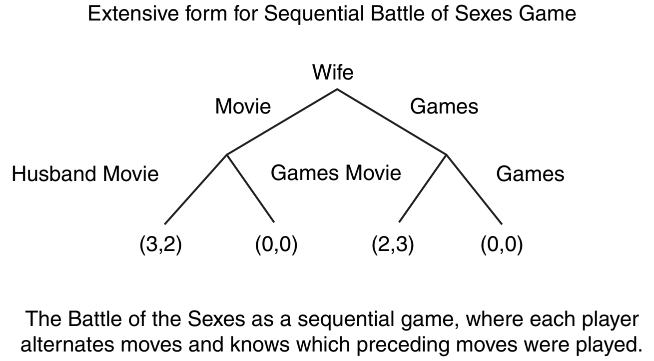


FIGURE 2.1. The Battle of the Sexes in extensive form.

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.