1 Introduction

On September 6, 1492, Christopher Columbus set off from the Canary Islands and sailed westward in an attempt to find a new trade route between Europe and the Far East. On October 12, after five weeks of sailing across the Atlantic, land was sighted. Columbus had never been to the Far East, so when he landed in Central America ("the West Indies") he believed that he had indeed discovered a new route to the Far East. Not until twenty-nine years later did Magellan finally discover the westward route to the Far East by sailing south around South America.

Columbus' decision to sail west from the Canary Islands was arguably one of the bravest decisions ever made by an explorer. But was it rational? Unlike some of his contemporaries, Columbus believed that the earth is a rather small sphere. Based on his geographical assumptions, he estimated the distance from Europe to East India to total 2,300 miles. The actual distance is about 12,200 miles, which is more than five times farther than Columbus thought. In the fifteenth century, no ship would have been able to carry provisions for such a long journey. Had America not existed, or had the earth been flat, Columbus would certainly have faced a painful death. Was it really worth risking everything for the sake of finding a new trade route?

This book is about decision theory. Decision theory is the theory of rational decision making. Columbus' decision to set off westwards across an unknown ocean serves as a fascinating illustration of what decision theory is all about. A *decision maker*, in this case Columbus, chooses an act from *a set of alternatives*, such as sailing westwards or staying at home. The *outcome* depends on the true *state of the world*, which in many cases is only partially known to the decision maker. For example, had the earth been a modest-sized sphere mostly covered by land and a relatively small and navigable sea, Columbus' decision to sail westwards would have made

him rich and famous, because the King and Queen of Spain had promised him 10% of all revenue gained from a new trade route. However, Columbus' geographical hypothesis turned out to be false. Although spherical, the earth is much bigger than Columbus assumed, and Europe is separated from the Far East by the huge American continents. Thus, in the fifteenth century the westward route was not a viable option for Europeans wishing to trade with the Far East. All this was unknown to Columbus. Despite this, the actual outcome of Columbus' decision was surprisingly good. When he returned to Spain he gained instant fame (though no financial reward). Another possible outcome would have been to never reach land again. Indeed, a terrible way to die!

The decision problem faced by Columbus on the Canary Islands in September 1492 can be summarized in the *decision matrix* shown in Table 1.1. Note that the outcome of staying at home would have been the same no matter whether his geographical hypothesis was true or not.

Since the second hypothesis turned out to be the true one, the actual outcome of sailing westwards was that Columbus got famous but not rich. However, it should be evident that the rationality of his decision depended on *all possible* outcomes – every entry in the matrix matters. But how should one use this basic insight for formulating more precise and useful theories about rational decision making? In this book we will consider a number of influential attempts to answer this question.

Roughly put, the ultimate aim of decision theory is to formulate hypotheses about rational decision making that are as accurate and precise as possible. If you wish to tell whether Columbus' decision to sail westwards was rational, or whether this is the right time to invest in the stock market, or whether the benefit of exceeding the speed limit outweighs the risk of getting caught, then this is the right subject for you. You think it is not worth the effort? Well, that is *also* a decision. So if you want to find out

Table 1.1

	Geographical hypothesis true	There is some other land westwards	There is no land westwards
Sail westwards	Rich and famous	Famous but not rich	Dead
Do not	Status quo	Status quo	Status quo

whether the decision not to learn decision theory is rational, you must continue reading this book. Don't stop now!

1.1 Normative and Descriptive Decision Theory

Decision theory is an interdisciplinary project to which philosophers, economists, psychologists, computer scientists and statisticians contribute their expertise. However, decision theorists from all disciplines share a number of basic concepts and distinctions. To start with, everyone agrees that it makes sense to distinguish between *descriptive* and *normative* decision theory. Descriptive decision theories seek to explain and predict how people *actually* make decisions. This is an empirical discipline, stemming from experimental psychology. Normative theories seek to yield prescriptions about what decision makers are *rationally required* – or *ought* – to do. Descriptive and normative decision theory are, thus, two separate fields of inquiry, which may be studied independently of each other. For example, from a normative point of view it seems interesting to question whether people visiting casinos in Las Vegas *ought* to gamble as much as they do. In addition, no matter whether this behavior is rational or not, it seems worthwhile to *explain* why people gamble (even though they know they will almost certainly lose money in the long run).

The focus of this book is normative decision theory. There are two reasons for this. First, normative decision theory is of significant philosophical interest. Anyone wishing to know what makes a rational decision rational should study normative decision theory. How people actually behave is likely to change over time and across cultures, but a sufficiently general normative theory can be expected to withstand time and cultural differences.

The second reason for focusing on normative decision theory is a pragmatic one. A reasonable point of departure when formulating descriptive hypotheses is that people behave rationally, at least most of the time. It would be difficult to reconcile the thought that most people most of the time make irrational decisions with the observation that they are in fact alive and seem to lead fairly good lives – in general, most of us seem to do pretty well. Moreover, if we were to discover that people actually behave irrationally, either occasionally or frequently, we would not be able to advise them how to change their behavior unless we had some knowledge about

normative decision theory. It seems that normative decision theory is better dealt with *before* we develop descriptive hypotheses.

That said, normative and descriptive decision theory share some common ground. A joint point of departure is that decisions are somehow triggered by the decision maker's beliefs and desires. This idea stems from the work of the Scottish eighteenth-century philosopher David Hume. According to Hume, the best explanation of why Columbus set off westwards was that he *believed* it would be possible to reach the Far East by sailing in that direction, and that he *desired* to go there more than he desired to stay at home. Likewise, a possible explanation of why people bet in casinos is that they *believe* that the chance of winning large amounts is higher than it actually is and that they have a strong *desire* for money. In the twentieth century much work in descriptive decision theory was devoted to formulating mathematically precise hypotheses about how exactly beliefs and desires trigger choices. Unsurprisingly, a number of philosophers, economists and statisticians also proposed theories for how beliefs and desires *ought* to be aggregated into rational decisions.

1.2 Rational and Right Decisions

A decision can be rational without being right and right without being rational. This has been illustrated through many examples in history. For instance, in the battle of Narva (on the border between Russia and what we now call Estonia) on November 20, 1700, King Carl of Sweden and his 8,000 troops attacked the Russian army, led by Tsar Peter the Great. The tsar had about ten times as many troops at his disposal. Most historians agree that the Swedish attack was irrational, since it was almost certain to fail. Moreover, the Swedes had no strategic reason for attacking; they could not expect to gain very much from victory. However, because of an unexpected blizzard that blinded the Russian army, the Swedes won. The battle was over in less than two hours. The Swedes lost 667 men and the Russians approximately 15,000.

Looking back, the Swedes' decision to attack the Russian army was no doubt right, since the *actual outcome* turned out to be success. However, because the Swedes had no *good reason* for expecting that they were going to win, the decision was nevertheless irrational. Decision theorists are primarily concerned with rational decisions rather than right ones.

In many cases it seems impossible to foresee, even in principle, which act is right until the decision has already been made (and even then it might be impossible to know what *would* have happened had one decided differently). It seems much more reasonable to claim that it is always possible to foresee whether a decision is rational. This is because theories of rationality operate on information available at the point in time the decision is made rather than on information available at some later point in time.

More generally speaking, we say that a decision is *right* if and only if its actual outcome is at least as good as that of every other possible outcome. Furthermore, we say that a decision is *rational* if and only if the decision maker chooses to do what she has most reason to do at the point in time at which the decision is made. The kind of rationality we have in mind here is what philosophers call *instrumental* rationality. Instrumental rationality presupposes that the decision maker has some *aim*, such as becoming rich and famous, or helping as many starving refugees as possible. The aim is external to decision theory, and it is widely thought that an aim cannot in itself be irrational, although it is of course reasonable to think that *sets* of aims can sometimes be irrational, e.g. if they are mutually inconsistent. Now, on this view, to be instrumentally rational is to do whatever one has most reason to expect will fulfill one's aim. For instance, if your aim is not to get wet and it is raining heavily, you are rational in an instrumental sense if you bring an umbrella or raincoat when going for a walk.

The instrumental, means-to-end notion of rationality has been criticized, however. Philosopher John Rawls argues that an aim such as counting the number of blades of grass on a courthouse lawn is irrational, at least as long as doing so does not help to prevent terrible events elsewhere. Counting blades of grass on a courthouse lawn is not important enough to qualify as a rational aim. In response to this point it could perhaps be objected that everyone should be free to decide for herself what is important in life. If someone strongly desires to count blades of grass on courthouse lawns, just for the fun of it, that might very well qualify as a rational aim.

1.3 Risk, Ignorance and Uncertainty

In decision theory, everyday terms such as *risk, ignorance* and *uncertainty* are used as technical terms with precise meanings. In decisions under risk the decision maker knows the probability of the possible outcomes, whereas in

decisions under ignorance the probabilities are either unknown or nonexistent. Uncertainty is used either as a synonym for ignorance, or as a broader term referring to both risk and ignorance.

Although decisions under ignorance are based on less information than decisions under risk, it does not follow that decisions under ignorance must therefore be more difficult to make. In the 1960s, Dr. Christiaan Barnard in Cape Town experimented on animals to develop a method for transplanting hearts. In 1967 he offered 55-year-old Louis Washkansky the chance to become the first human to undergo a heart transplant. Mr. Washkansky was dying of severe heart disease and was in desperate need of a new heart. Dr. Barnard explained to Mr. Washkansky that no one had ever before attempted to transplant a heart from one human to another. It would therefore be meaningless to estimate the chance of success. All Dr. Barnard knew was that his surgical method seemed to work fairly well on animals. Naturally, because Mr. Washkansky knew he would not survive long without a new heart, he accepted Dr. Barnard's offer. The donor was a 25-year-old woman who had died in a car accident the same day. Mr. Washkansky's decision problem is illustrated in Table 1.2.

The operation was successful and Dr. Barnard's surgical method worked quite well. Unfortunately, Mr. Washkansky died 18 days later from pneumonia, so he did not gain as much as he might have hoped.

The decision made by Mr. Washkansky was a decision under ignorance. This is because it was virtually impossible for him (and Dr. Barnard) to assign meaningful probabilities to the possible outcomes. No one knew anything about the probability that the surgical method would work. However, it was nevertheless easy for Mr. Washkansky to decide what to do. Because no matter whether the new surgical method was going to work on humans or not, the outcome for Mr. Washkansky was certain to be at least as good as if he decided to reject the operation. He had nothing to lose. Decision theorists say that in a case like this the first alternative (to have the operation) *dominates* the second alternative. The concept of

Table 1.2

	Method works	Method fails
Operation	Live on for some time	Death
No operation	Death	Death

dominance is of fundamental importance in decision making under ignorance, and it will be discussed in more detail in Chapter 3.

Ever since Mr. Washkansky underwent Dr. Barnard's pioneering operation, thousands of patients all over the world have had their lives prolonged by heart transplants. The outcomes of nearly all of these operations have been carefully monitored. Interestingly enough, the decision to undergo a heart transplant is no longer a decision under ignorance. Increased medical knowledge has turned this kind of decision into a decision under risk. Recent statistics show that 71.2% of all patients who undergo a heart transplant survive on average 14.8 years, 13.9% survive for 3.9 years, and 7.8% for 2.1 years. However, 7.1% die shortly after the operation. To simplify the example, we will make the somewhat unrealistic assumption that the patient's life expectancy after a heart transplant is determined entirely by his genes. We will furthermore suppose that there are four types of genes.

Group I: People with this gene die on average 18 days after the operation (0.05 years).
Group II: People with this gene die on average 2.1 years after the operation.
Group III: People with this gene die on average 3.9 years after the operation.
Group IV: People with this gene die on average 14.8 years after the operation.

Because heart diseases can nowadays be diagnosed at a very early stage, and because there are several quite sophisticated drugs available, patients who decline transplantation can expect to survive for about 1.5 years. The decision problem faced by the patient is summarized in Table 1.3.

The most widely applied decision rule for making decisions under risk is the principle of maximizing expected value. As will be explained in some detail in Chapter 4, this principle holds that the total value of an act equals the sum of the values of its possible outcomes weighted by the probability

Table 1.3

	Group I: 7.1%	Group II: 7.8%	Group III: 13.9%	Group IV: 71.2%
Operation	0.05 years	2.1 years	3.9 years	14.8 years
No operation	1.5 years	1.5 years	1.5 years	1.5 years

for each outcome. Hence, the expected values of the two alternatives are as follows.

Operation: $(0.05 \cdot 0.071) + (2.1 \cdot 0.078) + (3.9 \cdot 0.139) + (14.8 \cdot 0.712) \approx 11$
No operation: $(1.5 \cdot 0.071) + (1.5 \cdot 0.078) + (1.5 \cdot 0.139) + (1.5 \cdot 0.712) = 1.5$

Clearly, if the principle of maximizing expected value is deemed to be acceptable, it follows that having an operation is more rational than not having one, because 11 is more than 1.5. Note that this is the case despite the fact that 7.1% of all patients die within just 18 days of the operation.

1.4 Social Choice Theory and Game Theory

The decisions exemplified so far are all decisions made by a *single* decision maker *not* taking into account what other decision makers are doing. Not all decisions are like this. Some decisions are made collectively by a group, and in many cases decision makers need to take into account what others are doing. This has given rise to two important subfields of decision theory: social choice theory and game theory.

Social choice theory seeks to establish principles for how decisions involving more than one decision maker ought to be made. For instance, in many countries (but unfortunately not all) political leaders are chosen by democratic election. Voting is one of several methods for making social choices. However, as will be explained in Chapter 13, the voting procedures currently used in many democratic countries are quite unsatisfactory from a theoretical perspective, because they fail to meet some very reasonable requirements that such procedures ought to fulfill. This indicates that the voting procedures we currently use may not be the best ones. By learning more about social choice theory we can eventually improve the way important decisions affecting all of us are made. Naturally, the group making a social choice need not always be the people of a nation; it could also be the members of, say, a golf club or a family. The basic theoretical problem is the same: How do we aggregate the divergent beliefs and desires of a heterogeneous set of individuals into a collective decision? In order to avoid misunderstanding, it is worth keeping in mind that collective entities, such as governments and corporations, sometimes act as single decision makers. That is, not

every act performed by a group is a social choice. For example, once the government has been elected its decisions are best conceived of as decisions taken by a single decision maker.

Game theory is another, equally important subfield of decision theory. You are probably familiar with games such as chess and Monopoly, where the outcome of your decision depends on what others do. Many other decisions we make have the same basic structure. If your opponents are clever enough to foresee what you are likely to do, they can adjust their strategies accordingly. If you are rational, you will of course also adjust your strategy based on what you believe about your opponent. Here is an example, originally discussed by Jean-Jacques Rousseau: Two hunters can either cooperate to hunt stag (which is a rather large animal that cannot be caught by a single hunter) or individually hunt for hares. A hare is rather small and can easily be caught by a single hunter. If the hunters cooperate and hunt stag, each of them will get 25 lbs of meat; this is the best outcome for both hunters. The worst outcome for each hunter is to hunt stag when the other is hunting hare, because then he will get nothing. If the hunter decides to hunt hare he can expect to get a hare of 5 lbs. In Table 1.4 the numbers in each box refer to the amount of meat caught by the first and second hunter, respectively.

This game has become known as stag hunt. In order to analyze it, imagine that you are Hunter 1. Whether it would be better to hunt stag or hare depends on what you believe the other hunter will do. Note, however, that this also holds true for the other hunter. Whether it would be better for him to hunt stag or hare depends on what he believes you are going to do. If both of you were fully confident that the other would cooperate, then both of you would benefit from hunting stag. However, if only one hunter chooses to hunt stag and the other does not cooperate, the hunter will end up with nothing. If you were to hunt hare you would not have to worry about this risk. The payoff of hunting hare does not depend on what the other hunter chooses to do.

Table 1.4

		Hunter 2	
		stag	hare
Hunter 1	stag	25 lbs, 25 lbs	0 lbs, 5 lbs
	hare	5 lbs, 0 lbs	5 lbs, 5 lbs

The same point applies to the other hunter. If he suspects that you may not be willing to cooperate it is safer to hunt hare. Rational hunters therefore have to make a trade-off between two conflicting aims, viz. mutual benefit and risk minimization. Each hunter is pulled toward stag hunting by considerations of mutual benefit, and toward hare hunting by considerations of risk minimization. What should we expect two rational players to do when playing this game?

Many phenomena in society have a similar structure to the stag hunting scenario. In most cases we are all better off if we cooperate and help each other, but such cooperation can only occur if we trust our fellow citizens. Unfortunately, we sometimes have little or no reason to trust our fellow citizens. In such cases it is very likely that we will end up with outcomes that are bad for everyone. That said, there are of course also cases in which we tend to trust each other, even though the game has exactly the same structure as stag hunt. For instance, David Hume (1739: III) observed that, "Two men who pull at the oars of a boat, do it by an agreement or convention, tho' they have never given promises to each other." Arguably, the best outcome for both rowers is to cooperate, whereas the worst outcome is to row alone while the other is relaxing. Hence, from a game-theoretical point of view, stag hunting is similar to rowing. Why is it, then, that most people tend to cooperate when rowing but not when hunting stag? In Chapter 12 it will be explained that the answer has to do with the number of times the game is repeated.

Before closing this section, a note about terminology is called for. I – and many others – use the term *decision theory* both as a general term referring to all kinds of theoretical inquiries into the nature of decision making, including social choice theory and game theory, as well as a more narrow term referring only to individual decisions made by a single individual not considering the behavior of others. Whether the term is used in the general or narrow sense is determined by context.

1.5 A Very Brief History of Decision Theory

The history of decision theory can be divided into three distinct phases: the Old period, the Pioneering period and the Axiomatic period. As is the case for nearly all academic disciplines, the Old period begins in ancient Greece. However, the Greeks did not develop a *theory* of rational decision making.