PART I

INTUITIVE DEFINITIONS

SCOPE OF THIS BOOK

This book is about decision theory under uncertainty, namely, asking how do, and how should, people make decisions in situations of uncertainty. You probably already know the standard answer in economics, namely, that people do, and should, maximize expected utility. I don't think that this is a bad answer if we know what "utility" and "probability" mean.

The difficulty is that we often don't. Both concepts are problematic. We will discuss the notion of utility at various points, but these notes basically revolve around the meaning of probability. We ask, what is meant by saying "the probability of event A is p"? There are obviously many other ways to organize the material presented here, and, in particular, one can start with the meaning of utility. The present organization is but one way to relate the various topics to each other.

Naturally, the main question around which this book is organized has implications regarding its scope and content. There are important and interesting recent developments in decision under uncertainty that are not discussed here. In particular, there is relatively little discussion of nonexpected utility theory under risk, namely, with given probabilities, and almost no reference to models of choices from menus and to various models of behavioral economics. On the first topic (and, in particular, on cumulative prospect theory), the reader is referred to Wakker (2008), which also covers much of the classical theories covered here.

The first chapters are devoted to intuitive definitions of probabilities. We start with a few motivating examples and see how three intuitive definitions cope with these examples.

CHAPTER 1

Motivating Examples

Let us start with a concrete question such as what is the probability of another war in the Middle East erupting before December 31 of the current year? Of course, the focus should not be on this particular problem, but on the way we address it and, in fact, on its meaning. Similarly, I could ask you what is the probability of a stock market crash in the NY Stock Exchange during the next month. In both examples we have to define precisely what is meant by "war," or "stock market crash," so as to make the question well defined. Assume that we've done that. How would you approach the problems?

I believe that these are very complicated questions, not only because they have to do with complex processes, each involving many factors, but mostly because they are conceptually hard to deal with. To appreciate the difficulties, it may be useful to contrast this class of questions with three others. For concreteness, consider the following four questions, which, I will later argue, are in increasing order of conceptual difficulty:

1. I'm about to toss a coin. What is the probability of it coming up head?
2. I consider parking my car on the street tonight. What is the probability that it will be stolen overnight?
3. I am about to undergo a medical operation. What is the probability that I will survive it?
4. What is the probability of war in the Middle East this year?

It will prove useful to analyze these questions in light of the type of answers we can give them. From the very early days of probability theory (mid-seventeenth century), three ways of assigning probabilities to events can be documented. The first, sometimes called the "classical" approach, suggests that equal probabilities be assigned to all outcomes. The second, the "frequentist" approach, holds that the probability of an event should be the relative frequency with which it has been observed in the past. Finally, the "subjective" approach offers the mathematical machinery of probability theory as a way to model our vague beliefs and intuitions.

We will discuss these three methods in more detail. (In fact, the discussion of the subjective method will cover the entire classical theory of subjective expected utility maximization.) But before we do that, a word on free will may be in place.

CHAPTER 2

Free Will and Determinism

2.1 CAN FREE CHOICE BE PREDICTED?

In considering the four questions mentioned in Chapter 1, it is sometimes suggested that the fourth is conceptually more difficult than the others because it involves decisions of human beings, who have goals and desires, beliefs and ideas, and perhaps also free will. Can we hope to predict the behavior that results from all these? Will such prediction not be in conflict with the notion that individuals have free will?

These questions should bother anyone interested in the social sciences. If humans have free will, and if this means that their behavior cannot be predicted with any accuracy, we would have to declare social science an impossibility. The fact that we study economics probably indicates that we believe that some prediction is possible.[1] Indeed, the question of free will is usually not brought up in the context of the second question, namely, whether my car will be stolen. But cars are stolen by humans, and therefore any prediction regarding the car theft should also cope with the question of free will.

The fact is that there are many generalizations that are true of human beings and many predictions that can be made about them with a high degree of certainty. Whether individuals have free will or not, it is safe to predict that a $100 bill will not be lying on the sidewalk for 5 days in a row without being picked up by someone. Moreover, the individual who chooses to pick up the bill may well feel that she has free will in general, and even that she exercises it in this very act. This subjective experience of free will does not contradict our prediction.

By contrast, it may be very difficult to predict certain natural phenomena even if they do not involve the decision of any cognate entity to which free will can be ascribed. Combining the two, it appears that whether the behavior of a certain system can or cannot be predicted with confidence has little to do with the question of free will.

[1] This type of inference, from observed choices back to the beliefs that led to them, is what Part II of the book is about.

2.2 IS THE WORLD DETERMINISTIC?

A question that arises in the context of free will, but also in other contexts relating to probability and randomness, is whether the world we live in is deterministic, namely, one in which past events fully determine future ones. In the context of probability, the most popular version of determinism is causal or material determinism, which holds that knowledge of the laws of physics, coupled with exact knowledge of location and velocity of each particle, would suffice to fully determine the future. According to this view nothing is truly random, and probability can be used only when we do not have enough information or sufficient calculation capabilities.

This view is quite compelling in many situations. When I toss a coin, we all believe that an exact measurement of the angle at which my finger hits the coin, of the power exerted, and so forth could suffice for a precise calculation of the outcome of the toss. Indeed, if I let the coin simply slide from my finger down to the table, no one would accept the toss as "random," because it will be very easy to compute its outcome. Slightly more vigorous tosses, the claim goes, will differ from the languid slide only in degree, but not in kind.

However, it is clear that we can measure initial conditions only up to a certain accuracy. Even if we can exactly compute the outcome of a toss of a coin, it would be impossible to collect all the information needed to predict more complicated phenomena such as the exact impact of an earthquake or the result of a vote in Congress. This impossibility is not only a matter of practicality. Heisenberg's principle of uncertainty states that there are limits to the degree of accuracy with which mass and momentum can be simultaneously measured. Even within the scope of Newtonian physics, Chaos theory shows that deterministic systems can be sufficiently complex that all the information that can be gathered in reality will not suffice for complete prediction.[2]

One may model the world as deterministic, by introducing sufficiently many hidden variables (such as the world's future). But this model will be observationally equivalent to another model in which the world is nondeterministic. It follows that, as far as we will ever be able to know, the world is not deterministic.

2.3 IS FREE WILL OBSERVABLE?

Free will is a private phenomenon. I, for one, believe that I have free will, and I experience the exercise of free will whenever I make conscious decisions. I trust that all my readers also feel that they have free will. I can ask you to take a

[2] To be more concrete, even if we devoted all our resources to the production of measurement devices, the information gathered will be insufficient for perfect prediction. To complicate matters further, the measurement devices would be part of the system and would require to be measured themselves. (This brings to mind some of the discussions of Maxwell's demon, who can presumably decrease entropy of a system that does not contain itself.) That is, the impossibility of prediction that one can derive from Chaos theory goes a step beyond the commonsensical notion of "practicality."

few minutes, think whether you feel like turning the light on (or off), and then make a decision about it. You'd probably report a sensation of exercising free will and a feeling that you could have chosen differently. However, I do not have direct access to your sense of free will. I will not be able to distinguish between two individuals, one who feels that she has free will and another who only claims that he has, but who actually behaves in an automatic way without considering alternative courses of action. To put it bluntly, I know that I have free will, but whether anyone else has free will is a meaningless question.

This well-known observation is partly the reason that economists do not think that they should be interested in the phenomenon of free will. If it is unobservable, why should we care about it? If people make the same buying and selling decisions whether they do or do not have free will, who cares about the latter? Furthermore, one can argue that no scientific field should concern itself with the question of free will: observational equivalence implies that no amount of data, regarding economic behavior or otherwise, will ever suffice to determine whether an individual has free will. Free will is a metaphysical concept.

However, the sensation of free will, the subjective experience that one can choose among different options or could have chosen otherwise, is observable. Admittedly, only my own sensation of free will is directly observable to me, and that of others, only by their report. But this is the case with many psychological phenomena. Let us therefore refine our discussion: when we mention "free will" here, we refer to *psychological free will*, namely, the sensation of choice among several possible alternatives. This phenomenon has to do with our definition of the decision matrix and is therefore relevant to our analysis of our own decisions.

2.4 THE PROBLEM OF FREE WILL

We have hopefully agreed by now that (i) free will does not preclude prediction and (ii) the world is not deterministic. It would appear that free will should not bother us: whether people experience this sensation or not, there are no deterministic theories with which it may conflict. And the metaphysical notion of "real" free will certainly cannot conflict with observations. We can therefore go about our business, providing predictions of people's choices, without taking a position on the freedom of their will, whether this freedom is metaphysical or psychological. I claim, however, that a problem persists.

Much of the philosophical discussion of free will through the ages focuses on its conflict with determinism – causal determinism, as discussed earlier, or other notions, such as theological determinism, namely, the claim that the Almighty has already determined the future or at least that some entity (such as Fate) knows the future. Some recent contributions, relying on modern science, accept the view that determinism is the main challenge for the existence of free will. Penrose (1997) argues that the uncertainty at the level of elementary particles (Heisenberg's principle) may suffice to evolve into uncertainty about people's decisions, thus salvaging the notion of free will. Searle (2004) claims that our understanding of neurobiology at present does not yet prove that the

brain is deterministic and that, consequently, free will is an illusion, though he speculates that neurobiological research will get to this point.[3]

But to show the difficulty with free will one need not assume that *all* decisions are *predetermined*. It suffices that *one* decision be *known*. The logic is similar to suggesting a counterexample to a conjecture. The existence of one counterexample suffices. Similarly, if we can find one instance in which we have an undeniable sensation of free will on the one hand and practically certain knowledge of our choice on the other, we will have to admit that the sense of free will is illusory, at least in this example. In principle, one such example would suffice to put the notion of free will in doubt. In practice, I maintain that such examples abound.

Consider the following example: Sir Isaac Newton stands by a large window on the fourth floor. He contemplates the possibility of jumping out of the window. Should he jump, he considers two possibilities: he may hover in the air, enjoying the view, or crash to the ground. Being a rational decision maker, Newton contemplates the possibility of jumping and, given his knowledge of physics, concludes that crashing to the ground is a practical certainty. He now considers his own decision and decides not to jump. In so doing, he feels that he has made a decision and that he has exercised his free will. He could imagine choosing differently, and he decided not to.

Suppose that we are sitting with Sir Newton in his office throughout this process. Our limited knowledge of physics suffices for us to conclude, as does Newton, that a jump will result in a crash. With a lesser degree of certainty, but still quite confidently, we are willing to predict that Newton will not jump. We have seen many people next to many windows, and for the most part, they prefer to stay in their rooms. In short, we know Newton's choice with a high degree of certainty.[4]

But what about Sir Isaac Newton himself? Surely he knows himself at least as well as we know him. If we could conclude, based on our knowledge of human nature in general, that Newton will not jump, so can he. In fact, he is even in a privileged position to make predictions about himself.[5] Let us examine his reasoning process. A reasoned decision is supposed to take into account

[3] For a survey covering many other recent contributions, see Kane (2005).

[4] You may prefer to use the term "belief" in this context. The point is that this is a high degree of belief, which is probably as high as we can hope for in the social sciences and higher than our belief in, say, the weather forecast for the day after tomorrow. I do not think that the notion of free will can hinge on events that are possible but improbable, such as zero-probability events. One argument against a zero-probability event is aesthetic. It seems cheap. The other is more pragmatic: a zero-probability event will not be worth contemplating for even a negligible amount of time. The rational arguments given next can be restated when "knowledge" is replaced by "belief with very high probability."

[5] Some people have suicidal tendencies, but the majority do not. Our knowledge about Newton, based on statistics on a larger population, is less accurate than his own. Thus, for the majority of individuals it is true that they know that they are not suicidal with a higher degree of certainty than an outside observer would. Since we seek an example, we are justified in assuming that Newton is in this majority.

rules and regularities that are known to be quite accurate, to help us think about the consequences of our choices. We could imagine Newton drawing a decision tree and using all his knowledge to assign probabilities to the various branches in the tree, in particular, to cross out branches that he knows are practically impossible. This is how Newton concluded that, due to the gravitational force, he will not hover in the air should he jump. But, by the same logic, Newton can now cross out the branch "I jump" just as he previously crossed out the branch "I hover in the air" (conditional on jumping). By the time he finished the analysis there is no longer any decision to be made. Newton knows what his decision will be in the same sense that he knows what the outcomes of different decisions would be. When was a decision *taken* in this process? And how can Newton report an experience of free will if he cannot imagine a logically consistent world in which he chooses differently? How can we make sense of his claim "but I *could* have jumped"?

The paradoxical nature of free will stems from the co-occurrence of (i) the ability to imagine possible worlds that differ in terms of our choices and (ii) the fact that often our choices are practically known to us before we make them. Let me elaborate on these.

(i) Whatever free will is, it is closely related to the ability to conceive of different possible worlds, differing only in one's choice and its consequences. The ability to think of such different worlds, if not simultaneously then at least in the process of making a single decision, is essential to rational choice. And this ability is essential to, and maybe even a definition of, the sensation of free will. I feel that I exercise free will when I raise my arm, but not when my heart beats. The reason is that when consciously deciding to raise my arm, I can simultaneously imagine two worlds, one in which the arm is raised and the other in which it isn't. By contrast, I have never felt my heart stopping to beat, let alone decided to do so, and I cannot imagine a choice that would lead to this state of affairs. I therefore cannot argue that I exercised free will in letting my heart beat.

To see this point more clearly, suppose that you program a robot that will automatically make all the choices I make. Next, you allow the robot to speak and you want it to utter the statement "I hereby exercise free will" at the right moments, say, when I make such statements. Let us be slightly more demanding and require that the robot print out reasonable justifications for its choices. To this end, you will have to endow the robot with some reasoning ability and with the ability to distinguish between its own acts and the environment in which it lives. When facing an act, the robot will have to play around with some propositions of the form "If I do a, then the outcome will be x" and "conclude" that it prefers act a to act b. The robot will have to print several different such conditional statements for us to agree that it has exercised free will.

(ii) We typically know many things about ourselves. We know decisions that we have made, and we often have pretty good guesses about certain decisions that we are going to make. I know that I'm going to prefer coffee to tea. I know that I prefer not jumping out of the window to jumping. As a rational decision

maker, I gather data and make inferences. I cannot help observe regularities around me, and my own decisions in the past are included in the environment I study. Moreover, it is essential for rational choice that I learn things about myself. I need to know my "technical" capabilities, such as how fast I can run and how good my eyesight is. It will also be useful to know something about my mental capacities, such as how good my memory is and to what extent I follow my new year's resolutions. For this last purpose, I need to know my own choices in circumstances in which I felt that I was exercising free will. Finally, learning regularities about myself can be useful in predicting other people's behavior.

Let us consider the robot again. Will it know its own choices? Since you are programming it, you may try to avoid such knowledge. It is possible to restrict the inferences made by the robot to external events and to abort any calculation that refers to the robot's own choices. This will be somewhat artificial. Moreover, it will be inefficient, because the robot will not be able to use its own past decisions as guidance. Every time it will be offered coffee or tea it will have to make a calculation afresh. But the main difficulty with such a robot will be that it will not be as rational as I am. There will be some obvious inferences that it will fail to draw. Our own reasoning engines do not stop when it comes to our own choices in the past. We do learn about ourselves, and someone who fails to see obvious regularities in her own behavior is typically viewed as irrational.

We conclude that rationality makes two fundamental demands: first, we have to consider possible worlds that differ in terms of our choices. Second, we have to observe obvious regularities about ourselves, just like about any other relevant phenomenon. Taken together, we obtain the contradiction: we often need to consider as possible worlds that we know are impossible. Thus, the sensation of free will depends on our ability to suspend knowledge that we have about ourselves. Importantly, both the consideration of multiple possible worlds and the knowledge that some of them are impossible are dictated by rationality.

2.5 A RATIONAL ILLUSION

At the risk of belaboring obvious points, let me emphasize the following: not every decision will be known to the decision maker or to an outside observer before it has been taken. As long as the decision maker does not know what her choice is going to be, her sense of free will does not require that she suspend any knowledge she might have. In such a case the problem mentioned previously does not exist.

For example, assume that I have to choose between two quantities of a desirable good. We may think of tens of thousands of dollars or of years left to live – the point is that I truly prefer more to less. Consider now the following three choices:

(i) $\sqrt{17}$ or $(27/13)^2$
(ii) 2^3 or 3^2
(iii) 0 or 1.

In case (i) there is no difficulty. Reading the problem, it is not obvious to me which of the two numbers is larger. I therefore have to compute the outcome of both of my choices and then find out which one I prefer. An outside observer may have completed the calculation earlier and may already know what my choice will be. But I do not, and therefore my sense of free will does not contradict any knowledge I have at the time of starting my deliberation.

By contrast, case (iii) is one in which I know, more or less as soon as I read the problem, what my choice will be. I don't need a lengthy computation to figure out the meaning of 0 and 1. This is akin to Newton's problem, who stands by the window and has to decide whether to jump out or not. (The analogy is stronger if the numbers designate years one has to live, and 0 describes immediate death.) In both cases one needs to understand the two options and what they entail, but this understanding is quite trivial. The calculation that $1 > 0$ is about as immediate as the realization that jumping out of the window would result in death.

Case (ii) is brought as an intermediate case, suggesting that we cannot think of cases (i) and (iii) as qualitatively different. There is a range of difficulty levels, and a reasonable account of rational choice should describe a process that applies in all three cases. Thus, in all three cases we would like to assume that the decision maker makes a tentative assumption that she takes one option and thinks about the outcome. Then she does the same for the other option(s) and then she can make a reasoned decision. In case (i) there is no conflict with knowledge of her own choices, whereas in case (iii) there is. In cases such as (i) the decision maker may believe that she has free will, but in cases such as (iii) she has to admit that this was an illusion.

Efficiency of decision making might suggest that we need not compute our optimal choice anew every time. We may develop habits and rules that simplify our lives. It would therefore be tempting to categorize all decisions into two classes – the habitual decisions, such as in case (iii), in which there is no freedom of choice, but also no subjective sensation of free will; and the reasoned decisions, such as case (i), in which there is freedom of choice, but no a priori knowledge of what this choice is about to be. If such a dichotomy were possible, free will would not be such a pervasive problem: it would never clash with knowledge of one's own choice.

This, however, is not the case. Moreover, this could not be the case for rational individuals. First, however habitual a choice is, a rational individual should be able to ask herself whether she indeed wishes to stick to her habit. As soon as the question is posed, the individual will have to admit that she does know her choice, yet that she has a sensation of free will. Second, there will invariably be intermediate cases that are not regular enough to require no thought, yet sufficiently regular for the individual to know her own choice.

Rationality requires that we gather information and learn about the environment, our selves and our future selves included. Thus, we cannot escape knowledge of certain choices of ours. But rationality also requires that we be able to question these choices from time to time, and this means suspending