1

The logic of preference

The heart of Bayesian theory is the principle that rational choices maximize expected utility. This chapter begins with a statement of that principle. The principle is a formal one, and what it means is open to some interpretation. The remainder of the chapter is concerned with setting out an interpretation that makes the principle both correct and useful. I also indicate how I would defend these claims of correctness and usefulness.

1.1 EXPECTED UTILITY

If you need to make a decision, then there is more than one possible *act* that you could choose. In general, these acts will have different *consequences*, depending on what the true *state* of the world may be; and typically one is not certain which state that is. *Bayesian decision theory* is a theory about what counts as a rational choice in a decision problem. The theory postulates that a rational person has a probability function p defined over the states, and a utility function u defined over the consequences. Let $a(x)$ denote the consequence that will be obtained if act a is chosen and state x obtains, and let X be the set of all possible states. Then the *expected utility* of act a is the expected value of $u(a(x))$; I will refer to it as $EU(a)$. If X is countable, we can write

$$EU(a) = \sum_{x \in X} p(x)u(a(x)).$$

Bayesian decision theory holds that the choice of act a is rational just in case the expected utility of a is at least as great as that of any other available act. That is, rational choices maximize expected utility.

The principle of maximizing expected utility presupposes that the acts, consequences, and states have been formulated appropriately. The formulation is appropriate if the decision maker is (or ought to be) sure that

1. one and only one state obtains;
2. the choice of an act has no causal influence on which state obtains;[1] and
3. the consequences are sufficiently specific that they determine everything that is of value in the situation.

The following examples illustrate why conditions 2 and 3 are needed.

Mr. Coffin is a smoker considering whether to quit or continue smoking. All he cares about is whether or not he smokes and whether or not he lives to age 65, so he takes the consequences to be

Smoke and live to age 65
Quit and live to age 65
Smoke and die before age 65
Quit and die before age 65

The first-listed consequence has highest utility for Coffin, the second-listed consequence has second-highest utility, and so on down. And Coffin takes the states to be "Live to age 65" and "Die before age 65." Then each act–state pair determines a unique consequence, as in Figure 1.1. Applying the principle of maximizing expected utility, Coffin now reaches the conclusion that smoking is the rational choice. For he sees that whatever state obtains, the consequence obtained from smoking has higher utility than that obtained from not smoking; and so the expected utility of smoking is higher than that of not smoking. But if Coffin thinks that smoking might reduce the chance of living to age 65, then his reasoning is clearly faulty, for he has not taken account of this obviously relevant possibility. The fault lies in using states ("live to 65," "die before 65") that may be causally influenced by what is chosen, in violation of condition 2.

[1] Richard Jeffrey (1965, 1st ed.) maintained that what was needed was that the states be *probabilistically* independent of the acts. For a demonstration that this is not the same as requiring causal independence, and an argument that causal independence is in fact the correct requirement, see (Gibbard and Harper 1978).

	Live to 65	Die before 65
Smoke	Smoke and live to 65	Smoke and die before 65
Quit	Quit and live to 65	Quit and die before 65

Figure 1.1: Coffin's representation of his decision problem

Suppose Coffin is sure that the decision to smoke or not has no influence on the truth of the following propositions:

A: If I continue smoking then I will live to age 65.
B: If I quit smoking then I will live to age 65.

Then condition 2 would be satisfied by taking the states to be the four Boolean combinations of *A* and *B* (i.e., "*A* and *B*," "*A* and not *B*," "*B* and not *A*," and "neither *A* nor *B*"). Also, these states uniquely determine what consequence will be obtained from each act. And with these states, the principle of maximizing expected utility no longer implies that the rational choice is to smoke; the rational choice will depend on the probabilities of the states and the utilities of the consequences.

Next example: Ms. Drysdale is about to go outside and is wondering whether to take an umbrella. She takes the available acts to be "take umbrella" and "go without umbrella," and she takes the states be "rain" and "no rain." She notes that with these identifications, she has satisfied the requirement of act–state independence. Finally, she identifies the consequences as being that she is "dry" or "wet." So she draws up the matrix shown in Figure 1.2. Because she gives higher utility to staying dry than getting wet, she infers that the expected utility of taking the umbrella is higher than that of going without it, provided only that her probability for rain is not zero. Drysdale figures that the probability of rain is never zero, and takes her umbrella.

Since a nonzero chance of rain is not enough reason to carry an umbrella, Drysdale's reasoning is clearly faulty. The trouble

	Rain	No rain
Take umbrella	Dry	Dry
Go without	Wet	Dry

Figure 1.2: Drysdale's representation of her decision problem

	Rain	No rain
Take umbrella	Dry & umbrella	Dry & umbrella
Go without	Wet & no umbrella	Dry & no umbrella

Figure 1.3: Corrected representation of Drysdale's decision problem

is that carrying the umbrella has its own disutility, which has not been included in the specification of the consequences; this violates condition 3. If we include in the consequences a specification of whether or not the umbrella is carried, the consequences become those shown in Figure 1.3.

Suppose that these consequences are ranked by utility in this order:

Dry & no umbrella
Dry & umbrella
Wet & no umbrella

A mere positive probability for rain is now not enough to make taking the umbrella maximize expected utility; a small risk of getting wet would be worth running, for the sake of not having to carry the umbrella.

It is implicit in the definition of expected utility that each act has a unique consequence in any given state. This together with the previous conditions prevents the principle of maximizing expected utility being applied to cases where the laws of nature and the prior history of the world, together with the act chosen, do not determine everything of value in the situation (as might happen when the relevant laws are quantum mechanical).

In such a situation, taking the states to consist of the laws of nature and prior history of the world (or some part thereof) would not give a unique consequence for each state, unless the consequences omitted something of value in the situation. Including in the states a specification of what consequence will in fact be obtained avoids this problem but violates the requirement that the states be causally independent of the acts. There is a generalization of the principle of maximizing expected utility that can deal with decision problems of this kind; but I shall not present it here, because it introduces complexities that are irrelevant to the themes of this book. The interested reader is referred to (Lewis 1981).

1.2 Calculation

In many cases, it would not be rational to bother doing a calculation to determine which option maximizes expected utility. So if Bayesian decision theory held that a rational person would always do such calculations, the theory would be obviously incorrect. But the theory does not imply this.

To see that the theory has no such implication, note that doing a calculation to determine what act maximizes expected utility is itself an act; and this act need not maximize expected utility. For an illustration, consider again the problem of whether to take an umbrella. A fuller representation of the acts available would be the following:

t: Take umbrella, without calculating expected utility.
$\bar{t}$: Go without umbrella, without calculating expected utility.
c: Calculate the expected utility of t and $\bar{t}$ (with a view to subsequently making a choice that is calculated to maximize expected utility).[2]

Because calculation takes time, it may well be that t or $\bar{t}$ has higher expected utility than c; and if so, then Bayesian decision theory itself endorses not calculating expected utility.

[2] After calculating expected utility, one would choose an act without again calculating expected utility; thus the choice at that time will be between t and $\bar{t}$. So whether or not expected utility is calculated, one eventually chooses t or $\bar{t}$.

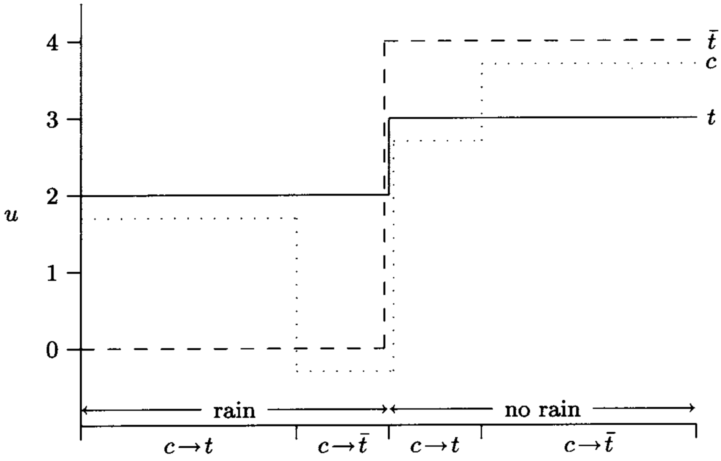

Figure 1.4: Acts of taking an umbrella (t), not taking umbrella ($\bar{t}$), and calculating (c) whether to choose t or $\bar{t}$.

Conversely, if c has higher expected utility than t or $\bar{t}$, then the theory holds that it is rational to do the calculation.

If we wanted to, we could do an expected utility calculation, to determine which of t, $\bar{t}$, and c maximizes expected utility. The situation is represented in Figure 1.4. Here "$c\rightarrow t$" means that choosing c (calculating the expected utility of t and $\bar{t}$) would lead to t being chosen;[3] and similarly for "$c\rightarrow \bar{t}$." When $c\rightarrow t$ is true, the utility of c is equal to that of t, less the cost of calculation; and when $c\rightarrow \bar{t}$ is true, the utility of c is equal to that of $\bar{t}$, less the same cost of calculation. Suppose that[4]

$$\begin{aligned} p(\text{rain}.c\rightarrow t) = p(\text{no rain}.c\rightarrow \bar{t}\,) &= .4 \\ p(\text{rain}.c\rightarrow \bar{t}\,) = p(\text{no rain}.c\rightarrow t) &= .1 \\ u(\bar{t}.\text{rain}) = 0; \qquad u(t.\text{no rain}) &= 3 \\ u(t.\text{rain}) = 2; \qquad u(\bar{t}.\text{no rain}) &= 4 \end{aligned}$$

[3] Assuming one will choose an act that is calculated to maximize expected utility, $c \rightarrow t$ includes all states in which calculation would show t to have a higher expected utility than $\bar{t}$. But it may also include states in which calculation would show t and $\bar{t}$ to have the same expected utility.

[4] Here the dot represents conjunction, and its scope extends to the end of the formula. For example, $p(\text{no rain}.c\rightarrow t)$ is the probability that there there is no rain and that $c\rightarrow t$.

Then

$$EU(t) = 2.5; \qquad EU(\bar{t}) = 2$$

and letting x be the cost of calculation,

$$EU(c) = 2.7 - x.$$

Thus Bayesian decision theory deems c the rational choice if x is less than 0.2, but t is the rational choice if x exceeds 0.2.

In the course of doing this second-order expected utility calculation, I have also done the first-order calculation, showing that $EU(t) > EU(\bar{t})$. But this does not negate the point I am making, namely that Bayesian decision theory can deem it irrational to calculate expected utility. For Bayesian decision theory also does not require the second-order calculation to be done. This point will be clear if we suppose that you are the person who has to decide whether or not to take an umbrella, and I am the one doing the second-order calculation of whether you should do a first-order calculation. Then I can calculate (using your probabilities and utilities) that you would be rational to choose t, and irrational to choose c; and this does not require you to do any calculation at all. Likewise, I could if I wished (and if it were true) show that you would be irrational to do the second-order calculation which shows that you would be irrational to do the first-order calculation.[5]

This conclusion may at first sight appear counterintuitive. For instance, suppose that t maximizes expected utility and, in particular, has higher expected utility than both $\bar{t}$ and c. Suppose further that you currently prefer $\bar{t}$ to the other options and would choose it if you do not do any calculation. Thus if you do no calculation, you will make a choice that Bayesian decision theory deems irrational. But if you do a calculation, you also do something deemed irrational, for calculation has a lower expected utility than choosing t outright. This may seem

[5] Kukla (1991) discusses the question of when reasoning is rational, and sees just these options: (a) reasoning is rationally required only when we *know* that the benefits outweigh the costs; or (b) a metacalculation of whether the benefits outweigh the costs is always required. Since (b) is untenable, he opts for (a). But he fails to consider the only option consistent with decision theory and the one being advanced here: That reasoning and metacalculations alike are rationally required just when the benefits *do* outweigh the costs, whether or not it is known that they do.

to put you in an impossible position. To know that t is the rational choice you would need to do a calculation, but doing that calculation is itself irrational. You are damned if you do and damned if you don't calculate.

This much is right: In the case described, what you would do if you do not calculate is irrational, and so is calculating. But this does not mean that decision theory deems you irrational no matter what you do. In fact, there is an option available to you that decision theory deems rational, namely t. So there is no violation here of the principle that 'ought' implies 'can'.

What the case shows is that Bayesian decision theory does not provide a means of guaranteeing that your choices are rational. I suggest that expecting a theory of rational choice to do this is expecting too much. What we can reasonably ask of such a theory is that it provide a criterion for when choices are rational, which can be applied to actual cases, even though it may not be rational to do so; Bayesian decision theory does this.[6]

Before leaving this topic, let me note that in reality we have more than the two options of calculating expected utility and choosing without any calculation. One other option is to calculate expected utility for a simplified representation that leaves out some complicating features of the real problem. For example, in a real problem of deciding whether or not to take an umbrella we would be concerned, not just with whether or not it rains, but also with how much rain there is and when it occurs; but we could elect to ignore these aspects and do a calculation using the simple matrix I have been using here. This will maximize expected utility if the simplifications reduce the computational costs sufficiently without having too great a probability of leading to the wrong choice. Alternatively, it might maximize expected utility to use some non-Bayesian rule, such as minimizing the maximum loss or settling for an act in which all the outcomes are "satisfactory."[7]

[6] Railton (1984) argues for a parallel thesis in ethics. Specifically, he contends that morality does not require us to always calculate the ethical value of acts we perform, and may even forbid such calculation in some cases.

[7] This is the Bayes/non-Bayes compromise advocated by I. J. Good (1983, 1988), but contrary to what Good sometimes says, the rationale for the compromise does not depend on probabilities being indeterminate.

1.3 Representation

Bayesian decision theory postulates that rational persons have the probability and utility functions needed to define expected utility. What does this mean, and why should we believe it?

I suggest that we understand attributions of probability and utility as essentially a device for interpreting a person's preferences. On this view, an attribution of probabilities and utilities is correct just in case it is part of an overall interpretation of the person's preferences that makes sufficiently good sense of them and better sense than any competing interpretation does. This is not the place to attempt to specify all the criteria that go into evaluating interpretations, nor shall I attempt to specify how good an interpretation must be to be sufficiently good. For present purposes, it will suffice to assert that if a person's preferences all maximize expected utility relative to some p and u, then it provides a perfect interpretation of the person's preferences to say that p and u are the person's probability and utility functions. Thus, having preferences that all maximize expected utility relative to p and u is a sufficient (but not necessary) condition for p and u to be one's probability and utility functions. I shall call this the *preference interpretation* of probability and utility.[8] Note that on this interpretation, a person can have probabilities and utilities without consciously assigning any numerical values as probabilities or utilities; indeed, the person need not even have the concepts of probability and utility.

Thus we can show that rational persons have probability and utility functions if we can show that rational persons have preferences that maximize expected utility relative to some such functions. An argument to this effect is provided by *representation theorems* for Bayesian decision theory. These theorems show that if a person's preferences satisfy certain putatively reasonable qualitative conditions, then those preferences are indeed representable as maximizing expected utility relative to some probability and utility functions. Ramsey (1926) and Savage (1954) each proved a representation theorem, and there have

[8] The preference interpretation is (at least) broadly in agreement with work in philosophy of mind, e.g., by Davidson (1984, pp. 159f.).

been many subsequent theorems, each making somewhat different assumptions. (For a survey of representation theorems, see [Fishburn 1981].)

As an illustration, and also to prepare the way for later discussion, I will describe two of the central assumptions used in Savage's (1954) representation theorem. First, we need to introduce the notion of *weak preference*. We say that you weakly prefer g to f if you either prefer g to f, or else are indifferent between them. The notation '$f \precsim g$' will be used to denote that g is weakly preferred to f. Now Savage's first postulate can be stated: It is that for any acts f, g, and h, the following conditions are satisfied:

Connectedness. *Either $f \precsim g$ or $g \precsim f$ (or both).*

Transitivity. *If $f \precsim g$ and $g \precsim h$, then $f \precsim h$.*

A relation that satisfies both the conditions of connectedness and transitivity is said to be a *weak* (or simple) *order*. So an alternative statement of this postulate is that the relation $\precsim$ is a weak order on the set of acts.

Savage's second postulate asserts that if two acts have the same consequences in some states, then the person's preferences regarding those acts should be independent of what that common consequence is. For example, in Figure 1.5, f and g have the same consequence on $\bar{A}$, and f' and g' are the result of replacing that common consequence with something else; so according to this postulate, if $f \precsim g$, then it should be that $f' \precsim g'$. Formally, the postulate is that for any acts f, g, f', and g', and for any event A, the following condition holds:[9]

Independence. *If $f = f'$ on A, $g = g'$ on A, $f = g$ on $\bar{A}$, $f' = g'$ on $\bar{A}$, and $f \precsim g$, then $f' \precsim g'$.*

[9] This postulate is often referred to as the *sure-thing principle*, a term that comes from Savage (1954, p. 21). But as I read Savage, what he means by the sure-thing principle is not any postulate of his theory, but rather an informal principle that motivates the present postulate. In Section 3.2.3 I will discuss that principle, and consider how well it motivates the postulate. So for my purposes, it would confuse an important distinction to refer to this postulate as "the sure-thing principle."