Chapter 4
DECISIONS UNDER RISK: UTILITY

4-1. Interval Utility Scales

Utilities are just as critical to our account of decisions under risk as probabilities since the rule of maximizing expected utility operates on them. But what are utilities? And what do utility scales measure? In discussing decisions under ignorance I hinted at answers to these questions. But such allusions will not suffice for a full and proper understanding of decisions under risk. So let us begin with a more thorough and systematic examination of the concept of utility.

The first point we should observe is that ordinal utility scales do not suffice for making decisions under risk. Tables 4-1 and 4-2 illustrate why this is so. The

4-1

A_1	6 ¼	1 ¾
A_2	5 ¼	2 ¾

expected utilities of A_1 and A_2 are 9/4 and 11/4, respectively, and so A_2 would be picked. But if we transform table 4-1 ordinally to table 4-2 by simply raising

4-2

A_1	20 ¼	1 ¾
A_2	5 ¼	2 ¾

the utility number 6 to 20, the expected utilities are now 23/4 and 11/4, which results in A_1 being picked. Thus two scales that are ordinal transformations of each other might fail to be equivalent with respect to decisions under risk.

This stands to reason anyway. Ordinal scales represent only the relative standings of the outcomes; they tell us what is ranked first and second, above and below, but no more. In a decision under risk it is often not enough to know that you prefer one outcome to another; you might also need to know whether you prefer an outcome *enough* to take the risks involved in obtaining it. This is reflected in our disposition to require a much greater return on an investment of \$1,000,000 than on one of \$10—even when the probabilities of losing the investment are the same.

Happily, interval scales are all we require for decisions under risk. In addition to recording an agent's ranking of outcomes, we need measure only the relative lengths of the "preference intervals" between them. To understand what is at stake, suppose I have represented my preferences for cola (C), ice cream (I), apples (A), and popcorn (P) on the following line.

least preferred			most preferred
P	A	C	I

If I now form a scale by assigning numbers to the line, using an ordinal scale would require me only to use numbers in ascending order as I go from left to right. But if I use an interval scale, I must be sure that the *relative* lengths of the intervals on the line are reflected as well. Thus I could not assign 0 to P, 1 to A, 2 to C, and 3 to I, for this would falsely equate the interval between, say P and A with that between C and I. More generally, if items x, y, z, and w are assigned utility numbers $u(x)$, $u(y)$, $u(z)$, and $u(w)$ on an interval scale, these numbers must satisfy the following conditions:

a. xPy if and only if $u(x) > u(y)$.
b. xIy if and only if $u(x) = u(y)$.
the preference interval between x and y is greater than or equal to that between z and w if and only if
$|u(x) - u(y)| \geq |u(z) - u(w)|$.

More than one assignment of numbers will satisfy these two conditions but every assignment that does is a positive linear transformation of every other one that does. To illustrate this point, suppose I assign numbers to my preferences as indicated here.

least preferred			most preferred
1	2	7	10
P	A	C	I

Then I have properly represented both the ordinal and interval information. But I could have used other numbers, such as the next set.

least preferred			most preferred
15	25	75	105
P	A	C	I

These are obtained from the first set by a positive linear transformation. (What is it?)

Two ordinal scales count as equivalent if and only if they can be obtained from each other by means of order-preserving (ordinal) transformations. Two interval scales will count as equivalent if and only if they can be obtained from each other by means of positive linear transformations.

Another type of scale with which you are familiar is ratio scales. We use these for measuring lengths (in yards, feet, meters) or weights (in pounds, grams, ounces). The scales all share two important features: First, they all have natural zero points (no length, no speed, no weight), and second, the scales are used to represent the ratio of a thing measured to some standard unit of measurement. (Something 10 yards long bears the ratio of 10 to 1 to a standard yardstick; the latter can be laid off ten times against the former.) In converting from one ratio scale to another we multiply by a positive constant. (Thus, to obtain inches from feet we multiply by 12.) This turns out to be the equivalence condition for ratio scales: Two ratio scales are equivalent to each other if and only if they may be obtained from each other by multiplying by positive constants. This is a special case of a positive linear transformation; thus ratio scales are a tighter kind of interval scale.

One way to appreciate the difference between ratio and interval scales is to think of changing scales in terms of changing the labels on our measuring instruments. If we had a measuring rod 9 feet long, labeled in feet, and relabeled it in yards, we would need fewer marks on the stick. If we labeled it in inches we would need more marks. But in either case the zero point would remain the same. This is not necessarily so with interval scales. If we had a thermometer marked in degrees Fahrenheit and changed it to degrees Celsius, we would use fewer marks (between the freezing and boiling points of water there are 100 Celsius units in contrast to 180 Fahrenheit units), and we would also shift the zero point upward. Because there is no fixed zero point on our temperature scales, we must be quite careful when we say that something is twice as hot as something else. Suppose that at noon yesterday, the outdoor temperature was at the freezing point of water and that today at noon it measured 64 degrees on the Fahrenheit scale. Was it twice as hot today at noon as it was yesterday? On the Fahrenheit scale, yes, but not on the Celsius scale. Suppose, for contrast, that I am driving at 60 miles per hour and you are driving at 30. Now convert our speeds to kilometers per hour. You will see that I would still be driving twice as fast as you.

Since we will use interval scales rather than ratio scales to represent preference intervals, we cannot assume that arithmetic operations that we perform freely on speeds, weights, or distances make sense for utilities. I will return to this point later.

We have already seen that decisions under risk require more than ordinal scales. Will interval scales suffice? Or must we move on to ratio scales? No, we need not, for *any two scales that are positive linear transformations of each other will produce the same ranking of acts in a decision table* and, thus, will yield the same decisions. In short, interval scales suffice for decisions under risk. Let us now prove this.

Let table 4-3 represent any decision table and any two acts in it. I will call the acts A_i and A_j, and for convenience I will write them next to each other, but in fact they might have many rows between them or above and below them.

4-3

A_i	u_1 p_1	u_2 p_2	. . .	u_n p_n
A_j	v_1 q_1	v_2 q_2	. . .	v_n q_n

The u's and v's are utility numbers and the p's and q's are probabilities. Since the states may be dependent on the acts, we cannot assume that the p's and q's are equal. However, we can assume that the probabilities across for each row sum to 1. The expected utilities for A_i and A_j are given by the formulas

$$EU(A_i) = u_1p_1 + u_2p_2 + \ldots + u_np_n$$
$$EU(A_j) = v_1q_1 + v_2q_2 + \ldots + v_nq_n.$$

Now a positive linear transformation of the scale used in table 4-3 would cause each u and each v to be replaced, respectively, by $au + b$ and $av + b$ (with $a > 0$). Thus after the equation the formulas for the expected utilities would be

$$EU_{\text{new}}(A_i) = (au_1 + b)p_1 + (au_2 + b)p_2 + \ldots + (au_n + b)p_n$$
$$EU_{\text{new}}(A_j) = (av_1 + b)q_1 + (av_2 + b)q_2 + \ldots + (av_n + b)q_n.$$

If we multiply through by the p's and q's and then gather at the end all the terms that contain no u's or v's we obtain

$$EU_{\text{new}}(A_i) = [au_1p_1 + au_2p_2 + \ldots + au_np_n] + [bp_1 + bp_2 + \ldots + bp_n]$$
$$EU_{\text{new}}(A_j) = [av_1q_1 + av_2q_2 + \ldots + av_nq_n] + [bq_1 + bq_2 + \ldots + bq_n].$$

If we now factor out the a's the expressions remaining in the left-hand brackets are the old expected utilities. On the other hand, if we factor out the b's the expressions remaining in the right-hand brackets are p's and q's that sum to 1. Thus our new utilities are given by this pair of equations:

$$EU_{\text{new}}(A_i) = aEU(A_i) + b$$
$$EU_{\text{new}}(A_j) = aEU(A_j) + b.$$

Now since $a > 0$, $aEU(A_i)$ is greater than (less than, or equal to) $aEU(A_j)$ just in case $EU(A_i)$ stands in the same relation to $EU(A_j)$. Furthermore, this relation is preserved if we add b to $aEU(A_i)$ and $aEU(A_j)$. In other words, $EU_{\text{new}}(A_i)$ and $EU_{\text{new}}(A_j)$ will stand in the same order as $EU(A_i)$ and $EU(A_j)$. This means that using expected utilities to rank the two acts A_i and A_j will yield the same results whether we use the original scale or the positive linear transformation of it. But our reasoning has been entirely general, so the same conclusion holds for all expected utility rankings of any acts using these two scales. In short, they are equivalent with respect to decision making under risk.

PROBLEMS

1. Suppose you can bet on one of two horses—Ace or Jack—in a match race. If Ace wins you are paid \$5; if he loses you must pay \$2 to the track. If you bet on Jack and he loses, you pay the track \$10. You judge each horse to be as likely to win as the other. Assuming you make your decisions on the basis of expected monetary values, how much would a winning bet on Jack have to pay before you would be willing to risk \$10?
2. Suppose the interval scale u may be transformed into u' by means of the transformation

$$u' = au + b \qquad (a > 0).$$

Give the transformation that converts u' back into u.

3. Suppose the u' of the last problem can be transformed into u'' by means of the transformation

$$u'' = cu' + d \qquad (c > 0).$$

Give the transformation for converting u into u''.

4. Suppose s and s' are equivalent ratio scales. Show that if $s(x) = 2s(y)$, then $s'(x) = 2s'(y)$.
5. Suppose you have a table for a decision under risk in which the probabilities are independent of the acts. Show that if you transform your utility numbers by adding the number b_i to each utility in column i (and assume that the numbers used in different columns are not necessarily the same), the new table will yield the same ordering of the acts.

4-2. Monetary Values vs. Utilities

A popular and often convenient method for determining how strongly a person prefers something is to find out how much money he or she will pay for it. As a general rule people pay more for what they want more; so a monetary scale can be expected to be at least an ordinal scale. But it often works as an interval scale too—at least over a limited range of items. A rough test of this is the agent's being indifferent between the same increase (or reduction) in prices over a range of prices, since the intervals remain the same though the prices change. Thus if I sense no difference between \$5 increases (e.g., from \$100 to \$105) for prices between \$100 and \$200, it is likely that a monetary scale can adequately function as an interval scale for my preferences for items in that price range. Within this range it would make sense for me to make decisions under risk on the basis of expected monetary values (EMVs).

Since we are so used to valuing things in terms of money—we even price intangibles, such as our own labor and time or a beautiful sunset, as well as necessities, such as food and clothing—it is no surprise that EMVs are often used as a basis for decisions under risk. My earlier insurance and car purchase examples typify this approach. Perhaps this is the easiest and most appropriate method for making business decisions, for here the profit motive is paramount.

It is both surprising and disquieting that a large number of nonbusiness decisions are made on the basis of EMVs. For example, it is not unusual for government policy analysts to use monetary values to scale nonmonetary outcomes such as an increase or decrease in highway deaths or pollution-induced cancers. To make a hypothetical example of a historical case, consider the federal government's decision in 1976 to vaccinate the population against an expected epidemic of swine flu. The *actual relevant outcomes* might have been specified in terms of the number of people contracting the disease, the number of deaths and permanent disabilities, and the cost and inconvenience of giving or receiving the vaccine, but instead of assigning values to these outcomes, I will suppose (as is probably the case) that policy analysts turned to their economic consequences. This meant evaluating alternatives first in terms of lost working days and then in terms of a reduced gross national product (GNP), and so on until the cost of administering the vaccine could be compared to the expected benefit to be derived from it. To illustrate this in a simplified form using made-up figures, suppose a flu epidemic will remove five million people from the work force for a period of five days. (Some will be sick, others will care for the sick. I am also talking about absences over and above those expected in normal times.) Further, suppose the average worker contributes \$200 per five days of work to the GNP. Finally, suppose the GNP cost of the vaccine program is \$40,000,000, that without it there is a 90% chance of an epidemic and with it only a 10% chance. Then we can set up decision table 4-4, which values out-

4-4	Epidemic	No Epidemic
Have Vaccine Program	−\$1,040 million .1	−\$40 million .9
No Program	−\$1,000 million .9	0 .1

comes in terms of dollar costs to the GNP. The EMV of having the program is −\$140 million and that of not having it is −\$900 million; thus, under this approach at least, the program should be initiated since it minimizes the cost to the GNP. (This example also illustrates an equivalent approach to decisions under risk: Instead of maximizing expected gains one minimizes expected losses. See exercises 1–3 in the next Problems section.)

Perhaps the EMV approach to large-scale decisions is the only practical alternative available to policy analysts. After all, they should make some attempt to factor risks into an analysis of costs and benefits, and that will require at least an approximation to an interval scale. Monetary values provide an accessible and publicly understandable basis for such a scale.

But few people would find EMVs a satisfactory basis for every decision; they will not even suffice for certain business decisions. Sometimes making greater (after-tax) profits is not worth the effort. Just as we are rapidly reaching the point where it is not worth bending over to pick up a single penny, it might not be worth a company's effort to make an extra $20,000. More money would actually have less utility; so money could not even function as an ordinal scale. Of course, this completely contravenes the way an EMVer sees things.

Furthermore, some apparently rational businesspeople gladly sacrifice profits for humanitarian, moral, or aesthetic considerations—even when those considerations cannot be justified in greater profits in the long run. Many companies sponsor scholarships for college students, knowing full well that the associated tax benefits, improvements in corporate image, and recruitment have but a small probability of producing profits in excess of the costs. These people cannot base their decisions solely on EMVs.

On a personal level, too, the "true value" of an alternative is often above or below its EMV. Thus I might pay more for a house in the mountains than its EMV (calculated, say, by real estate investment counselors) because the beauty and solitude of its setting make up the extra value *to me*, or I might continue to drive the old family car out of sentiment long after it has stopped being economical to do so. To see the divergence between EMVs and true values in simple decision problems, consider these examples.

Example 1. True value below EMV. After ten years of work you have saved $15,000 as a down payment on your dream house. You know the house you want and need only turn over your money to have it. Before you can do that your stockbroker calls with a "very hot tip." If you can invest your $15,000 for one month, he can assure you an 80% chance at a $50,000 return. Unfortunately, if the investment fails, you lose everything. Now the EMV of this investment is $37,000—well above the $15,000 you now have in hand. Your broker points out that you can still buy the house a month from now and urges you to make the investment. But you do not, because you feel you cannot afford to risk the $15,000. For you, making the investment is worth less than having $15,000 in hand, and thus it is worth less than its EMV.

Example 2. True value above EMV. Suppose you have been trying to purchase a ticket for a championship basketball game. Tickets are available at $20 but you have only $10 on you. A fellow comes along and offers to match your $10 on a single roll of the dice. If you roll snake eyes, you get the total pot; otherwise he takes it. This means that your chances are 1 in 12 of ending up with the $20 you need and 11 in 12 of losing all you have. The EMV of this is −$7.50—definitely below the $10 you have in hand. But you take the bet, since having the $20 is worth the risk to you. Thus the EMV of this bet is below its true value to you.

In addition to practical problems with EMVs there is an important philosophical difficulty. Even if you are guided solely by the profit motive, there is no logical connection between the EMV of a risk and its monetary value. Consider this example. You alone have been given a ticket for the one and only lottery your state will have. (Although the lottery is in its very first year, the legislature has already passed a bill repealing it.) The ticket gives you one chance in a million of winning $1,000,000. Since you lose nothing if you fail to win, the EMV of the ticket is $1. After the lottery is drawn you win nothing or

\$1,000,000 – but never \$1. Thus how can we connect this figure with a cash value for the bet? Why – assuming money is all that counts – would it be rational for you to sell your ticket for \$2? One is tempted to answer in terms of averages or long runs: If there were many people in your situation, their winnings would average \$1; if this happened to you year after year, your winnings would average \$1. But this will not work for the case at hand, since, by hypothesis, you alone have a free ticket and there will be only one lottery. With a one-shot decision there is nothing to average; so we have still failed to connect EMVs with cash values.

PROBLEMS

1. Given a utility scale u, we can formulate a disutility scale, $d(u)$, by multiplying each entry on the u-scale by -1. The expected disutility of an act is calculated in the same way as its expected utility except that every utility is replaced by its corresponding disutility. Reformulate the rule for maximizing expected utilities as a rule involving expected disutilities.
2. Show that the expected disutility of an act is equal to -1 times its expected utility.
3. Show that using a disutility scale and the rule you formulated in problem 1 yields the same rankings of acts as maximizing expected utilities does.
4. The St. Petersburg game is played as follows. There is one player and a "bank." The bank tosses a fair coin once. If it comes up heads, the player is paid \$2; otherwise the coin is tossed again with the player being paid \$4 if it lands heads. The game continues in this way with the bank continuing to double the amount set. The game stops when the coin lands heads.

   Consider a modified version of this game. The coin will be tossed no more than two times. If heads comes up on neither toss, the player is paid nothing. What is the EMV of this game?

   Suppose the coin will be tossed no more than n times. What is the EMV of the game?

   Explain why an EMVer should be willing to pay any amount to play the unrestricted St. Petersburg game.
5. Consider the following answer to the one-shot lottery objection to EMVs: True, there is only one lottery and only one person has a free ticket. But in a hypothetical case in which there were many such persons or many lotteries, we would find that the average winnings would be \$1. Let us identify the cash value of the ticket with the average winnings in such hypothetical cases. It follows immediately that the cash value equals the EMV.

   Do you think this approach is an adequate solution to the problem of relating EMVs to cash values?

4-3. Von Neumann-Morgenstern Utility Theory

John Von Neumann, a mathematician, and Oskar Morgenstern, an economist, developed an approach to utility that avoids the objections we raised to EMVs.

Although Ramsey's approach to utility antedates theirs, today theirs is better known and more entrenched among social scientists. I present it here because it separates utility from probability, whereas Ramsey's approach generates utility and subjective probability functions simultaneously.

Von Neumann and Morgenstern base their theory on the idea of measuring the strength of a person's preference for a thing by the risks he or she is willing to take to receive it. To illustrate that idea, suppose that we know that you prefer a trip to Washington to one to New York to one to Los Angeles. We still do not know how *much more* you prefer going to Washington to going to New York, but we can measure that by asking you the following question: Suppose you were offered a choice between a trip to New York and a lottery that will give you a trip to Washington if you "win" and one to Los Angeles if you "lose." How great a chance of winning would you need to have in order to be indifferent between these two choices? Presumably, if you prefer New York quite a bit more than Los Angeles, you will demand a high chance at Washington before giving up a guaranteed trip to New York. On the other hand, if you only slightly prefer New York to Los Angeles, a small chance will suffice. Let us suppose that you reply that you would need a 75% chance at Washington—no more and no less. Then, according to Von Neumann and Morgenstern, we should conclude that the New York trip occurs 3/4 of the way between Washington and Los Angeles on your scale.

Another way of representing this is to think of you as supplying a ranking not only of the three trips but also of a lottery (or gamble) involving the best and worst trips. You must be indifferent between this lottery and the middle-ranked trip. Suppose we let the expression

$L(a, x, y)$

stand for the lottery that gives you a chance equal to a at the prize x and a chance equal to $1 - a$ at the prize y. Then your ranking can be represented as

Washington
New York, L(3/4, Washington, Los Angeles)
Los Angeles.

We can use this to construct a utility scale for these alternatives by assigning one number to Los Angeles, a greater one to Washington, and the number 3/4 of the way between them to New York. Using a 0 to 1 scale, we would assign 3/4 to New York—but any other scale obtained from this by a positive linear transformation will do as well.

Notice that since the New York trip and the lottery are indifferent they are ranked together on your scale. Thus the utility of the lottery itself is 3/4 on a 0 to 1 scale. But since on that scale the utilities of its two "prizes" (the trips) are 0 and 1, the expected utility of the lottery is also 3/4. We seem to have forged a link between utilities and expected utilities. Indeed, Von Neumann and Morgenstern showed that if an agent ranks lotteries in the manner of our example, their utilities will equal their expected utilities. Let us also note that the Von

Neumann-Morgenstern approach can be applied to any kind of item—whether or not we can sensibly set a price for it—and that it yields an agent's personal utilities rather than monetary values generated by the marketplace. This permits us to avoid our previous problems with monetary values and EMVs.

(You might have noticed the resemblance between the Von Neumann-Morgenstern approach and our earlier approach to subjective probability, where we measured degrees of belief by the amount of a valued quantity the agent was willing to stake. Ramsey's trick consisted in using these two insights together without generating the obvious circle of defining probability in terms of utility and utility in turn in terms of probability.)

The Von Neumann-Morgenstern approach to utility places much stronger demands on agents' abilities to fix their preferences than do our previous conditions of rationality. Not only must agents be able to order the outcomes relevant to their decision problems, they must also be able to order all lotteries involving these outcomes, all compound lotteries involving those initial lotteries, all lotteries compounded from those lotteries, and so on. Furthermore, this ordering of lotteries and outcomes (I will start calling these *prizes*) is subject to constraints in addition to the ordering condition. Put in brief and rough form, these are: (1) Agents must evaluate compound lotteries in agreement with the probability calculus (reduction-of-compound-lotteries condition); (2) given three alternatives A, B, C with B ranked between A and C, agents must be indifferent between B and some lottery yielding A and C as prizes (continuity condition); (3) given two other lotteries agents will prefer the one giving the better "first" prize—if everything else is equal (better-prizes condition); (4) given two otherwise identical lotteries, agents will prefer the one that gives them the best chance at the "first" prize (better-chances condition). If agents can satisfy these four conditions plus the ordering condition of chapter 2, we can construct an interval utility function u with the following properties:

(1) $u(x) > u(y)$ if and only if xPy
(2) $u(x) = u(y)$ if and only if xIy
(3) $u[L(a, x, y)] = au(x) + (1 - a)u(y)$
(4) Any u' also satisfying (1)–(3) is a positive linear transformation of u.

You should recognize (1) and (2) from our discussion of decisions under ignorance (chapter 2). They imply that u is at least an ordinal utility function. But (3) is new. It states that the utility of a lottery is equal to its expected utility. We can also express this by saying that u has the *expected utility property*. The entire result given by (1)–(4) is known as the *expected utility theorem*. Let us now turn to a rigorous proof of it.

First we must specify lotteries more precisely than we have. The agent is concerned with determining the utilities for some set of outcomes, alternatives, or prizes. Let us call these *basic prizes*. Let us also assume that the number of basic prizes is a finite number greater than 1 and that the agent is not indifferent between all of them. Since we can assume that the agent has ranked the prizes, some will be ranked at the top and others at the bottom. For future reference, let us select a top-ranked prize and label it "B" (for best). Let us also select a bottom-ranked prize and label it "W" (for worst).

I will now introduce compound lotteries by the following rule of construction. (In mathematical logic this is called an inductive definition.)

Rule for Constructing Lotteries:

1. Every basic prize is a lottery.
2. If L_1 and L_2 are lotteries, so is $L(a, L_1, L_2)$, where $0 \leq a \leq 1$.
3. Something is a lottery if and only if it can be constructed according to conditions 1 and 2.

Thus lotteries consist of basic prizes, simple lotteries involving basic prizes, further lotteries involving basic prizes or simple lotteries, ad infinitum. This means that B and W are lotteries, $L(1/2, B, W)$ and $L(3/4, W, B)$ are, and so are $L(2/3, B, L(1/2, B, W))$ and $L(1, L(0, B, W), W)$

PROBLEMS

1. Why can we not assume that there is a single best prize?
2. Why have I assumed that there is more than one basic prize? That the agent is not indifferent between all the prizes? If these assumptions were not true, would the expected utility theorem be false?
3. What chance at B does each of the following lotteries give?
   a. $L(1, B, W)$
   b. $L(1/2, L(1, W, B), L(1/2, B, W))$
   c. $L(a, B, B)$
4. Why should an agent be indifferent between $L(a, B, W)$ and $L(1 - a, W, B)$?
5. Show how to construct one of our lotteries that is equivalent to the lottery with *three* prizes, A, B, and C, that offers a 50% chance of yielding A and 25% chances of yielding B and C.

Now that we have a precise characterization of lotteries, let us turn to precise formulations of the "rationality" conditions the agent must satisfy. The first of these is the familiar *ordering condition*, applied this time not just to basic prizes but also to all lotteries. This means that conditions O1–O8 (discussed in chapter 2) apply to all lotteries. An immediate consequence of this is that we can partition the lotteries into ranks so as to rank together lotteries between which the agent is indifferent while placing each lottery below those the agent prefers to it.

The next condition is called the *continuity condition* because one of its consequences is that the ordering of the lotteries is continuous. It is formulated as follows:

> For any lotteries x, y, and z, if xPy and yPz, then there is some real number a such that $0 \leq a \leq 1$ and $y\ I\ L(a, x, z)$.

In less formal terms this says that if the agent ranks y between x and z, there is some lottery with x and z as prizes that the agent ranks along with y.

The *better-prizes condition* is next. Intuitively, it says that other things being equal, the agent prefers one lottery to another just in case the former involves better prizes. Put formally:

For any lotteries x, y, and z and any number a $(0 \leq a \leq 1)$, xPy if and only if $L(a, z, x)\ P\ L(a, z, y)$ and $L(a, x, z)\ P\ L(a, y, z)$.

There is also the *better-chances condition*, which says roughly that, other things being equal, the agent prefers one lottery to another just in case the former gives a better chance at the better prize. Put in precise terms:

For any lotteries x and y and any numbers a and b (both between 0 and 1, inclusively), if xPy, then $a > b$ just in case $L(a, x, y)\ P\ L(b, x, y)$.

The final condition is called the *reduction-of-compound-lotteries condition* and requires the agent to evaluate compound lotteries in accordance with the probability calculus. To be exact, it goes:

For any lotteries x and y and any numbers a, b, c, d (again between 0 and 1 inclusively), if $d = ab + (1 - a)c$, then $L(a, L(b, x,y), L(c, x, y))\ I\ L(d, x, y)$.

To get a better grip on this condition, let us introduce the *lottery tree* notation (figure 4-1), which is similar to the decision tree notation.

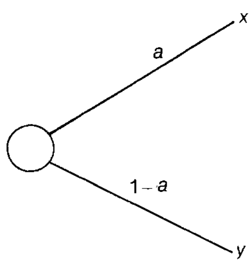

Figure 4-1

This diagram represents the simple lottery that yields the prize x with a chance of a and the prize y with a chance of $1 - a$. Compound lotteries can be represented by iterating this sort of construction, as figure 4-2 illustrates. This is a two-stage lottery whose final prizes are x and y. What are the chances of getting x in this lottery? There is a chance of a at getting into lottery 2 from lottery 1 and a chance of b of getting x. In other words, there is an ab chance of getting x through lottery 2. Similarly, there is a $(1 - a)c$ chance of getting x through lottery 3. Since these are the only routes to x and they are mutually exclusive, the chances for x are $ab + (1 - a)c$. The same type of reasoning shows that the chances for y are $a(1 - b) + (1 - a)(1 - c)$. If we set $d = ab + (1 - a)c$, a little algebra will show that $1 - d = a(1 - b) + (1 - a)c$. The reduction-of-compound-lotteries condition simply tells us that the agent must be indifferent between the compound lottery given earlier and the next simple one (figure 4-3). Notice, by the way, that the condition is improperly named, since the agent is indifferent

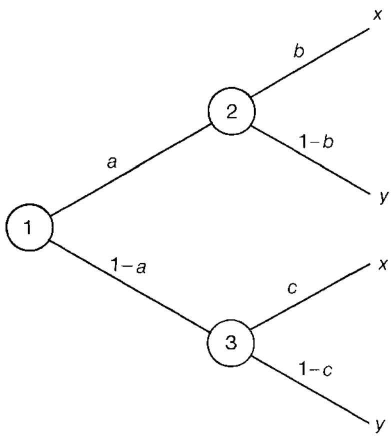

Figure 4-2

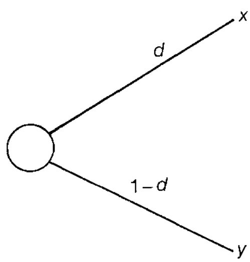

Figure 4-3

to both reductions and *expansions* of lotteries so long as they accord with the probability calculus.

Let us turn now to the proof of the expected utility theorem. I will divide the proof into two parts. The first part will establish that, given that the agent satisfies the ordering, continuity, better-prizes, better-chances, and reduction-of-compound-lotteries conditions, there is a utility function, u, satisfying the expected utility property that represents his preferences. This is called the *existence* part of the proof, since it proves that there exists a utility function having the characteristics given in the theorem. The second part of the proof is called the *uniqueness* part, because it establishes that the utility function constructed in the first part is unique up to positive linear transformations; that is, it is an interval utility function.

Directing ourselves now to the proof of the existence of u, recall that we have already established that there are at least two basic prizes B and W, where the agent prefers B to W and regards B as at least as good as any basic prize and every basic prize as at least as good as W. Since all lotteries ultimately pay in basic prizes and are evaluated by the agent in terms of the probability calculus, there is no lottery ranked above B or below W. (See exercises 4-8 in the next problems section.) Accordingly, fixing the top of our utility scale at 1 and the bottom at 0, we stipulate that

$u(B)=1$ and $u(x)=1$ for all lotteries x indifferent to B,
$u(W)=0$ and $u(x)=0$ for all lotteries x indifferent to W.

Having taken care of the lotteries at the extremes we must now define u for those in between. So let x be any lottery for which

BPx and xPW

holds. Applying the continuity axiom to this case, we can conclude that there is a number a, where $0 \leq a \leq 1$, such that

$x\ I\ L(a, B, W)$.

If there is just one such a, we will be justified in stipulating that $u(x)=a$. So let us assume that $a' \neq a$ and $x\ I\ L(a', x, y)$ and derive a contradiction. Since a and a' are assumed to be distinct, one is less than the other. Suppose $a < a'$. Then by the better-chances condition,

$L(a', B, W)\ P\ L(a, B, W)$,

but this contradicts the ordering condition since both lotteries are indifferent to x. A similiar contradiction follows from the alternative that $a' < a$. Having derived a contradiction from either alternative, we may conclude that $a = a'$.

We are now justified in stipulating that

$u(x)=a$,

where a is the number for which $x\ I\ L(a, B, W)$. Note that by substituting equals for equals we obtain

$(*)\ x\ I\ L(u(x), B, W)$;

that is, the agent is indifferent between x and the lottery that gives a $u(x)$ chance at B and a $1-u(x)$ chance at W.

So far we have simply established the existence of a function u that assigns a number to each lottery. We must also show that this is an (interval) utility function that satisfies the expected utility property.

Let us first show that, for all lotteries x and y,

(1) xPy if and only if $u(x) > u(y)$.

By the better chances condition we have

(a) $L(u(x), B, W)\ P\ L(u(y), B, W)$ if and only if $u(x) > u(y)$.

By $(*)$, above, we have

(b) $x\ I\ L(u(x), B, W)$ and $y\ I\ L(u(y), B, W)$.

Using the ordering condition, we can easily prove

(c) for all lotteries x, y, z, and w, if xIy and zIw, then xPz if and only if yPw.

This together with (a) and (b) immediately yields (1).

It is now easy to prove

(2) xIy if and only if $u(x) = u(y)$, for all lotteries x and y.

For if xIy and $u(x) > u(y)$, then, by (1) xPy—a contradiction; so if xIy, then not $u(x) > u(y)$. Similarly, if xIy, then it is false that $u(y) > u(x)$. Thus if xIy, $u(x) = u(y)$. On the other hand, if $u(x) = u(y)$, we can have neither xPy nor yPx without contradicting (1) and the ordering condition. This means that if $u(x) = u(y)$, then xIy, since the ordering condition implies that either xIy, xPy, or yPx.

The rest of this part of our proof will be concerned with showing that for all lotteries x and y,

(3) $u(L(a, x, y)) = au(x) + (1 - a)u(y)$.

To prove this, however, it will be convenient for us to first prove that the following condition follows from the others.

Substitution-of-Lotteries Condition: If $x\ I\ L(a, y, z)$, then both
(a) $L(c, x, v)\ I\ L(c, L(a, y, z), v)$
(b) $L(c, v, x)\ I\ L(c, v, L(a, y, z))$.

This condition states that if the agent is indifferent between a prize (lottery) x and some lottery $L(a, y, z)$, the agent is also indifferent to substituting the lottery $L(a, y, z)$ for x as a prize in another lottery.

Turning to the derivation of the substitution condition, let us abbreviate "$L(a, y, z)$" by "L". Now assume that xIL. We want to show that (a) and (b) hold. I will derive (a) and leave (b) as an exercise. By the ordering condition, $L(c, x, v)$ is indifferent to $L(c, L, v)$ or one is preferred to the other. If $L(c, x, v)$ is preferred to $L(c, L, v)$, then by the better-prizes condition, xPL. But that contradicts xIL. Similarly, if $L(b, L, w)$ is preferred to $L(b, x, w)$, then LPx—again contradicting xIL. So the only alternative left is that the two lotteries are indifferent. This establishes (a).

With the substitution condition in hand we can now turn to the proof of (3). Let us use "L" this time to abbreviate "$L(a, x, y)$." By $(*)$ we have

i) $L\ I\ L(u(L), B, W)$
ii) $x\ I\ L(u(x), B, W)$
iii) $y\ I\ L(u(y), B, W)$.

Thus by substituting $L(u(x), B, W)$ for x and $L(u(y), B, W)$ for y in the lottery L and applying the substitution condition, we obtain

$L\ I\ L(a, L(u(x), B, W), L(u(y), B, W))$.

We can reduce the compound lottery on the right to a simple lottery, thus obtaining

$$L(a, L(u(x), B, W), L(u(y), B, W)) \; I \; L(d, B, W),$$

where $d = au(x) + (1 - a)u(y)$. But then by the ordering condition we must have

$$L \; I \; L(d, B, W),$$

which together with (i) yields

$$L(u(L), B, W) \; I \; L(d, B, W).$$

If $u(L) > d$ or $d > u(L)$, we would contradict the better-chances condition. So $u(L) = d$. But this is just an abbreviated form of (3).

This completes the existence part of the expected utility theorem.

PROBLEMS

1. Using the existence part of the expected utility theorem show
   a. $L(1, x, y) \quad I \, x$
   b. $L(0, x, y) \quad I \, y$
   c. $L(a, x, y) \quad I \, L(1 - a, y, x)$
   d. $L(a, x, x) \quad I \, x$
2. Derive part (b) of the substitution-of-lotteries condition.
3. Using just the ordering condition, prove:

   If xIy and zIw, then xPz if and only if yPw.
4. In this and the following exercises *do not* appeal to the expected utility theorem. Instead reason directly from the rationality conditions of the theorem.
   a. There is no number a or basic prize x distinct from B for which $L(a, x, B) \; P \; L(a, B, B)$ or $L(a, B, x) \; P \; L(a, B, B)$.
   b. There is no number a or basic prizes x and y distinct from B for which $L(a, x, y) \; P \; L(a, B, B)$.
5. Define the degree of a lottery as follows:

   All basic prizes are of degree zero.

   Let n be the maximum of the degrees of L_1 and L_2, then the degree of $L(a, L_1, L_2)$ is equal to $n + 1$.

   Suppose no lottery of degree less than n is preferred to B. Show that

   There is no number a and lottery L of degree less than n for which $L(a, B, L) \; P \; L(a, B, B)$ or $L(a, L, B) \; P \; L(a, B, B)$.

   There is no number a and no lotteries L_1 and L_2 of degree less than n for which $L(a, L_1, L_2) \; P \; L(a, B, B)$.

   It follows from this and the previous exercise that no lottery of degree greater than 0 is preferred to $L(a, B, B)$ for any number a.
6. Show that for no number a, $L(a, B, B) \; P \; B$. Hint: Apply the conclusion of exercise 5 to $L(a, L(a, B, B), L(a, B, B))$.
7. Show that for no number a, $B \; P \; L(a, B, B)$.

   Exercises 5–7 establish that no lottery is preferred to B. We can similarly show that no lottery is less preferred than W.
8. Show that if BPx and xPW, then $L(a, x, x) \; I \; x$ for any number a.

Let us turn now to the uniqueness part of the proof. Now our task is to show that if u' is a utility function defined for the same preferences as u that satisfies

(1) $u'(x) > u'(y)$ if and only if xPy,
(2) $u'(x) = u'(y)$ if and only if xIy,
(3) $u'[L(a, x, y)] = au'(x) + (1-a)u'(y)$,

then there are numbers c and d with $c > 0$ such that

$u'(x) = cu(x) + d$.

Since the two functions u and u' give rise to utility scales for the same preference ordering, we can picture the situation as follows (figure 4-4). (We

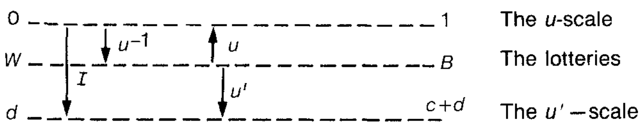

Figure 4-4

know that the end points of the u'-scale will be d and $c+d$ if our proof is correct, since

$cu(W) + d = c0 + d = d$,
$cu(B) + d = c1 + d = c + d$.)

The function u assigns utilities on the u-scale, the function u' assigns them on the u'-scale, and the function I converts assignments on the u-scale into assignments on the u'-scale. Given a number e on the u-scale, the function I first selects a lottery L for which $u(L) = e$ [this is $u^{-1}(e)$], then I applies the function u' to L to find the number f [$= u'(L)$]. In short, we have

(a) $I(e) = u'[u^{-1}(e)] = f$.

Now let k and m be any numbers on the u-scale. Note that for any number a, such that $0 \le a \le 1$, the number $ak + (1-a)m$ is between k and m or one of them. So the number $ak + (1-a)m$ is also on the u-scale. Substituting this number in (a), we obtain

(b) $I[ak + (1-a)]m = u'\{u^{-1}[ak + (1-a)m]\}$.

But $u^{-1}[ak + (1-a)m]$ is a lottery whose utility on the u-scale is $ak + (1-a)m$. Since k and m are also on the u-scale, they are utilities of some lotteries x and y; that is, $u(x) = k$ and $u(y) = m$. But then by (3), the expected utility condition, we have

(c) $u[L(a, x, y)] = au(x) + (1-a)u(y) = ak + (1-a)m$.

From which it follows that

(d) $I[ak + (1-a)m] = u'[L(a, x, y)]$.

Since u' also satisfies the expected utility condition we have

(e) $u'[L(a, x, y)] = au'(x) + (1-a)u'(y)$,

and since $u(x) = k$ and $u(y) = m$ we must have

(f) $I(k) = u'(x)$ and $I(m) = u'(y)$.

Putting these in (e) and (d) we obtain

(g) $I[ak + (1-a)m] = aI(k) + (1-a)I(m)$.

With (g) in hand (which tells us that I mimicks the expected utility property) we can complete the proof. Since each number k on the u-scale is $u(x)$ for some lottery x, we have

(h) $I[u(x)] = u'(x)$.

But by simple algebra

(i) $u(x) = u(x)1 + [1-u(x)]0$

Thus by (g), (h), and (j)

(j) $u'(x) = I[u(x)] = I\{u(x)1 + [1-u(x)]0\}$
$= u(x)I(1) + [1-u(x)]I(0)$
$= u(x)[I(1) - I(0)] + I(0)$.

Thus by setting

(k) $c = I(1) - I(0)$ and $d = I(0)$

and substituting in (j), we have

$u'(x) = cu(x) + d$.

To finish our proof we need only show that $c > 0$. That is left as an exercise.

PROBLEMS

1. Prove that c as defined in (k) above is greater than zero.
2. Prove that given any number k on the u-scale, there is some lottery x for which $u(x) = k$.
3. Show how to transform a 0 to 1 scale into a 1 to 100 scale using a positive linear transformation. Similarly, show how to transform a -5 to $+5$ scale into a 0 to 1 scale.
4. If we measure an agent's preferences on a Von Neumann-Morgenstern utility scale, does it make sense to say that the agent prefers a given prize twice as much as another?

4-3a. Some Comments and Qualifications on the Expected Utility Theorem

Now that we have concluded the proof of the expected utility theorem, let us reflect on what it has accomplished for us. The theorem is a *representation theorem*; that is, it shows that a certain nonnumerical structure can be represented numerically. Specifically, it tells us that if an agent's preferences have a sufficiently rich structure, that structure can be represented numerically by means of an interval utility function having the expected utility property. We proved the theorem by assigning numbers to each prize and lottery and then verifying that the resulting numerical scale had the desired properties. However, if the agent's preferences had failed to satisfy any one of the conditions of the theorem, then our construction would have failed to have the desired properties. For example, without the continuity condition we could not be assured of a numerical assignment for each lottery or prize, and without the reduction-of-compound-lotteries condition we could not have established the expected utility property. In a sense, then, the theorem merely takes information already present in facts about the agent's preferences and reformulates it in more convenient numerical terms. It is essential to keep this in mind when applying the theorem and discussing its philosophical ramifications.

How might we apply the theorem? Recall that we needed an interval utility scale for use with the rule of maximizing expected utility. Monetary scales proved unsatisfactory, because monetary values sometimes part company with our true preferences and because EMVs cannot ground our one-time decisions.

By contrast, utility scales do assign "true values" in the sense that utilities march along with an agent's preferences. Furthermore, each act in a decision under risk is itself a lottery involving one or more of the outcomes of the decision. Thus we can expect our agent to rank all the acts open to him along with all the prizes and lotteries. When he applies the rule of maximizing expected utility, he chooses an act whose expected utility is maximal among his options. But the utility of that act, since it is a lottery, is equal to its expected utility. Thus the agent chooses an act whose utility is highest. If there were an act he preferred to that one, it would have a higher utility. Hence in picking this act, the agent is simply taking his most preferred option. This is true even in the case of a one-shot decision. So we now know what justifies the use of expected utilities in making decisions under risk—in particular one-time decisions. It is this: In choosing an act whose expected utility is maximal an agent is simply doing what he wants to do!

Closer reflection on these facts about the theorem may cause you to wonder how it can have any use at all. For the theorem can be applied only to those agents with a sufficiently rich preference structure; and if they have such a structure, they will not need utility theory—because they will already prefer what it would advise them to prefer.

Still, decision theory can be useful to us mortals. Although the agents of the theorem are ideal and hypothetical beings, we can use them as guides for our own decision making. For example, although we may find that (unlike the ideal agents) we must calculate the expected utility of an act before we can rank it, this still does not prevent us from ranking one act above another if its expected utility is higher. We also can try to bring our preferences into conformity with the conditions of the expected utility theorem. In practice we might construct our personal utility functions by setting utilities for some reasonably small number of alternatives and then obtain a tentative utility function from these points by extrapolation and curve fitting. This tentative function can be modified by checking its fit with additional alternatives and a new function can be projected from the results, and so on, until we obtain a function satisfactory for our current purposes.

Furthermore, utility functions are useful even for ideally rational agents—for the same reason that arabic numerals are preferable to roman numerals. Utility functions facilitate the manipulation of information concerning preferences—even if the manipulator is an ideally rational being.

But, whether they are our own or those of ideally rational beings, we must approach such manipulations with caution. Suppose an agent assigns a utility of 2 to having a dish of ice cream. Can we conclude that the agent will assign a utility of 4 to having two dishes? No, utility is not an additive quantity; that is, there is no general way of combining prizes with the result that the utility of the combination equals the sum of the utilities of the components. As a result, it does not make sense to add, subtract, multiply, or divide utilities. In particular, we have no license to conclude that two dishes of ice cream will be worth twice the utility of one to our agent. If eating the second dish would violate his diet, then having two dishes might even be worth less to him than having one.

It would also be fallacious to conclude, for example, that something assigned a utility of 2 on a given scale is twice as preferable to something assigned a 1 on the same scale. For suppose that the original scale is a 1 to 10 scale. If we transform it to a 1 to 91 scale by the permissible transformation of multiplying every number by 10 and then subtracting 9, the item originally assigned 1 will continue to be assigned 1 but the one assigned 2 will be assigned 11. Thus its being assigned twice the utility on the first scale is simply an artifice of the scale and not a scale-invariant property of the agent's preferences.

As a general rule, we must be cautious about projecting properties of utility numbers onto agents preferences. A utility scale is only a numerical representation of the latter. Consequently, agents have no preferences *because* of the properties of their utility scales; rather their utility scales have some of their properties because the agents have the preferences they have.

PROBLEMS

1. Suppose that yesterday the highest temperature in New York was 40 degrees Fahrenheit whereas in Miami it was 80 degrees Fahrenheit. Would it be correct to say that Miami was twice as hot as New York?
2. To graph utility against money, we represent amounts of money on the x-axis and utilities for amounts of money on the y-axis and draw utility graphs in the usual way. One utility graph for an EMVer is the straight line given by the equation $y = x$. This graphs the function $u(x) = x$. All the other utility functions of the EMVer are positive linear transformations of this one. Describe their graphs.
3. Suppose you have an aversion to monetary risks; that is, you prefer having an amount of money for certain to having a lottery whose EMV is that amount. What does your utility graph for money look like in comparison to the graph $y = x$?

4. Suppose you welcome monetary risks in the sense that you would rather take a gamble whose EMV was a certain amount than have that amount for certain. Now what does your utility graph look like?
5. What could we conclude about the preferences of someone whose utility function for money was given by

$$u(x) = x^2$$

in the \$0 to \$100,000 range and by

$$u(x) = (100{,}000)^2$$

for amounts greater than or equal to \$100,000?

4-4. Criticisms of Utility Theory

A number of thinkers have criticized utility theory and the expected utility theorem. I will begin by discussing some criticisms of the better-chances condition, the reduction-of-compound-lotteries condition, and the continuity condition. Then I will turn to three paradoxes that have been used to criticize the theory as a whole.

Some have objected to the better-chances condition as a requirement of rationality on the grounds that if you prefer life to death, the condition requires you to prefer any act that does not increase your risk of death to one that does—no matter how small the increase might be. But such preferences could lead to total and irrational paralysis. You would not even get out of bed in the morning for fear of dying. The answer to this criticism is that the better-chances condition has no such implications. When, for instance, I compare an afternoon in front of the TV with one hang gliding, I am not comparing one pure life-and-death lottery with another. I am comparing a safe but *dull* afternoon with a dangerous but *thrilling* one. For the better-chances condition to apply the two lotteries must involve the same outcomes, and here they plainly do not. I might die during the afternoon while I watch TV just as I might die while hang gliding. But dying while watching TV is surely a different and less noble outcome than dying while hang gliding. (You might think it is irrational to go hang gliding, or, if you prefer, to play Russian roulette, because the chances of dying are so great. Remember that decision theory concerns itself only with the form of an agent's preferences and not with their specifics. Decision theory will not tell you not to prefer Russian roulette to ordinary roulette, but it will tell you that if you prefer one to the other, you cannot also be indifferent between them.)

The standard objection to the reduction-of-compound-lotteries condition is not so easily dismissed. Since utility theory forces an agent to regard compound lotteries as indifferent to certain simple ones, it (or so the objection goes) abstracts from the pleasures and anxieties associated with gambling. The avid gambler will not be indifferent between a single-stage lottery and a multistage one, whereas someone who regards gambling as wrong, will want to do it as little as possible. Utility theory, with its reduction-of-lotteries condition, has no place for these preferences and aversions.

To respond to this criticism we must distinguish between those multistage lotteries that are simply theoretical idealizations introduced for calibrating utility scales and real-life multistage lotteries that may take weeks or months to be completed. We can think of theoretical lotteries as being run and paid off in an instant. Given this it would not make sense for an agent to distinguish between having, say, a \$100 ticket to an instantaneous simple lottery ticket paying \$100 unconditionally and an instantaneous multistage lottery with the same outcome. In this case we need not take the objection seriously.

On the other hand, if a multistage lottery takes place over time, we cannot use this reply. Suppose, for example, that in situation A an agent is faced with the prospect of a 25% chance of selling her house today and a 75% chance of never selling it. In situation *B*, the same agent is faced with a 50% chance of agreeing today on an option to sell the house within the week and a 50% chance of never selling it. However, if the option is taken, the agent faces a 50% chance that the prospective buyer will obtain a loan and buy the house within the week and a 50% chance that the deal will collapse and the house will never be sold. Using lottery trees we might represent the two situations as follows (figure 4-5).

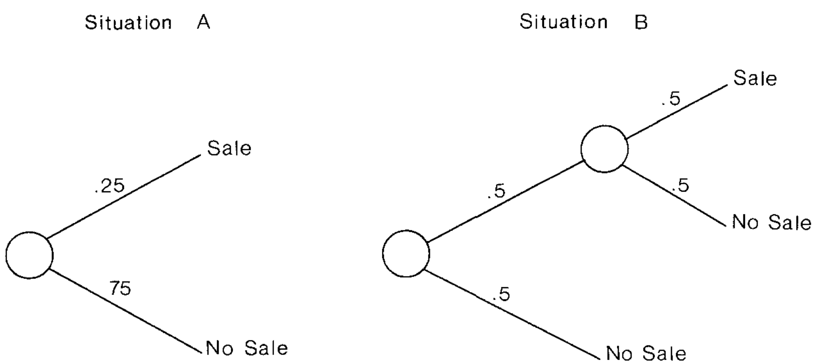

Figure 4-5

If the outcomes are as these trees represent them, the compound lottery in B reduces to the simple lottery in A in accordance with the reduction-of-compound-lotteries condition. That would constitute a genuine objection to a theory since an agent who sorely needs to sell her house today would rightly prefer A to B. But are the outcomes really as the trees represent them? I think not, for in situation B the sale is not made today but within the week. Someone who needs his money now or dislikes the suspense of waiting as much as one week will not view the outcomes in A and B as the same. Thus the compound lottery in B and the simple one in A do not satisfy the antecedent of the reduction condition and, accordingly, do not qualify as counterexamples. I would expect that other "over time" counterexamples could be dissolved similarly.

The objection to the continuity condition is similar to that to the better-chances condition. Suppose I prefer living another ten years to receiving a gift of one cent right now to dying immediately. Then the continuity condition requires me to be indifferent between receiving the penny right now and accepting a lottery that gives me some positive chance of dying on the spot. One might object that it is irrational for me to risk an additional chance of dying for a mere penny. But note that the continuity condition requires only that the chance be positive – not that it constitute a significant increase in my chances of dying. So all I am required to do is to be indifferent between receiving the penny and positively increasing my chances of dying by *as little as I want*. There need be nothing irrational in that at all.

Although most decision theorists do not take the objection to the continuity axiom seriously, some have developed utility theories that do not use the continuity condition. These are multidimensional or multiattribute utility theories in which outcomes are evaluated along several dimensions (or with respect to several attributes). For example, in choosing between having and not having an operation to relieve chronic back pain, a person could consider several attributes of each of the outcomes, such as the risk of death, the attendant pain, the length of the recovery period, the costs, and so on. It may well be that this person can order the outcomes in terms of priorities – for example, minimizing disability might have the highest priority, minimizing long-term pain the next, minimizing the risk of death the next, and so on. Yet this person might be unable or unwilling to make "trade-offs," for example, to trade more disability for less pain, and thus unable or unwilling to combine the many dimensions into one. This would violate the continuity condition since that condition requires all the outcomes to be ranked on the same numerical scale. But this does not seem irrational in itself. A properly constructed multidimensional utility theory should be able to deal with this question. But I will not be able to explore this matter further here.

4-4a. Allais's Paradox

I will now turn to three "paradoxes" directed at utility theory as a whole. The first, proposed by the contemporary French economist Maurice Allais, presents us with two situations. In situation A we offer an agent a choice between receiving $1,000,000 for certain and a lottery that furnishes him a .1 chance of winning $5,000,000, a .89 chance of winning $1,000,000 and a .01 chance of receiving nothing at all. In situation B the choice is between two lotteries. One offers a .1 chance at $5,000,000 and a .9 chance at nothing, the other offers a .11 chance of $1,000,000 and a .89 chance of nothing. Figure 4-6 presents the decision trees for these two situations. The paradox is this. Many presumably rational and reflective people find that they would prefer *a* (the bird in the hand) to *b* if they were in situation A and would prefer *c* to *d* in situation B. But no matter what your utility for money is, the choices of *a* in A and *c* in B, or *b*

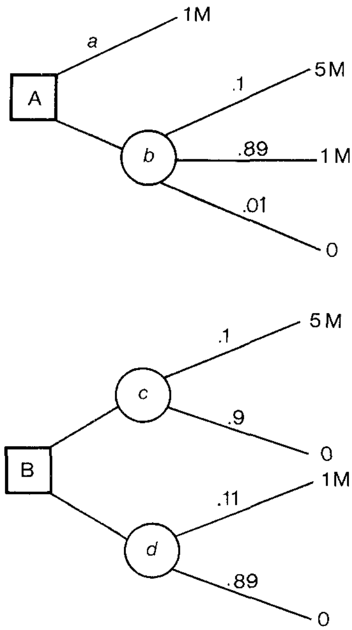

Figure 4-6

in A and d in B contradict utility theory. To see how, let us compute the utilities for a, b, c, and d. They are:

$$u(a) = u(1\text{M})$$
$$u(b) = .1u(5\text{M}) + .89u(1\text{M}) + .01u(0)$$
$$u(c) = .1u(5\text{M}) + .9u(0)$$
$$u(d) = .11u(1\text{M}) + .89u(0).$$

But it then follows by simple arithmetic that

$$u(a) - u(b) = .11u(1\text{M}) - [.1u(5\text{M}) + .01u(0)]$$
$$u(d) - u(c) = .11u(1\text{M}) - [.1u(5\text{M}) + .01u(0)].$$

Now if you prefer a to b, then $u(a) > u(b)$, so $u(a) - u(b) > 0$. But then $u(d) > u(c)$, so you must prefer d to c to conform to utility theory. A similar argument shows that if you prefer b to a, you ought to prefer c to d. But, of course, many people balk at these recommendations.

Several resolutions of this paradox have been advanced and each has been criticized. Two deserve examination here. The first follows the line I used in responding to the counterexamples to the better-chances and reduction conditions. It consists in arguing that the formal representation of situations A and B is incorrect and that, consequently, nobody who chooses a in A and c in B contravenes utility theory. The argument turns on the claim that anyone who chooses b in situation A and ends up with nothing has done so *when he could have had $1,000,000 for certain.* No one would be passing up a certain $1,000,000 in situation B. Thus we have misrepresented the two situations by using 0 in both to stand for the outcome in which the agent receives nothing. The outcome designated as 0 in situation A presumably has a lower utility than its correspondent in B. Once this is acknowledged the rest of the argument breaks down, since we can no longer count on the truth of $u(a) - u(b) = u(d) - u(c)$.

Another resolution, proposed by the statistician Leonard Savage, tries to persuade us that the reasoning that leads people to choose *a* in A and *c* in B is fallacious. Savage asks us to represent the two situations as involving choices between lotteries with 100 tickets where the payoffs are given according to table 4-5. (Here is how row *b* is obtained: the .01 chance of 0 is converted to a payoff of 0 for 1 ticket out of 100 [ticket number 1]; the chance of .1 at $5,000,000 is converted to a payoff of $5,000,000 for 10 tickets out of 100 [tickets 2-11], and the .89 chance at $1,000,000 is converted to a payoff of $1,000,000 for 89 tickets of the 100. The rest of table 4-5 is obtained similarly.)

4-5

		Ticket Number 1	2-11	12-100
A	*a*	1M	1M	1M
	b	0	5M	1M
B	*c*	0	5M	0
	d	1M	1M	0

Now in choosing lotteries in either situation tickets 12–100 can be ignored, because they give the same prizes in their respective situations. Thus the decisions between *a* and *b* and between *c* and *d* should be based on the first two columns of the table. However, the first two columns of *a* and *d* are identical and so are those of *b* and *c*. Accordingly, it is obviously mistaken to choose *a* in A and not *d* in B or *b* in A and not *c* in B.

In short, Savage claims that those who succumb to the Allais paradox have simply made a mistake in reasoning. This is no different, he urges, than the mistakes otherwise intelligent people often make in estimating probabilities or in carrying out complicated mathematical arguments. In each case all we need do to correct them is to point out their errors clearly and perspicuously. I will let you decide whether you find this or the previous resolution of the Allais paradox convincing.

4-4b. Ellsberg's Paradox

The Allais paradox derives its force from our tendency to prefer having a good for certain to having a chance at a greater good. The next paradox, developed by Daniel Ellsberg, appeals to our preference for known risks over unknown ones.

Here is a version of the Ellsberg paradox that is quite similar to Allais's paradox. An urn contains ninety uniformly sized balls, which are randomly distributed. Thirty of the balls are yellow, the remaining sixty are red or blue. We are not told how many red (blue) balls are in the urn—except that they number anywhere from zero to sixty. Now consider the following pair of situations. In each situation a ball will be drawn and we will be offered a bet on its color. In situation A we will choose between betting that it is yellow or that it is red. In situation B we will choose between betting that it is red or blue or that it is yellow or blue. The payoffs are given in table 4-6.

4-6		Yellow	Red	Blue
		30	60	
A	*a*	100	0	0
	b	0	100	0
B	*c*	0	100	100
	d	100	0	100

When confronted with this decision people frequently reason as follows. In situation A, I will pick bet *a* (yellow), since the chance of the ball's being red might be quite small. But in situation B, I will pick bet *c* (red or blue). That gives me sixty chances out of ninety at winning, whereas if I took bet *d* (yellow or red), I might have only thirty out of ninety chances.

Now suppose that we reasoned this way. Then our choices could not consistently reflect the expected utilities of the bets. For let p be the probability of getting a blue ball given that the ball is red or blue and let A be the utility of 100 and B the utility of 0. Then the expected utilities are:

$$
\begin{aligned}
EU(a) &= (1/3)A + (2/3)pB + (2/3)(1-p)B = (1/3)A + (2/3)B \\
EU(b) &= (1/3)B + (2/3)pA + (2/3)(1-p)B = B + (2/3)p(A-B) \\
EU(c) &= (1/3)B + (2/3)pA + (2/3)(1-p)A = (1/3)B + (2/3)A \\
EU(d) &= (1/3)A + (2/3)pB + (2/3)(1-p)A = A + (2/3)p(B-A).
\end{aligned}
$$

Now subtract $EU(b)$ from $EU(a)$ and $EU(d)$ from $EU(c)$. You will find that

$$EU(a) - EU(b) = -[EU(c) - EU(d)].$$

This means that our preferences will accord with expected utilities if and only if our preferences for *a* and *b* are the reverse of those for *c* and *d*. So we cannot pick *a* in situation A and *c* in situation B without contravening utility theory.

I will present two responses to the Ellsberg paradox, although, as with the paradox by Allais, there have been numerous discussions of it. The first response mimics Savage's resolution of the Allais paradox. Notice that the figures in the third column for a and b in table 4-6 are identical and so are those for c and d. Hence, or so Savage would claim, the third column can be ignored in making our decisions. If we do that, a and d are equivalent and b and c are equivalent. It follows immediately that we should prefer a to b just in case we prefer d to c.

Furthermore, there is a mistake in the reasoning that leads people to choose a and c. They say that they will pick a over b because the probability of getting a red ball might be quite small. In fact, it has to be less than 1/3 for the choice of a to be better than that of b. This would mean that the probability of getting a blue ball would be 1/3 or greater. This is inconsistent with the reasoning they give for choosing c over d. For that choice makes sense only if the probability of getting a blue ball is less than 1/3.

The other resolution consists in noting that we do not know the probability p used in comparing the expected utilities of the four bets. Thus these are decisions under ignorance rather than under risk, so the rule of maximizing expected utilities does not apply.

If you accept the subjective theory of probability, however, this alternative is not open to you. For the mere fact that nobody has told us the ratio of red to blue balls will not prevent us from assigning a subjective probability to the color of the ball that is drawn. This being so, choosing bets a and c would contradict the combined theories of utility and subjective probability. Perhaps this speaks against the combined theories, but those who believe the combined approach is the correct one could make a virtue out of this vice. They could argue that the combined theories present simultaneous constraints on our assignments of probabilities and our preference orderings. Obeying them is a difficult and complex task, so we should expect that from time to time our initial hunches will need to be corrected by the theory. To take a well-known example from probability theory, since one spin of a roulette wheel is independent of any other, it would be wrong for you to conclude after a long losing streak at the roulette table that you were almost certain to win on the next spin. Drawing such a conclusion is called the gambler's fallacy. Having been educated in the ways of probability, we can appreciate why it is a fallacy, but I am sure that each of us either has committed it or has been sorely tempted to do so. Someone like Savage would urge that the Allais and Ellsberg paradoxes play a similar role in our intellectual life. Those of us uneducated in decision theory react as Allais and Ellsberg tempt us to react. But once we have learned the theory and have correctly reflected on its morals, the paradoxes join the ranks of old mistakes along with the gambler's fallacy.

4-4c. The St. Petersburg Paradox

The next paradox, known as the St. Petersburg paradox, does not apply to the version of utility theory presented earlier, because it depends on the assumption that there is no upper bound to the agent's utility scale. This happens when for any basic prize there is always a better one. The purpose of the paradox is to show that under this assumption the agent will assign something an infinite utility—even if he assigns only finite utilities to the basic prizes.

Before passing to the derivation of the paradox let me note that unbounded utility scales can be constructed along the same lines as we used in constructing bounded utility scales. Although the construction is mathematically more complicated, it requires no further conditions on the rationality of the agent's preferences.

The St. Petersburg paradox is based on the St. Petersburg game, which is played as follows. A fair coin is tossed until it lands heads up. If the first toss is heads, then the player is given a prize worth 2 on his utility scale (2 utiles, for short). If it lands heads on the second toss, he is given a prize worth 4 utiles, and so on, with heads on the nth toss paying 2^n utiles. The paradox is simply that playing the St. Petersburg game has an infinite utility to the agent. Thus he should be willing to surrender any basic prize as the price of admission to the game.

To see why, note that the probability of heads on the first toss is 1/2, that of heads on the second but not the first is 1/4, . . . , that on the nth but not on the first $n-1$ tosses is $1/2^n$. Thus the expected utility of the game is

$$(1/2)2 + (1/4)4 + \ldots (1/2^n)(2^n) + \ldots = 1 + 1 + 1 + \ldots$$

which is larger than any finite number. And utility theory tells us that the expected utility of the game is its utility.

One might object that we have not made sense of infinite lotteries in our development of utility theory and that the St. Petersburg game is essentially an infinite lottery. True, but many decision theorists would not be moved by this. For utility theory is easily extended to cover infinite lotteries, and it must be in order to handle more advanced problems in statistical decision theory. From the mathematical and logical point of view, the derivation of the St. Peterburg paradox is impeccable.

The only alternative then, short of modifying utility theory, is to question the assumption on which it is built—that is, that the agent's utility scale is unbounded. Now it must be admitted that this assumption simplifies the mathematical treatment of many advanced applications of utility theory. Also there is nothing irrational on the face of it in having an unbounded utility scale. So responding to the paradox by prohibiting such scales is not an attractive option. My inclination is to see the paradox as simply showing us the music agents must face if they do not bound their preferences. Although I see nothing irrational in unbounded preferences per se, the St. Petersburg paradox favors avoiding them.

On the other hand, I see no reason for the St. Petersburg paradox to arise in practice. No one is insatiable; there can be too much of anything—even life, money, and power. When the saturation point with respect to a "commodity" is reached there is a disutility to having more of it. (What would you do with all the money in the world? What value would it have if you had it all?) If this is correct, the paradox is based on a theoretically possible but unrealistic assumption. On the other hand, in a theory of ideal rational beings theoretical possibilities cannot be treated lightly. Some decision theorists have taken the paradox quite seriously and have proposed modifications of utility theory to avoid it.

PROBLEMS

1. Ellsberg presented another paradox. If we offer people a bet on a fair coin, most are indifferent between betting on heads and betting on tails. They are similarly indifferent when offered a bet on biased coins when they are not told the bias. However, when offered the choice

   bet heads on the fair coin vs. bet tails on the biased one

   they prefer to bet on the fair coin. Suppose that whichever coin is involved in a winning bet pays $1, whereas a losing one pays nothing. Show that the choices most people make contradict the combined theories of utility and subjective probability.
2. Of course, if you are a fan of utility theory and have reservations about subjective probabilities, the Ellsberg paradox confirms your suspicions. Explain why.
3. In both the Allais and Ellsberg paradoxes we have appealed to what "most people" think or to what "many apparently rational and thoughtful people" think. What bearing, if any, does this have on decision theory, construed as a theory of ideally rational beings?
4. Utility theory originated in attempts to deal with a version of the St. Petersburg game played with monetary prizes. In 1730, Daniel Bernoulli proposed that monetary values be replaced with their logarithms. If we use logarithms to the base 10, then the expected value of this version of the St. Petersburg game is

   $$(1/2)\log_{10}(2) + (1/4)\log_{10}(4) + (1/8)\log_{10}(8) + \ldots ,$$

   which converges to a finite sum. Notice, however, that $\log_{10}(10) = 1$, $\log_{10}(100) = 2$, $\log_{10}(1{,}000) = 3$, etc., so that we can reinstate the St. Petersburg paradox by increasing the payoffs. Give a detailed explanation of how this can be done.

4-5. The Predictor Paradox

Here is a paradox, introduced by a physicist named Newcomb, which has generated more controversy than all the others combined.

Suppose a being with miraculous powers of foresight, called *the Predictor*, offers you the following choice. There are two shoe boxes on the table between the two of you. One is red, the other blue. The tops are off the boxes and you can see that the red box has $1,000 in it and that the blue box is empty. The Predictor tells you that you are to leave the room and return in five minutes. When you return you will be offered the choice of having the contents of the blue box or that of both boxes. The Predictor also tells you that while you are out of the room he will place $1,000,000 in the blue box if he predicts that you will take just the blue box, and he will leave it empty if he predicts that you will take both boxes. You leave the room and return after five minutes. The top is still off the red box and the $1,000 is still in it. The blue box is now closed. You have heard about the Predictor before. You know that almost everyone who has taken both boxes has received just $1,000 while almost everyone who has taken just the blue box has received $1,000,000. What is your choice?

Suppose you use decision theory to help with your choice and your utility for money is equal to its EMV. Then you will have decision table 4-7. The

4-7	Blue Box Empty	Not Empty
Take Blue Box	0 .1	1M .9
Take Both Boxes	$1,000 .9	1M + $1,000 .1

probabilities are determined as follows: You reckon that if you take just the blue box there is a 90% chance that the Predictor will have predicted that you would and will have put $1,000,000 in that box. You also think that there is only a 10% chance that he will have predicted incorrectly and left the box empty. Similarly you reckon that if you take both boxes, there is a 90% chance that he will have predicted that and left the blue box empty. It is easy to see that, given this problem specification, you will maximize expected utility by taking just the blue box. Recalling all your rich friends who took just the blue box, you are about to reach for it.

But doubt strikes. You notice that the dominant choice is to take both boxes. What then should you do? Decision theory's two most venerable rules offer you conflicting prescriptions!

Then you remember that the dominance principle is supposed to yield to the principle of maximizing expected utility when the probabilities of one or more states are dependent on the choices made. This made good sense in the nuclear disarmament example because disarming made an attack by the other side highly likely. But does it make good sense in this case? You argue to yourself: The $1,000,000 is already in the blue box or it is not. My choice cannot change things now. If there is $1,000,000 in the blue box, then taking both will yield $1,001,000 which is better than taking just the blue box. If the $1,000,000 is not there, I can get $1,000 by taking both boxes, and that's better than taking just the blue box and getting nothing. So in either case I am better off taking both boxes.

Yet next you remember that your friends who reasoned in this way are still poor whereas those who took the blue box are rich. But that does not seem to make sense either. Can the Predictor exercise mind control and influence your choice, thereby making his prediction come true? But how? He has not even told you what his prediction is. Or can your choice somehow influence his prediction? But again how? You choose *after* he predicts.

The dilemma is this. If we use the dominance principle to decide, we must ignore all the empirical data pointing to the folly of choosing both boxes; but if we follow the data and maximize expected utility, we are at a loss to explain why the data are relevant. So what is your choice?

If you do not know what to do, take comfort: You are not alone. Many decision theorists think you should take both boxes but many others think you should take just the blue one. Moreover, this paradox has shaken decision theory to its foundations.

Before we consider some of the growing number of responses to the Predictor paradox, let us bring it down to Earth. (Not that this fanciful example is not useful, for, even if it might not arise in practice, it points to a conceptual weakness in decision theory.) Here are two real-life problems closely related to the Predictor paradox.

The first is derived from the religious views of the sixteenth-century Protestant theologian John Calvin. According to one interpretation of Calvin's views, God has already determined who will go to Heaven and nothing we do during our lifetimes – whether we be saints or sinners – will change this. On the other hand, although no mortal can know whom God has chosen, Calvinists believe that there is a high correlation between being chosen and being devout. Now suppose you are a dyed-in-the-wool Calvinist and are sorely tempted to commit some sin. You consider two arguments. According to the first you should pick the dominant act. Go ahead and sin, since that is not going to affect your going to Heaven anyway, and an earthly life with pleasure is better than one of Calvinist abstinence. But the other argument tells you that you should restrain yourself. As a devout Calvinist, you are certain that your sinning would be a clear sign that you are not among the elect.

Now for a contemporary example. Smoking cigarettes is highly correlated with heart disease. Smokers are much more likely to have heart troubles than nonsmokers and heavy smokers are at greater risk than light ones. Heart disease also seems to run in families and to be associated with ambitious, serious, and competitive individuals, known as type A personalities. Today most people researching the causes of heart disease regard smoking, genetic predisposition, and the type A personality as independent contributing causes of heart disease. However, suppose the causal story went as follows instead: Type A personality is inherited (just as breeding has made race-horses high strung and draft horses calm), and people with type A personalities have a greater need to smoke than do other people. However, smoking does not cause heart disease, rather the cause is the same genetic defect that produces the type A personality. Thus the cause of smoking and heart disease is at bottom one and the same – a genetic defect. Let us suppose that the same story held for the other diseases – for example, cancer, emphysema – in which smoking has been implicated. If you believed this story and were trying to decide whether to smoke, you might be in a dilemma similar to the Predictor paradox. For the high correlation between smoking and various diseases bestows a higher expected utility on not smoking. But you would either have a disease-causing genetic defect or not. If you had it, smoking would not increase your risk of disease, and it would (or so I will suppose) lead to a more pleasurable life than would abstaining from smoking. So dominance considerations would argue for smoking. Once again the rule of maximizing expected utility and the dominance principle would conflict.

PROBLEMS

1. I formulated the Predictor paradox using .9 as the probability of the correct predictions, but it need not be that high. Assuming, as indicated in table 4-7, that $P(S_1/A_1) = P(S_2/A_2)$, how low can it get subject to the condition that taking the blue box remains the act with greatest expected utility?
2. Set up a decision table for the Calvinist example that will establish it as a variant of the Predictor paradox.
3. Do the same for the smoking example.

4-6. Causal Decision Theory

The Predictor paradox, the Calvinist example, and our fanciful speculations on the causes of smoking-related diseases remind us that from time to time there can be a high correlation between two sorts of phenomena without any causal relation between them. Almost every time the thermometers go up in my house it is hot. The same thing happens in every other house I know. So the probability that it is hot given that the thermometer readings are high is close to 1. Does that mean that the high thermometer readings cause it to get hot? Obviously not. Everything we know about temperatures and thermometers tells us that the causal relation goes oppositely. Because of this it would be silly to try to heat a room by holding a match under the thermometer on the wall. But would it not be just as silly for someone who really believed my fantasy about the causes of heart disease to try to erase its genetic cause by not smoking? Or for my Calvinist to try to get into Heaven by being devout? Of course, the Calvinist does not know whether he is among the elect and the potential victim of heart disease might not know whether she has unlucky genes. But that does not matter since, under the terms of the stories, the woman considering smoking knows that her doing so will not alter her chances of getting heart disease and the Calvinist knows that his sin will not consign him to Hell.

These examples show that even though a state is highly probable given an act, the act itself need not be an effective method for bringing about the state. Given everything we know about causal relationships, taking just the blue box is not an effective method for ensuring that there will be $1,000,000 in it. Perhaps we were mistaken then in using conditional probabilities in setting up the decision table for the Predictor paradox. Let us try again using the unconditional probabilities of the states. Since we do not what they are, I will denote them by "p" and "$1-p$." This yields table 4-8. Then whatever the value of p, taking both boxes has the higher expected utility, which agrees with the recommendation of the dominance principle. Applying the same strategy to the smoking and Calvinist cases yields similar results. This is how *causal decision theory* solves the Predictor paradox.

4-8

	Blue Box Empty	Not Empty
Take Blue Box	0 p	1M $1-p$
Take Both Boxes	\$1,000 p	1M + \$1,000 $1-p$

However, there is more to causal decision theory than I have indicated so far. It does not advocate an unqualified return to unconditional probabilities. If you believe, as most of us do, that smoking does cause disease, then, despite smoking's dominating, causal decision theory recommends that you not smoke. Given our current beliefs about the *causal* relationship between smoking and disease, causal decision theory tells us to calculate expected utilities in the old way, that is, by using probabilities conditional on the acts of smoking and not smoking.

To have a more general formulation of causal decision theory, let us define the *conditional causal probability* of a state given an act as *the propensity for the act to produce or prevent the state.* Notice that if an act has a propensity to bring about a state, the conditional causal probability of that act will be greater than the unconditional probability of the state. If, on the other hand, the act has a propensity to prevent the state, the latter probability will be higher than the former. Finally, if an act has no propensity to affect the state, the probabilities will be the same.

Causal decision theory uses causal conditional probabilities in place of unqualified conditional probabilities. Warmer rooms do have a propensity to bring about higher thermometer readings, but the converse is not so. Thus, in applying causal decision theory, we can use the probability that a thermometer will rise given that it is located in a room that is becoming hotter. However, we must replace the converse conditional probability by the absolute probability, because the thermometer's rising has no propensity to bring about a rise in the room's temperature.

We can now see the rationale for causal decision theory's solution to the Predictor paradox. In making decisions we select acts in virtue of their power to produce the outcomes we desire (and hence the states that foster those outcomes). In view of this, it would be wrong for us to endow our decision theoretic framework with indicators of the efficacy of our acts that we know to be misleading. In the Predictor, Calvinist, and smoking examples, there is, as far as we can tell, no propensity for the acts to bring about the states. Thus standard decision theory used the wrong indicators in treating them. No wonder, then, that it produced an apparently inexplicable contradiction between the principle of maximizing expected utility and the dominance principle.

Sometimes a decision maker will be quite certain that an act will affect a state, but be uncertain about its ultimate outcomes because of other states she is unable to affect by her choices. A physician, for example, might be certain that giving an injection of penicillin is an effective way to cure a patient's infected toenail, but be uncertain whether it will cause a dangerous penicillin reaction—something over which she has no control. She should use conditional probabilities for evaluating the consequences of giving penicillin as far as the infection is concerned and unconditional probabilities for the allergic reaction. We could put this in the form of decision table 4-9 as follows. (For the sake of generating probability numbers, I will assume that the patient has a 50% chance of being allergic, that penicillin is an effective cure 75% of the time, and that the patient has a 30% of "curing himself" if no penicillin is given. Being cured of the infection is independent of having a reaction, so I will multiply the probabilities in calculating the probabilities of the various conjunctive states. I will also omit utilities.)

4-9

	Patient Allergic Infection Cured	Patient Allergic Infection Not Cured	Not Allergic Infection Cured	Not Allergic Infection Not Cured
Give Penicillin	Cure & Reaction .375	No Cure & Reaction .125	Cure & No Reaction .375	No Cure & No Reaction .125
Do not	Cure & No Reaction .15	No Cure & No Reaction .35	Cure & No Reaction .15	No Cure & No Reaction .35

Another sort of case is one in which the decision maker is unsure of the propensity of a given act to bring about a given state. Consider our physician again. Let us suppose she knows that her patient is not allergic to penicillin, but this time the patient has a staph infection. Here the physician might vacillate between thinking that penicillin has only a 50% chance of curing the patient and thinking that it has a 75% chance, while assigning a probability of 50% to each hypothesis. Suppose her other choice is to use antibiotic X in place of penicillin and that she is certain this is 60% effective. Then her decision could be represented as follows in table 4-10. (Again, I will omit utilities.)

4-10

	Pen. 50% Effective Infection Cured	Pen. 50% Effective Infection Not Cured	Pen. 75% Effective Infection Cured	Pen. 75% Effective Infection Not Cured
Give Penicillin	Cure .25	No Cure .25	Cure .375	No Cure .125
Give X	Cure .30	No Cure .20	Cure .30	No Cure .20

The general form of these two examples is this. We first set out a mutually exclusive and exhaustive list of states that represent those factors relevant to the decision that will not be affected by the acts. This may include possibilities concerning the propensity of one or more of our acts to affect other states. Then under each of these states we set up a mutually exclusive and exhaustive list of relevant states that will be affected by the act, repeating this division under each state from the first group. Using absolute probabilities for states in the first group and causal conditional probabilities for those in the second group, we calculate the probabilities for each square of decision table 4-10 by multiplying the probabilities of the two states that determine it. Expected utilities are then calculated in the usual way: Multiply each utility in a square by the corresponding probability and sum across the rows.

PROBLEMS

1. Suppose in the first penicillin example the physician's utilities are as follows: u(cure & no reaction) = 1, u(cure & reaction) = .5, u(no cure & no reaction) = .25, u(no cure & reaction) = 0. Should the physician use penicillin?
2. Suppose in the second penicillin example the physician is also uncertain about the effectiveness of antibiotic X. She thinks there is a 50% chance that it is 70% effective and a 50% chance that it is 40% effective. Set up a decision table for this example, supplying probabilities for each square.

4-6a. Objections and Alternatives

Causal decision theory is certainly a reasonable and attractive alternative to the standard theory of expected utility maximization. But I would be remiss if I did not point out that it too must deal with philosophical difficulties of its own. The most prominent of these is its appeal to causal necessities (as opposed to mere statistical regularities), which have been philosophically suspect ever since Hume's critique of causality. Hume's thesis was a simple one: What pass for causal necessities are in fact nothing but statistical regularities; for, try as we may, we will never find evidence for causal necessities but only for statistical regularities. Since Hume's claim might seem very implausible to you, let us consider an example where causal necessity seems to be plainly at work. You grab a handful of snow, pack it into a ball, and throw it at one of your friends. The snowball hits his back and splatters. On the face of it causes abound here: Your exerting force on the snow causes it to form a hard ball, your hurling it causes it to fly through the air, its impacting your friend's back causes it to splatter, and so on. Let us focus on your packing the snow, since here you seem to able to even *feel* causal powers at work. Now Hume would not deny that whenever you or anyone you know has packed snow into a ball, shortly thereafter there has been a snowball in your or his hand. He would simply deny that this gives us any evidence that you *caused* the snow to form a ball, in the sense that your action *necessitated* the observed effect. If the cause must necessarily produce the effect it is inconceivable for the cause to occur without the effect's occurring.

But we can easily conceive of cases in which we apply the "cause" to the snow and no snowball results. For example, as we pack the ball, someone might shine a heat lamp on the snow, "causing" it to melt. Now you might think that all this shows is that we can cause the snow to form a ball so long as no *interfering causes* prevent it. But given any putative list of interfering causes, it is still easy to conceive of our packing the snow and the snow's still failing to form a ball despite the absence of each of the interfering causes in our list. Besides, we have helped ourselves freely to the notion of causality in describing the example. If we take a neutral stance, we should describe ourselves as having observed that whenever we have had "snowball-packing" sensations we have found a snowball in our hands. But once we look at the example in this way, Hume's point becomes almost obvious. For we can have snowball-packing sensations without even having anything in our hands!

If Hume is correct, there are no causal powers or laws—just highly confirmed regularities—and causal decision theory is without foundation. Viewing the heart disease example from the Humean point of view, we will find no basis for distinguishing the high correlation between smoking and heart disease and between the genetic defect and heart disease. Given that you smoke, it is highly probable that you will get heart disease. Given that you have the defect, it is highly probable that you will smoke and have heart disease. But without appealing to the concept of causality (or something similar) we have no grounds for claiming that your smoking will *affect* your chances of having heart disease. Similarly, we cannot, strictly speaking, say that your taking both boxes will have no effect on whether there is $1,000,000 in the blue box; for the language of cause and effects is not available to us. We are left then with nothing but correlations. And given the correlations, we are best advised not to smoke and to take just the blue box.

There is a practical and epistemological aspect to Hume's problem too. For even those who believe in causality have been hard pressed to specify methods for distinguishing genuine causal relations from mere statistical regularities. This problem bears on the applicability of causal decision theory, since it requires us to segregate states over which we have causal control from those that are connected to our acts by mere statistical regularities.

Although these difficulties favor a more Humean approach to decision theory, causal theorists have a biting response. The point of decision theory, they argue, is to develop methods for choosing acts that yield desirable consequences. If our acts have no effect on the outcomes, what is the point of choosing? We might as well just watch whatever happens, happen. Indeed, if there are no causes, we act, if we can call it that, by being passive observers. How, then, can we continue to speak of choosing between such *acts* as, say, *replacing the window myself* or *hiring someone to do it*?

Some Humean theorists, or as they are now called, *evidential decision theorists*, although acknowledging this point, respond that the true role for decision theory is to help us determine our preferences for *news items*. Pushed to the limit, this would mean that when we decide between, say, marrying at eighteen or jilting our sweetheart and joining the Foreign Legion, we are determining whether we would prefer to learn the news that we had married at eighteen or instead that we had jilted our sweetheart and joined the Foreign Legion. Making a choice is to pick from among such news items.

As long as one is consistent about this, decision theory retains its point. Suppose that, in casual terms, I buy a brand-new Mercedes to ensure that people will think I am rich. The evidential theorist can recast this as my choosing to learn that I own a new Mercedes on the grounds that (1) I know there is a high correlation between owning a Mercedes and being thought rich and (2) I desire to learn that people think I am rich.

In the Predictor case causal decision theory advises us to take both boxes because doing so cannot affect their contents. Evidential decision theory advises us to pay attention to the evidence: There can be no evidence of causality since there is no causality; but there is plenty of evidence that after we pick just the blue box we are almost certain to learn that we are rich. So we ought to take just the blue box. And so the debate goes.

Recently, Ellery Eells has indicated a way in which one can arrive at the same recommendations as causal decision theory without having to base decision theory on the distinction between those states that are under our causal control and those that are not. However, unlike the dyed-in-the-wool Humean, Eells does assume that it makes sense to speak of causes. Consider the smoking example again. Eells describes it in causal terms as we did before: A birth defect is the common cause of both smoking and heart disease, and that is why there is a high correlation between smoking and heart disease. However, Eells parts company with the causal decision theorist at this point, for he does not simply dismiss the high correlation between smoking and heart disease on the grounds that we have no control over whether we have the genetic cause of both smoking and heart disease. Instead he argues that if you are sufficiently rational and intelligent, you will see that the high correlation in no way entails that the probability that you, in particular, will get heart disease given that you smoke is greater than the probability that you will get it given that you do not smoke. Eells bases his reasoning on the assumption that you do not know whether you have the birth defect and that you believe it is the common cause of both smoking and heart disease. Given this, you might think that you should assign a high probability to getting heart disease given that you smoke, since smoking is an indication that you have the birth defect. But Eells parries this move by introducing another assumption, namely, that you believe the common cause can cause you to decide to smoke only by causing you to have desires and beliefs that entail that smoking is an optimal act for you. Given this, you will believe that the variables determining your choice to smoke or not are fully captured in the particular beliefs and desires you have when you make the decision. Eells assumes that you know what your beliefs and desires are, although you do not know whether you have the defect leading to smoking. But this means that you would believe that any rational agent who had your beliefs and desires would, regardless of having or not having the defect, arrive at the same decision. Thus *you* should assign the same probability to your deciding to smoke (and doing so) given that you have the defect as you would assign to your deciding to smoke (and doing so) given that you do not have the defect. Once you do that, however, the dominant act is the rational one for you to choose.

Eells then argues that the same idea applies to the Predictor. Assuming that as a rational agent you do not believe in backward causation, mind control, or miracles, Eells concludes that you must believe that some common cause is responsible for both your choice and the Predictor's prediction. Perhaps there is something about our appearance that distinguishes those who will take just the blue box from those who will take both. Perhaps the Predictor is not even conscious of this. But no matter; you do not need to know what the common cause is or whether you have it. You simply need to believe that the Predictor case follows the same model as the smoking decision. Then the same reasoning shows that you should assign the same probability to there being $1,000,000 in the blue box given that you chose it alone as you would to there being $1,000,000 in it given that you took both.

Although the Eells solution to the Predictor paradox avoids importing causality directly into the formal scaffolding of decision theory, it still depends heavily on causal reasoning in analyzing the problematic decisions. If you do not believe in causality, then you could hardly accept Eells's assumptions to the effect that rational agents in the Predictor situation will believe that there must be a common cause that links their choice and the Predictor's prediction. A full reconciliation between causal and evidential decision theory is thus still not in the offing.

(A final note: There are variants of causal decision theory that reduce causal necessities to something else, for example, to natural laws. But Humean objections apply to these views too. Also I have only stated a response to Hume that happens to be particularly suited to our concern with decision theory. Philosophers have propounded several subtle and profound responses to Hume. Yet the debate over the reality of causality rages on, with the advent of causal decision theory adding fuel to the fire.)

4-6b. Concluding Remarks on the Paradoxes

We have just reviewed a potpourri of paradoxes and have found definitive resolutions to none of them. Does this mean that utility theory is doomed? That its foundations are rotten and crumbling? How should we respond to the paradoxes?

First, we must determine the type of evidence that is to count as relevant to utility theory. The fact that most people have an aversion to betting on coins of unknown bias or prefer the bird in the hand, facts on which the Ellsberg and Allais paradoxes depend, need not be relevant to abstract decision theory. For the theory claims to describe the behavior of ideally rational beings rather than that of ordinary people. On the other hand, we cannot determine what an ideally rational agent would do by asking one or observing one in a psychological laboratory. Such beings are, after all, merely hypothetical extrapolations based on what we think we would do if we had better memories, longer attention spans, quicker brains, sharper mathematical powers, and so forth. Thus our judgment that a given act is irrational could be entirely relevant to a theory of an ideally rational agent. In confronting the Ellsberg and Allais paradoxes, then, one of the chief issues will be whether we should take our aversion to the choices recommended by utility theory as failures on our part to be less than fully rational or as evidence that utility theory has incorrectly characterized rational choice under risk.

Decision theorists have taken both approaches to those paradoxes. Savage, you will recall, tried to show us that the Allais paradox is simply a trick that is easily explained away when seen in a clear light. But others have complicated and revised decision theory to discount the utility of acts involving systems whose biases are unknown and to mark up those involving certainties. If we decide to reject Savage's or similar resolutions, we should take such theories seriously and see whether they harbor their own paradoxes or anomalies. Since these theories are even more complicated than standard utility theory, I must forgo that task here.

The St. Petersburg and Predictor paradoxes, by contrast, are intended to show not that utility theory conflicts with our untutored views of rationality but rather that it is internally incoherent: that it offers, as in the Predictor paradox, conflicting recommendations, or that it commits us to something of infinite utility so long as there is no upper bound to our preferences. Again, there are those who have taken these paradoxes to heart and have offered revised versions of utility theory. Causal decision theory is an example. And again, there are experts who believe that these paradoxes dissolve when carefully analyzed. Eells is a case in point.

There has been an additional and most interesting reaction to the Predictor paradox, however. Some people believe that it reveals that there are *two types of rationality*. One type is captured by decision theories that apply the dominance rule to the Predictor problem; the other is captured by theories that follow the rule of expected utility maximization. Thus, just as we have had alternative geometries for over a hundred years and alternative logics and set theories for the last fifty years, we might now have alternative decision theories.

In sum, reactions to the paradoxes range from dismissing them as intellectual illusions to proposing revisions of utility theory to entertaining the possibility of alternative conceptions of rational choice. Plainly, it will be a while before a consensus on the paradoxes is obtained.

4-7. References

The classic presentation of the version of utility theory expounded in this chapter is contained in *Von Neumann and Morgenstern*. *Raiffa* is a clear introductory account, which focuses more on the relation between monetary values and utility than I have here. *Savage* contains a brief history of the concept of utility. For the Ellsberg and Allais paradoxes see *Ellsberg* and *Allais*, respectively. Replies may be found in *Savage*, *Raiffa*, and *Eells*. *Jeffrey* has a discussion of and response to the St. Petersburg paradox. *Eells* is also a good general review of utility theory and causal decision theory. See also Skyrm's *Causal Necessity*, *Gibbard and Harper*, and *Lewis*. *Nozick* announced the Predictor paradox to the philosophical world. *Eells* lists much of the extensive literature on that paradox.