Chapter Two
Utility Theory

Game theory is based on utility theory, a simple mathematical theory for representing decisions. In utility theory, we assume that actors are faced with choices from a set of available actions. Each action provides a probability of producing each possible outcome. Utility is a measure of an actor's preferences over the outcomes that reflects his or her willingness to take risks to achieve desired outcomes and avoid undesirable outcomes. The probabilities of obtaining each outcome after taking an action represent uncertainty about the exact consequences of that action.

We calculate an expected utility for an action by multiplying the utility of each possible outcome by the probability that it will occur if the action is chosen, and then summing across all possible outcomes. Utilities for outcomes are chosen so that the magnitude of expected utilities concur with preferences over actions. Actions with larger expected utilities are preferred. Given the probabilities that actions produce outcomes and preferences over actions, we can calculate utilities over outcomes so that actions with larger expected utilities are preferred.

Utility theory is closely tied to probability theory and is almost as old. As in the case of probability theory, the rigorous analysis of gambling problems drove the early development of utility theory. Daniel Bernoulli[1] first worked on utility theory to explain why the attractiveness of gambles did not necessarily equal the gambler's monetary expectation. After this initial observation, utility theory lay dormant until Jeremy Bentham advanced utilitarianism as a philosophy in the 1800s. Bentham's utility theory was, mathematically speaking, quite sloppy and is not useful for developing a rigorous theory of decision. Consequently, utility was rejected as a useful concept until the middle of the twentieth century.

Von Neumann and Morgenstern revived utility theory by providing a firm mathematical foundation for the concept in an appendix to their *Theory of Games and Economic Behavior* ([1943] 1953). Several rigorous versions of utility theory were produced after the publication of that book. Since then, economists have reformulated economic theory using utility theory and game theory as the description of individual behavior.

This chapter begins with the concept of rationality. The characteristics of rational preferences are presented, followed by a discussion of some common misconceptions about rationality. The elements of a formal decision problem are described, and the idea of expected utility is introduced. Two examples,

one frivolous and one historical, i
mal basis of utility theory, the Exp
that a utility function can be found
those preferences observe six cond
utility theory are rebutted. I next
ferent reactions to risk and prefere
simple examples, one concerning
people should vote. I end the cha
utility theory.

The Concept of Rationali

Game theory assumes rational b
In everyday parlance, rational be
thoughtful, or reflective behavior t
not think that someone who drives
is rational. But rational behavior
common meaning of the term. Pu
the best means to gain a predeter
consistency of choices and not o
fixed goals and not of the morality

In utility theory, rational actors
can choose. They then choose the
point, I will be vague about what "
behavior is goal directed; actors
rather than less desired outcomes.

But how do we know what an
tors' goals from observing their p
assume that actors will continue
pursued in the past. We fix actors'
have and the situation they face to

To specify these goals formally
(I use these two words interchang
sult of the actor's decision. It incl
set of outcomes is exhaustive and
come occurs. The definition of the
It can be as simple or complex as
presidential election could be des
Democratic candidate won. The co
cies that the winning candidate e
the margin of victory or whether
facet of an election in the eyes of

Chapter Two
Utility Theory

mathematical theory for repre-
ıat actors are faced with choices
/ides a probability of producing
an actor's preferences over the
ake risks to achieve desired out-
probabilities of obtaining each
ertainty about the exact conse-

on by multiplying the utility of
it will occur if the action is cho-
omes. Utilities for outcomes are
ies concur with preferences over
are preferred. Given the proba-
eferences over actions, we can
ns with larger expected utilities

theory and is almost as old. As
analysis of gambling problems
. Daniel Bernoulli[1] first worked
eness of gambles did not neces-
on. After this initial observation,
ıam advanced utilitarianism as a
eory was, mathematically speak-
ing a rigorous theory of decision.
l concept until the middle of the

ıtility theory by providing a firm
an appendix to their *Theory of*
53). Several rigorous versions of
ication of that book. Since then,
ory using utility theory and game
ior.
ationality. The characteristics of
by a discussion of some common
nts of a formal decision problem
ity is introduced. Two examples,

one frivolous and one historical, illustrate these ideas. I then present the formal basis of utility theory, the Expected Utility Theorem. That theorem shows that a utility function can be found to represent preferences over actions when those preferences observe six conditions. Some common misconceptions about utility theory are rebutted. I next consider how utility functions represent different reactions to risk and preferences over time. I apply utility theory in two simple examples, one concerning deterrence and the other concerning when people should vote. I end the chapter with a discussion of the limitations of utility theory.

The Concept of Rationality

Game theory assumes rational behavior. But what is meant by rationality? In everyday parlance, rational behavior can mean anything from reasonable, thoughtful, or reflective behavior to wise, just, or sane actions. We generally do not think that someone who drives eighty miles per hour on narrow side streets is rational. But rational behavior for our purposes means much less than the common meaning of the term. Put simply, rational behavior means choosing the best means to gain a predetermined set of ends. It is an evaluation of the consistency of choices and not of the thought process, of implementation of fixed goals and not of the morality of those goals.

In utility theory, rational actors have specified goals and a set of actions they can choose. They then choose the action that will best attain those goals (at this point, I will be vague about what "best" and "attain" mean precisely). Rational behavior is goal directed; actors are trying to create more desired outcomes rather than less desired outcomes.

But how do we know what an actor's goals are? In general, we deduce actors' goals from observing their prior behavior or by experimentation. We then assume that actors will continue to pursue the goals we have deduced they pursued in the past. We fix actors' preferences and allow the information they have and the situation they face to change, creating variation in their actions.

To specify these goals formally, we begin with **consequences**, or **outcomes** (I use these two words interchangeably). A consequence is a possible final result of the actor's decision. It includes all relevant effects for the decider. The set of outcomes is exhaustive and mutually exclusive; one and only one outcome occurs. The definition of the consequences is the choice of the modeler. It can be as simple or complex as you choose. For example, the outcomes of a presidential election could be described as simply whether the Republican or Democratic candidate won. The consequences of the election could be the policies that the winning candidate enacts once in office. We could also consider the margin of victory or whether a particular individual voted as a relevant facet of an election in the eyes of an actor.

Actors have preferences over the set of outcomes. Let C be the set of all consequences, with specific consequences indicated by subscripts. Preferences are given by a relation, R, between pairs of consequences, with C_iRC_j read as "Consequence i is at least as good as consequence j."[2] R is called the **weak preference relation** because the actor weakly prefers the first outcome to the second. In other words, the first is at least as good as and maybe better than the second. **Indifference,** I, occurs when both C_iRC_j and C_jRC_i, and is written C_iIC_j. Because both outcomes are at least as good as each other, they are equally desirable in the actor's eyes. The **strong preference relation,** P, means the first outcome is better than the second. In other words, C_iPC_j is equivalent to (C_iRC_j and not C_jRC_i).

We assume each actor has a set of complete and transitive preferences over the set of outcomes. Completeness means that actors can make comparisons across all pairs of consequences.

> **Definition**: An ordering is **complete** iff (read if and only if) for all pairs of outcomes C_i and C_j, either C_iRC_j or C_jRC_i or both.

An actor can say that it prefers one outcome to the other or that it is indifferent between the two. Completeness implies that the preference order is **reflexive:** for all C_i, C_iRC_i. In other words, all consequences are at least as good as themselves. Preferences are also transitive.

> **Definition**: An ordering is **transitive** iff (C_iRC_j and C_jRC_k) implies C_iRC_k.

Transitivity means that if one outcome (C_i) is at least as good as a second (C_j), and the second is at least as good as a third (C_k), then the first must be at least as good as the third.

Completeness and transitivity are the basic elements of a preference ordering. Each actor can rank the outcomes from best to worst, allowing for indifference between outcomes. Because complete and transitive preferences order the outcomes, they are called **ordinal preferences**. These two assumptions are necessary for rationality. Without complete preferences, actors are unable to choose between noncomparable outcomes. But situations with noncomparable preferences are odd. For example, many of us would be hard pressed to state whether we prefer eating a bushel of pears or seeing Senator Kennedy elected president.[3] We do not think of those outcomes as comparable. But neither is it a plausible choice. Typically, we exclude situations with noncomparable consequences and restrict the set of consequences so that the actors have complete preferences over the whole set. Because we can choose the set of outcomes to fit the problem we are modeling, we can define the outcomes to avoid incomplete preferences.

Transitive preferences eliminat
ple, C_iPC_j and C_jPC_k and C_kPC_i.
choices. What outcome does the a
cannot say even though it prefers
choice between the outcomes in
the outcomes are presented. If we
the result to C_k, we find that C_k i
then comparing the result to C_i, w
say that these preferences are wel
The classic example of prefer
sweetness of coffee. One grain of
grains of sugar, two grains is at le
case, I am indifferent because I ca
makes to the sweetness of my cof
coffee are transitive, then I should
my coffee. But I am not indiffere
like best, and several pounds of s
cup. Less sugar than that and I pre
I prefer not adding one more grain
the addition of one grain of sugar
cannot distinguish which cup has
ison. We can avoid these situatio
distinguish, eliminating this para
Preferences over outcomes are
ing the course of the decision bei
erence occur. We do not allow p
of its explanatory power. Prefere
by observing actions. But shifts i
tempting to say that preferences
ior at variance with a model. The
ences shift when behavior change
ignored by claiming that prefere
preferences are fixed and that ch
the situation and the information
The assumption of fixed prefe
But it is not as restrictive as it
to suit our purposes. We distingu
preferences over actions (or stra
tions are choices that could produ
outcomes are assumed to be fixed
actors gain new information abou
the outcomes carefully, "shifts in
actions, rather than outcomes. Fo

Transitive preferences eliminate cycles among strong preferences, for example, C_iPC_j and C_jPC_k and C_kPC_i. Strong preference cycles prevent consistent choices. What outcome does the actor prefer among the three in this cycle? We cannot say even though it prefers one outcome in each pair of outcomes. The choice between the outcomes in the cycle depends on the order in which the outcomes are presented. If we compare C_i and C_j first, and then compare the result to C_k, we find that C_k is preferred. Comparing C_j and C_k first, and then comparing the result to C_i, we find that C_i is preferred, and so on. Can we say that these preferences are well-defined?

The classic example of preferences that violate transitivity concerns the sweetness of coffee. One grain of sugar in my coffee is at least as good as no grains of sugar, two grains is at least as good as one grain, and so on. In each case, I am indifferent because I cannot tell the difference that one grain of sugar makes to the sweetness of my coffee. If my preferences over the sweetness of coffee are transitive, then I should be indifferent about the amount of sugar in my coffee. But I am not indifferent between no sugar, the amount of sugar I like best, and several pounds of sugar per cup. I like about two teaspoons per cup. Less sugar than that and I prefer adding another grain; more than that, and I prefer not adding one more grain of sugar. So I am not really indifferent about the addition of one grain of sugar to my coffee. The paradox occurs because I cannot distinguish which cup has the additional grain of sugar in each comparison. We can avoid these situations by choosing outcomes that the actors can distinguish, eliminating this paradox of indifference.

Preferences over outcomes are assumed to be fixed. They do not change during the course of the decision being examined. Some argue that shifts in preference occur. We do not allow preference shifts because they rob the theory of its explanatory power. Preferences are unobservable; we infer preferences by observing actions. But shifts in preferences cannot be confirmed then. It is tempting to say that preferences have changed whenever we observe behavior at variance with a model. The model cannot be tested if we claim preferences shift when behavior changes because any disconfirming evidence can be ignored by claiming that preferences have changed. Instead, we assume that preferences are fixed and that changes in behavior are caused by changes in the situation and the information available to the actors.

The assumption of fixed preferences restricts the situations we can model. But it is not as restrictive as it seems because we can select the outcomes to suit our purposes. We distinguish between preferences over outcomes and preferences over actions (or strategies). Outcomes are the final results; actions are choices that could produce one of several outcomes. Preferences over outcomes are assumed to be fixed. Preferences over actions can change as the actors gain new information about the efficacy of different actions. By choosing the outcomes carefully, "shifts in preferences" are shifts in preferences among actions, rather than outcomes. For example, the outcomes of an election could be modeled as which candidate is elected, say, Bush or Dukakis for the 1988 presidential election. But then voters who change their minds about which candidate to vote for during the campaign appear to shift their preference over the outcomes. Instead, let the outcomes of the election be the policies that the winning candidate will adopt. Then voters could change their preference between the candidates as they learn what policies each candidate is likely to adopt if he or she is elected. A shift in preference between the candidates is now seen as a shift in preference over the actions of voting for Bush and voting for Dukakis.

Ordinal preferences can be represented by a descending sequence of numbers. The largest number is assigned to the most preferred outcome, the second-largest number to the next outcome in the preference order, and so on down to the least preferred outcome. Indifference between two outcomes requires equal numbers. Any sequence where the order of the numbers assigned to consequences mirrors the preference ordering is a representation of that ordering. We call these numbers **utilities**, u, and the function that maps from consequences to numbers that represent an individual's preferences over those outcomes is a **utility function**. With ordinal preferences, the larger the number the better the outcome, but the difference between the numbers assigned to two outcomes is meaningless. For example, both the function $u(C_1) = 1, u(C_2) = 2$, and $u(C_3) = 0$ and the function $u(C_1) = 1, u(C_2) = 104$, and $u(C_3) = -4$ are acceptable utility functions for the preference ordering, $C_2PC_1PC_3$. Only the order of the numbers in the representation reflects the preferences; hence the name *ordinal preferences*. If the intervals between these numbers were meaningful, we would have a cardinal preference order. Later, I show that cardinal utilities can be calculated that represent not only ordinal preferences over outcomes, but also the risks that an actor is willing to take to obtain preferred outcomes.

Some Misconceptions about the Notion of Rationality

Because the game-theoretic definition of rationality is narrower than the intuitive one, it is frequently misunderstood. Some of the common misinterpretations and the proper responses follow. First, we do not assume the decision process is a series of literal calculations. Instead, people make choices that reflect both their underlying goals and the constraints of the situation, and we can create a utility function that represents their actions given those constraints. Utility theory is not an attempt to explain the cognitive processes of individuals. It is an attempt to capture the important considerations underlying decisions in a general framework that can be manipulated mathematically while allowing for variation across the choices of individuals. Our purpose is not to explain cognition, but rather to understand political acts. We use the abstract model of choice here to represent individuals' choices in political settings. Strategic logic is quite complex even with this simple model of cognition. Rational choice models attempt to capture key facets of a situation and examine actors' decisions as a consequence of their preferences in conjunction with the constraints of a situation. To do so, we use the rational choice model to simplify away from the complexity of actual cognition.

Second, rationality tells us nothing about an actor's preferences over outcomes—only about its choices given those preferences and the situation that confronts it. The classic example here is Adolf Hitler; according to the common idea of rationality, he was crazy. He pursued abhorrent goals and took immense political risks that eventually led to his own destruction and that of Nazi Germany. But from the perspective of utility theory, his behavior can be explained rationally. He consistently pursued German nationalist expansion and responded to the environment he faced and the opportunities it presented him. In many ways, Hitler understood the international climate of the 1930s better than any other leader did.

Third, rational actors may not and probably will not all reach the same decision when faced with the same situation. Rational actors can differ in their preferences over the outcomes. A chess master playing with his or her child is unlikely to play purely to win the game. Instead, he or she strives to make the game enjoyable for the child to play. Moves that would appear irrational if the master were playing against a competitive opponent may be quite rational when playing against the child because the master's objective in playing the game is different. Moreover, even if two actors have the same ordinal preferences, they can have different reactions to risk and uncertainty that lead them to evaluate the available actions differently. Different reactions to risk are captured in the utility functions, and different reactions to uncertainty are captured in the actors' subjective probability distributions of outcomes that follow from each action. Everyone prefers winning at the roulette wheel to losing there. But some are willing to run the risk of the game while others are not. Both types prefer winning to losing, but nonplayers avoid the risks of the game, and this preference is captured in their utility functions over the outcomes. Later, I discuss attitudes to risks and how they are captured in utility functions. Finally, different actors could have different information that leads them to believe that different courses of action lead to preferred outcomes. Even risk-averse individuals might bet on a horse race if they knew which horse would win a fixed race.

Fourth, rational actors can make errors, that is, achieve undesirable outcomes, for three reasons. Rationality does not mean error-free decisions. First, situations are risky. When one takes a chancy course of action, bad outcomes may occur because one is unlucky. Second, the information available to actors is limited. Actors cannot determine the consequences of their actions and must make judgments that may be incorrect. Third, actors may hold incorrect beliefs about the consequences of their actions. They may believe that ineffective means of achieving their goals are effective. The critical point here is that hindsight is always correct, but actors do not know what the consequences of their actions will be when they must choose. Decisions can be judged properly only in the light of the conditions under which they are made. Bad decisions are obvious after their consequences are known. But actors cannot know the future; they make judgments about the likely future and act on those judgments. Utility theory attempts to capture this problem.

The difference between **ex ante facto** (before the fact) and **ex post facto** (after the fact) reasoning leads to a final observation. When we can choose the goals of an actor to fit its actions, it is very easy to produce circular explanations. For example, the statement "He jumped off the building because he wanted to" is not a very satisfying explanation of a suicide. To avoid this type of circular reasoning, we either make assumptions about preferences or base them on sufficient prior information. Frequently, both approaches are used together. Certain general assumptions are made about actors' preferences, allowing for terms that vary across actors that can be fit by examining the actors' prior behavior. Economists generally assume that "more is better" when it comes to consumer goods, but how much more of which good is an empirical question for specific individuals out to make purchases.

How Do Utility Functions Predict Actions?

Any decision problem can be described formally as follows:

1) A set of **acts**, A, one of which will be chosen as the decision.

2) A set of **states of the world**, S. The states are mutually exclusive and exhaustive—only one can occur, and one of them must occur. The "world" is defined to encompass all matters relevant to the problem beyond the control of the decider. An **event** is a subset of the states.

3) A set of **consequences** or **outcomes**, C, with one consequence for each pair of acts and states.

4) A **preference ordering** over the consequences, P. These preferences are assumed to be complete, transitive, and fixed.

Outcomes are produced by the decider's chosen action and other factors outside the decider's control. We summarize the latter factors as the **state of the world**. The state of the world consists of all those factors that influence the outcome but lie beyond the control of the decider. Generally, actors do not know the state of the world. Instead, they face several possible states of the world. If they knew the state of the world, actors could determine the consequences of their actions and choose the action that produces the most preferred outcome. But because they do not know the state of the world, deciders cannot determine the consequences of their actions. They must assess which action is most likely to produce the best final outcome, considering that each action could produce

y are made. Bad decisions are
actors cannot know the future;
act on those judgments. Utility

e the fact) and **ex post facto**
tion. When we can choose the
sy to produce circular expla-
d off the building because he
a suicide. To avoid this type of
about preferences or base them
approaches are used together.
tors' preferences, allowing for
examining the actors' prior be-
re is better" when it comes to
good is an empirical question

ctions?

lly as follows:

e chosen as the decision.

e states are mutually exclusive
nd one of them must occur. The
tters relevant to the problem be-
nt is a subset of the states.

es, C, with one consequence for

nsequences, P. These preferences
and fixed.

osen action and other factors out-
e latter factors as the **state of the**
hose factors that influence the out-
ler. Generally, actors do not know
eral possible states of the world. If
ld determine the consequences of
luces the most preferred outcome.
world, deciders cannot determine
t assess which action is most likely
ing that each action could produce

desirable or undesirable outcomes depending upon the state of the world. This assessment requires some judgment of the likelihood of each state and a more detailed scale of the actor's preferences than just ordinal preferences over the outcomes.

This more detailed scale is a **Von Neumann–Morgenstern utility function**. Such utility functions measure the attractiveness of outcomes to an actor by its willingness to take risks to obtain preferred outcomes. A probability distribution over the states capture an actor's assessment of the likelihood of each state. We calculate the expected utility for an action by multiplying the probability of each state's occurring by the utility of the outcome that results from that state and the action, and then summing these products over all the possible states. The available action with the highest expected utility is the choice. In mathematics, we have the following:

$$EU(A) = \sum_{\text{all } S} p(S)u[C(S, A)] \text{ and}$$

$$\text{Choose } A \text{ such that } EU(A) \text{ is maximized}$$

where EU is expected utility, A is an available action, p is probability, S is a state, u is utility, and C(S,A) is the consequence that results when S is the state and A the action. Each action is evaluated both for the likelihood of the consequences it could produce and for the attractiveness of those outcomes. The action with the highest expected utility among the set of available actions is the choice.

A simple example may help clarify these ideas.[4] After working hard studying decision theory, you need a break to revive yourself. You head down to the soft-drink machine, which offers Classic Coke, Diet Coke, and Sprite (all registered trademarks of the Coca-Cola Company). You need caffeine and sugar to restore your alertness. If you cannot have both caffeine and sugar, you prefer caffeine to sugar. Your ordinal preferences are Classic Coke P Diet Coke P Sprite. You deposit your money, and as you are about to press the Classic Coke button, a friend of yours walks by and says, "Do you know that the man who services the machine got the Cokes and Sprites mixed up together? He dropped a case of each, and they broke open, spilling cans of both all over the place. He was in a hurry, so he just stuffed cans in the Coke and Sprite slots without looking to see what each can was. Mary got a Sprite when she pressed the Coke button." Now your choice is not so simple. The Classic Coke button might produce a Sprite. You know the Diet Coke button will produce a Diet Coke, so you are choosing between a Diet Coke for certain and a chance at either a Classic Coke or a Sprite.

The outcome is the drink you get after pushing a button, Classic Coke, Diet Coke, or Sprite. The actions available are the three buttons on the machine.[5]

You choose an action, and the combination of your choice and other factors outside of your control produce the outcome. These other factors are called the state of the world. Here, the state of the world is which drinks are next in each slot. Classic Coke in the Classic Coke slot, Diet Coke in the Diet Coke slot, and Classic Coke in the Sprite slot is one possible state of the world.

To decide which button to push, you need to judge how likely each state of the world is and what risk of getting a Sprite you are willing to accept to try to get a Classic Coke versus taking a Diet Coke for certain. The former is summarized in a probability distribution over the states, and the latter in a utility function over the outcomes. We use these two together to calculate an expected utility for each action. The action with the highest expected utility is the choice. Assume the following utility function and probability distribution where the states are read as drink in the Classic Coke slot, drink in the Diet Coke slot, and drink in the Sprite slot; all other possible states have probability 0:

$$\text{u(Classic Coke)} = 1 \quad \text{u(Diet Coke)} = .4 \quad \text{u(Sprite)} = 0$$

$$\text{p(Classic Coke, Diet Coke, Classic Coke)} = .15$$

$$\text{p(Classic Coke, Diet Coke, Sprite)} = .3$$

$$\text{p(Sprite, Diet Coke, Classic Coke)} = .2$$

$$\text{p(Sprite, Diet Coke, Sprite)} = .35$$

We calculate the expected utility of each action. If you press the Classic Coke button, you get a Classic Coke if either of the first two states is the state of the world and a Sprite if either of the latter two states is the state of the world. Thus the probability of getting a Classic Coke if you press the Classic Coke button is the sum of the probabilities of the first two states, p(Classic Coke, Diet Coke, Classic Coke) + p(Classic Coke, Diet Coke, Sprite) = .15 + .3 = .45, and the probability of getting a Sprite is the sum of the probabilities of the last two states, p(Sprite, Diet Coke, Classic Coke) + p(Sprite, Diet Coke, Sprite) = .2 + .35 = .55. The probability of getting a Diet Coke if you press the Classic Coke button is 0 because no state produces a Diet Coke when you press the Classic Coke button. Such probabilities are conditional probabilities; what is the chance of an outcome given that you take an action. Conditional probabilities are written as p(outcome| action). The expected utility of pressing the Classic Coke button is the sum of the utility of each outcome times the conditional probability of that outcome's occurring if you press the Classic Coke button. The calculation is as follows:

EU(Press Classic Coke)

$$= p(\text{Classic Coke}|\text{Press Classic Coke})u(\text{Classic Coke})$$
$$+ p(\text{Diet Coke}|\text{Press Classic Coke})u(\text{Diet Coke})$$
$$+ p(\text{Sprite}|\text{Press Classic Coke})u(\text{Sprite})$$
$$= (.45)(1) + (0)(.4) + (.55)(0) = .45.$$

Similarly, the expected utility of pressing the Diet Coke button is .4, and the expected utility of pressing the Sprite button is .35. Your choice is the action with the highest expected utility, pressing the Classic Coke button in this case.

An Example: Nixon's Christmas Bombing

Consider a simple example abstracted from an actual decision, Nixon's Christmas bombing of North Vietnam in December of 1972. After an initial agreement to end U.S. intervention in the Vietnam War, a disagreement over the exact nature of the settlement between the United States and North Vietnam emerged. From the point of view of the Nixon administration, the North Vietnamese government was trying to extract additional concessions by holding up the signing of the accord. However, it was also possible that an honest misunderstanding over the interpretation of the original agreement had arisen. We refer to each of these possibilities as a **state of the world**. When the Nixon administration had to decide what to do, it did not know what the actual state of the world was. It had to make a judgment about which state was more likely. We call the first state "Vietnamese Bluff," or S_1, and the second "No Bluff," or S_2. The states are mutually exclusive and exhaustive; the North Vietnamese could not be both bluffing and not bluffing at the same time, but one of the two had to occur.

We consider two available **actions**: (1) a demonstration of military power through aerial bombing and (2) agreeing to the additional concessions requested by the North Vietnamese government. Obviously, there were more options available than just these two, but I have chosen to simplify the situation this way for illustrative purposes. Call the first action "Bomb," or A_1, and the second "Do Not Bomb," or A_2. The states and the actions together create the **outcomes** or **consequences**. If the Nixon administration selects A_2, we assume that it will grant additional concessions and a revised settlement will be reached regardless of whether or not the North Vietnamese were bluffing. Call this consequence C_2. If the Nixon administration chooses A_1, the outcome will depend upon how the North Vietnamese respond to the bombing. If they were bluffing (i.e., if S_1 is the state of the world), we assume that the original agreement will be reaffirmed. Call this consequence C_1. If a simple misunderstanding had occurred, however, we assume that the bombing will

Table 2.1

Acts	States: Bluff (S_1)	States: No Bluff (S_2)
Bomb (A_1)	North Vietnam returns to the table; quick agreement reached (C_1)	Talks break down; the war continues (C_3)
Do Not Bomb (A_2)	Agreement reached with additional concessions (C_2)	Agreement reached with additional concessions (C_2)

Acts, States, and Consequences in the Christmas Bombing of North Vietnam

provoke the North Vietnamese to break off the talks and resume the ground war (we assume for simplicity that the renewed ground war will not involve U.S. ground troops). Call this consequence C_3. Table 2.1 arrays the choices, states, and consequences.

In the eyes of the Nixon administration (we assume), the consequences rank in the order of their subscripts: $C_1PC_2PC_3$. The Nixon administration preferred not making additional concessions to granting those concessions, but a continued war was worse than making the additional concessions. This order gives a set of ordinal preferences over the outcomes. But a utility function over the outcomes also specifies what degree of risk an actor will accept to gain preferred outcomes. The Nixon administration viewed the additional concessions as highly undesirable and was willing to take a chance to avoid making them. Let us assume that $u(C_1) = 1, u(C_2) = .3$, and $u(C_3) = 0$. The probability distribution over the states summarizes the beliefs about what the Nixon administration thought the North Vietnamese would do in response to its own action. It believed that the North Vietnamese government was bluffing. Assume that $p(S_1) = .7$ and $p(S_2) = .3$ (these two probabilities must sum to one because the two states are exhaustive). Calculating expected utilities, we can show that the Nixon administration preferred A_1 to A_2:

$$EU(A_1) = p(S_1)u(C_1) + p(S_2)u(C_3) = (.7)(1) + (.3)(0) = .7,$$

$$EU(A_2) = p(S_1)u(C_2) + p(S_2)u(C_2) = (.7)(.3) + (.3)(.3)$$

$$= .21 + .09 = .3,$$

and

$$EU(A_1) > EU(A_2), \text{ so } A_1PA_2.$$

The decision to bomb Hanoi an
United States. *The New York Time*
tion for the bombing. Although it
the Nixon administration for the t
administration's willingness to tak
estimation that the North Vietname

Exercise 2.1: Show that eit
cient to reverse the decider
above example: (1) $u(C_2)$ =

The final result of an expected u
attached to each outcome and the
either, the choice may change, as i
The definition of the outcome
sion. How you choose to repres
the consequences of those choice
I have deliberately made the abov
tions, states, and outcomes to it.
this simple model. Decision prob
complexity. The choice of the ap
question in the design of any mo
and the problem becomes intract
and the results are trivial. The e
say with your model and how yo
complexity of a model is a choi
more detailed model and the clear
els are adequate to establish ma
complexity and mathematical sop
insight.

Exercise 2.2: Recast the al
other action, A_3, reintervei
sulting consequences if A_3
the world and C_4, a renew
as North and South Vietnar
is the state of the world. C_3
example above for $u(C_1)$, u

a) Solve for the expected u

b) Find a value for $u(C_4)$ s

The decision to bomb Hanoi and Haiphong was quite controversial in the United States. *The New York Times* heavily criticized the Nixon administration for the bombing. Although it shared the assumed ordinal preferences of the Nixon administration for the three outcomes, it did not share either the administration's willingness to take the risk of restarting the war or Nixon's estimation that the North Vietnamese were bluffing.

> Exercise 2.1: Show that either of the following changes is sufficient to reverse the decider's preference over the actions in the above example: (1) $u(C_2) = .8$ or (2) $p(S_1) = .2$ and $p(S_2) = .8$.

The final result of an expected utility comparison depends on both the utility attached to each outcome and the probability of each state. When we vary either, the choice may change, as in the exercise above.

The definition of the outcomes, actions, and states is a modeling decision. How you choose to represent the choices available to an actor and the consequences of those choices is the most important step in modeling. I have deliberately made the above example simple. One could add more actions, states, and outcomes to it. Of course, these additions will complicate this simple model. Decision problems can be structured at many levels of complexity. The choice of the appropriate level of complexity is a critical question in the design of any model. Set the level of complexity too high, and the problem becomes intractable. Set the level of complexity too low, and the results are trivial. The essential question here is what you want to say with your model and how you want to make your points. The level of complexity of a model is a choice between the additional complexity of a more detailed model and the clear exposition of a simple model. Simple models are adequate to establish many substantively important points; greater complexity and mathematical sophistication may not lead to greater political insight.

> Exercise 2.2: Recast the above decision problem by adding another action, A_3, reintervening with U.S. ground troops. The resulting consequences if A_3 is chosen are C_1 if S_1 is the state of the world and C_4, a renewed ground war involving U.S. as well as North and South Vietnamese troops (a new consequence) if S_2 is the state of the world. C_3PC_4. Continue to use the values in the example above for $u(C_1)$, $u(C_2)$, $u(C_3)$, $p(S_1)$, and $p(S_2)$.
>
> **a)** Solve for the expected utility of A_3. Let $u(C_4) = x$.
>
> **b)** Find a value for $u(C_4)$ such that A_2PA_3.

As a final note on this example, it should seem odd to you that we treat the actions of the North Vietnamese as fixed and determined by the state of the world. It is odd. The North Vietnamese government should be treated as a separate actor in its own right. To do so creates a game between the Nixon administration and the government in Hanoi. There is a natural parallel between decision theory and game theory that I want to suggest through this example. In game theory, the other players' actions parallel the states of the world in decision theory. But because other players also choose their actions to maximize their expected utility, each side's decision in a game must depend upon its anticipations about the other side's chosen action. This is a more complicated problem than the choice of an action in decision theory. But both types of decisions begin with the concept of utility presented in this chapter.

Certainty, Risk, and Uncertainty

The probability distributions reflect the decider's beliefs about which actions will lead to what consequences. These beliefs arise from the actor's knowledge about the situation. Decisions can be made under three different conditions reflecting the decider's knowledge about the states of the world:

1) Certainty—the state of the world is known prior to choosing an act. Then the act whose consequence is most preferred will be chosen. From the perspective of decision theory, decisions under certainty are trivial. The actual calculations that determine the outcome that results from each act may be so complicated that decisions under certainty may be quite computationally complex.

2) Risk—each state has a known probability of occurring, which obeys all the laws of probability. These probabilities can be assumed to be based on known frequencies over many repetitions. The condition of risk occurs in most gambling games, craps or roulette, for example, where the probabilities of different outcomes can be calculated and are well known.

3) Uncertainty—the probabilities of each state's occurring are either unknown or meaningless in the sense of reflecting any long-run frequency of occurrence.

Most decisions are made under the condition of uncertainty. Although some prior information about the likely state of the world is available, no long-run frequency of outcomes can be used to determine the chance of future outcomes. For example, people generally understand that flipping a fair coin generates a 50-50 probability of heads or tails. Thus decisions about how to bet on coin flips are made under risk. Other games of chance that people bet on, such as football games and horse races, do not have well-established, long-run frequencies.

Still, people do have prior inform
form beliefs about what outcomes
outcomes are made under uncerta
tive probability estimates that refle
state of the world. The difference
certainty lies in the probability di
Under risk, all deciders must hol
that distribution is known; under
ferent probability distributions be
underlying state of the world.

The choice of the condition of
you wish to capture in your mode
litical scientists are made under u
exist for the events, a model of dec
strate your point. For example, el
for both voters and candidates. Alt
and polls collected and analyzed,
which different actors may hold c
as a problem under risk may be su
toral competition and behavior. T
to reflect the effects of uncertainty

Utility Theory under the C

I have not yet explained the logi
ity function over the outcomes and
the outcomes, we can determine pr
tion come from? Utility functions
that an actor will accept to gain m
over outcomes are insufficient to
function. We need preferences ov
represent risky choices over the o
selected from a fixed set of conseq
each outcome. If an individual ca
quences and those preferences ove
ditions, then a utility function can

Definition: A **lottery** (or g
probabilities $(p_1, p_2, \ldots, p_n)$

eem odd to you that we treat
nd determined by the state of
ernment should be treated as a
a game between the Nixon ad-
re is a natural parallel between
suggest through this example.
allel the states of the world in
ɔ choose their actions to maxi-
n in a game must depend upon
action. This is a more compli-
decision theory. But both types
resented in this chapter.

er's beliefs about which actions
ırise from the actor's knowledge
ınder three different conditions
ıtates of the world:

known prior to choosing an act.
t preferred will be chosen. From
sions under certainty are trivial.
ıe outcome that results from each
ns under certainty may be quite

ability of occurring, which obeys
ilities can be assumed to be based
titions. The condition of risk oc-
roulette, for example, where the
be calculated and are well known.

ıch state's occurring are either un-
reflecting any long-run frequency

ɔn of uncertainty. Although some
ıe world is available, no long-run
ıine the chance of future outcomes.
hat flipping a fair coin generates a
sions about how to bet on coin flips
e that people bet on, such as foot-
-established, long-run frequencies.

Still, people do have prior information about those games that allows them to form beliefs about what outcomes are likely. Decisions on bets on these future outcomes are made under uncertainty. Under uncertainty, actors form subjective probability estimates that reflect their degree of belief about the underlying state of the world. The difference between the condition of risk and that of uncertainty lies in the probability distributions the deciders hold over the states. Under risk, all deciders must hold the same probability distribution because that distribution is known; under uncertainty, different deciders may hold different probability distributions because they hold different beliefs about the underlying state of the world.

The choice of the condition of information in a model is a matter of what you wish to capture in your model. Although most decisions of interest to political scientists are made under uncertainty because no long-run frequencies exist for the events, a model of decision under risk may be sufficient to demonstrate your point. For example, elections present decisions under uncertainty for both voters and candidates. Although long-run tendencies can be calculated and polls collected and analyzed, the actual outcome is a unique event, about which different actors may hold divergent beliefs. Still, modeling an election as a problem under risk may be sufficient to establish some points about electoral competition and behavior. The assumption of risk can be loosened later to reflect the effects of uncertainty.

Utility Theory under the Condition of Risk

I have not yet explained the logical basis of utility functions. Given a utility function over the outcomes and probability distributions from the actions to the outcomes, we can determine preferred actions. Where does the utility function come from? Utility functions capture the risks of less preferred outcomes that an actor will accept to gain more preferred outcomes. Ordinal preferences over outcomes are insufficient to create a Von Neumann–Morgenstern utility function. We need preferences over all possible risky choices. Abstractly, we represent risky choices over the outcomes as **lotteries**, where one outcome is selected from a fixed set of consequences with known probabilities of selecting each outcome. If an individual can rank all possible lotteries over the consequences and those preferences over the lotteries observe certain regularity conditions, then a utility function can be calculated to reflect those preferences.

Definition: A **lottery** (or **gamble**) is a matched pair of a set of probabilities $(p_1, p_2, \ldots, p_n)$ with

$$\sum_{i=1}^{n} p_i = 1$$

and a set of prizes $(Q_1, Q_2, \ldots, Q_n)$ (these prizes could be either consequences or other lotteries), where p_i gives the probability of obtaining prize Q_i. A lottery with only consequences as prizes is a **simple lottery**; one with other lotteries as prizes is a **compound lottery**. A lottery is written $(p_1Q_1, p_2Q_2, \ldots, p_nQ_n)$.

Actions are represented as lotteries, with the probability of each consequence's being the prize representing the probability that that consequence will occur given that action. Compound lotteries provide a way to consider decision problems where some risks lead to further risks. If we have an actor's preferences over all lotteries over the consequences, we can summarize those preferences in a utility function. The Expected Utility Theorem gives the consistency conditions that individuals' preferences over lotteries must observe for us to calculate a utility function to represent those preferences.

Theorem (the Expected Utility Theorem): If

1) the preference ordering over consequences is transitive and complete (for convenience, the subscripts on consequences give the preference ordering: $C_1PC_2PC_3 \ldots PC_n$),

2) compound lotteries can be reduced to simple lotteries (assuming each lottery is independent),

3) for each consequence C_i, there exists a lottery $\hat{C}_i$ involving just C_1 and C_n (the most and least preferred consequences) such that the actor is indifferent between C_i and $\hat{C}_i$,

4) $\hat{C}_i$ can be substituted for C_i in any lottery,

5) preference among lotteries is complete and transitive, and

6) (the Sure Thing Principle) $[pC_1, (1-p)C_n]$ is preferred to $[p'C_1, (1-p')C_n]$ iff $p > p'$,

then there exists a set of numbers $(u_1, u_2, \ldots, u_n)$ associated with a set of consequences $(C_1, C_2, \ldots, C_n)$ such that for any two lotteries L and L′, the magnitudes of the expected values $(p_1u_1 + p_2u_2 + \ldots + p_nu_n)$ and $(p'_1u_1 + p'_2u_2 + \ldots + p'_nu_n)$ reflect the preference between the lotteries.

If the six conditions above are satisfied and we have an actor's preferences over lotteries, we can calculate a utility function such that the magnitude of expected utilities gives the actor's preferences among lotteries.

But what do these six conditic
says that actors have preference o
assumption earlier. The second cc
ity to reduce compound lotteries t
prizes. Some actions lead to multi
lottery. Reducing compound lotte
over them because the other cond
ies. Reduction of compound lotte
bility, under the assumption that
dent. For example, let $L_1 = (.4$
$(.6L_1, .4L_2)$ is equivalent to L'_3
$.2)C_1, .36C_2, .2C_3] = (.44C_1, .36$

Exercise 2.3: Reduce the c
simple lottery among C_1, C
defined as in the example d

The third condition allows us
worst outcome that makes the de
consequence for certain. The fou
can be substituted for the consequ
us to take any lottery, compound
between the best and the worst c
to a simple lottery in all the con
lottery for each outcome in the si
comes. Finally, reduce that new c
just the best and worst outcomes.
But the third condition also co
the best outcome is not so desir
with a nonzero chance of the best
and the worst outcome is not so
other outcome for certain over a
outcome. Put another way, the
outcomes. Infinite utilities produ
argue that nuclear war has a neg
tion, then you must prefer any c
own personal nonnuclear nightm
any nonzero probability of nucle
to a lottery in which everyone on
billion–plus coins come up heads
comes up tails, world peace (or
I find it hard to believe that you
certain over the lottery. My point

se prizes could be either
$_i$ gives the probability of
onsequences as prizes is
as prizes is a **compound**
$\ldots, p_nQ_n)$.

ne probability of each conse-
bability that that consequence
ries provide a way to consider
further risks. If we have an
e consequences, we can sum-
n. The Expected Utility Theo-
dividuals' preferences over lot-
lity function to represent those

n): If

quences is transitive and
ipts on consequences give
$PC_n)$,

simple lotteries (assuming

s a lottery $\hat{C}_i$ involving just
ed consequences) such that
$\hat{C}_i$,

ottery,

lete and transitive, and

$-p)C_n]$ is preferred to $[p'C_1$,

$u_2, \ldots, u_n)$ associated with a
uch that for any two lotteries
cted values $(p_1u_1 + p_2u_2 +$
$p'_nu_n)$ reflect the preference

d we have an actor's preferences
nction such that the magnitude of
ces among lotteries.

But what do these six conditions commit us to accept? The first condition says that actors have preference orders over the consequences. I discussed this assumption earlier. The second condition allows us to use the laws of probability to reduce compound lotteries to simple lotteries, those with just outcomes as prizes. Some actions lead to multiple risks and can be thought of as a compound lottery. Reducing compound lotteries to simple lotteries gives us preferences over them because the other conditions provide preferences over simple lotteries. Reduction of compound lotteries is done according to the laws of probability, under the assumption that the probabilities in each lottery are independent. For example, let $L_1 = (.4C_1, .6C_2)$ and $L_2 = (.5C_1, .5C_3)$. Then $L_3 = (.6L_1, .4L_2)$ is equivalent to $L'_3 = [.6(.4C_1, .6C_2), .4(.5C_1, .5C_3)] = [(.24 + .2)C_1, .36C_2, .2C_3] = (.44C_1, .36C_2, .2C_3)$.

> Exercise 2.3: Reduce the compound lottery $(.3L_1, .4L_2, .3L_3)$ to a simple lottery among C_1, C_2, and C_3 where each of the lotteries is defined as in the example directly above.

The third condition allows us to create a lottery between the best and the worst outcome that makes the decider indifferent between this lottery and the consequence for certain. The fourth condition states that the equivalent lottery can be substituted for the consequence. Together, these two assumptions allow us to take any lottery, compound or simple, and reduce it to a simple lottery between the best and the worst outcome. First, reduce any compound lottery to a simple lottery in all the consequences. Second, substitute the equivalent lottery for each outcome in the simple lottery other than the best and worst outcomes. Finally, reduce that new compound lottery to a simple lottery between just the best and worst outcomes.

But the third condition also contains a very important pair of assumptions—the best outcome is not so desirable that the decider will accept any lottery with a nonzero chance of the best outcome over any other outcome for certain, and the worst outcome is not so undesirable that the decider will accept any other outcome for certain over any lottery with a nonzero chance of the worst outcome. Put another way, the third condition excludes infinite utilities for outcomes. Infinite utilities produce very bizarre behavior.[6] For example, some argue that nuclear war has a negative infinite utility. If one accepts this position, then you must prefer any outcome without nuclear war—imagine your own personal nonnuclear nightmare here—for certain to any lottery that gives any nonzero probability of nuclear war. The nonnuclear nightmare is preferred to a lottery in which everyone on earth flips a coin simultaneously and if all five billion–plus coins come up heads, a nuclear war results, while if even one coin comes up tails, world peace (or your own personal idea of bliss) is realized.[7] I find it hard to believe that you would choose the nonnuclear nightmare for certain over the lottery. My point here is not that infinite utilities do not "exist," but, rather, that infinite utilities have some bizarre consequences for the choices predicted.

In the fifth condition, we assume transitivity in preferences among lotteries. The sixth condition is the Sure Thing Principle. If we compare two simple lotteries between the best and the worst outcomes, the lottery that gives the larger chance of the best outcome (and thus the lower probability of the worst outcome) should be preferred. It is a "sure thing." The Sure Thing Principle allows us to compare any lotteries. We reduce them to a simple lottery of the best and the worst outcomes, using the equivalent lotteries for each outcome. The lottery that produces the greater chance of the most preferred outcome is preferred according to the Sure Thing Principle.

These six conditions together allow us to estimate utilities as follows. First, we find a lottery equivalent to each consequence, using the third condition. Using the second and fourth conditions, we can substitute these lotteries for the consequences in any lottery and reduce the resulting compound lottery to a simple lottery between the best and worst outcomes, C_1 and C_n. Then the rank of all lotteries is given by the probability of obtaining the best outcome in these reduced lotteries by the Sure Thing Principle. Set the utility of the best consequence at 1, $u(C_1) = 1$, and the utility of the worst consequence at 0, $u(C_n) = 0$. The utility of each consequence is just the probability of obtaining C_1 in its equivalent lottery from the third condition.

The Expected Utility Theorem provides us with a way to represent deciders' preferences on a cardinal scale. The differences in utility among outcomes allow us to judge what risks deciders will accept. Because utilities are calculated from deciders' willingness to take risks, they measure relative preference among outcomes by the probabilities of obtaining different outcomes. Such utility functions are called **Von Neumann–Morgenstern utility functions** in honor of the two men who originally proved the Expected Utility Theorem.

> <u>Exercise 2.4</u>:
>
> a) Calculate a Von Neumann–Morgenstern utility function consistent with the following preferences: $C_1PC_2PC_3PC_4$ with equivalent lotteries $\hat{C}_2 = (.65C_1, .35C_4)$ and $\hat{C}_3 = (.4C_1, .6C_4)$ for C_2 and C_3.
>
> b) Using this utility function, determine which of the following two lotteries is preferred: $L_1 = (.3C_1, .2C_2, .2C_3, .3C_4)$ or $L_2 = (.03C_1, .4C_2, .5C_3, .07C_4)$.

Utility functions reflect an individual's preferences for lotteries over the possible consequences. We can predict an individual's decisions if we have a utility function representing the decider's preferences and the probability of each

possible consequence for each ac
action—the action with the great

Some Common Miscon

The concept of utility may seen
This section covers some simple
conception, and then explain wh

Misconception 1: L *is pref*
of L is greater than the exp

This fallacy has the causati
greater because L is preferred
preferences; preferences do no
works with utility theory slips i
close identification between util
els. Nevertheless, utilities are co
not a simulacrum of the decisi
proaches that contend that actors
are incorrect on their face. The
process within individuals. Rat
are consistent with observed beh

Misconception 2: Let A, B,
Suppose that u(A) + u(D
should be preferred to L′
same expected utility beca

If two lotteries have identical
different between them. Otherw
utility of L is $\frac{1}{2}$u(B) + $\frac{1}{2}$u(C).
which equals the expected utilit
be indifferent between these tw
represent reactions to risky situ
chose the utility of an outcome
accept in the lottery between th
to that outcome. The other assun
that these values represented all
are consistent. Then all reactio
pected utilities. The variance of
theory.

possible consequence for each action. We calculate the expected utility for each action—the action with the greatest expected utility is the choice.

Some Common Misconceptions about Utility Theory

The concept of utility may seem simple, but it is frequently misunderstood. This section covers some simple misunderstandings of utility. I state each misconception, and then explain why it is incorrect.

> *Misconception 1:* L *is preferred to* L′ *because the expected utility of* L *is greater than the expected utility of* L′.

This fallacy has the causation backwards—the expected utility of L is greater because L is preferred to L′. Utilities are constructed to represent preferences; preferences do not arise from utilities. Almost everyone who works with utility theory slips into this fallacy at some point because of the close identification between utility calculations and decisions in utility models. Nevertheless, utilities are constructs to represent preferences over actions, not a simulacrum of the decision process. Criticisms of rational choice approaches that contend that actors do not perform the calculations in the theory are incorrect on their face. There is no claim that utilities reflect cognitive process within individuals. Rather, utility functions can be constructed that are consistent with observed behavior.

> *Misconception 2: Let* A, B, C, *and* D *be outcomes with* APBPCPD. *Suppose that* $u(A) + u(D) = u(B) + u(C)$. *Then* $L = (\frac{1}{2}B, \frac{1}{2}C)$ *should be preferred to* $L' = (\frac{1}{2}A, \frac{1}{2}D)$, *even though they have the same expected utility because the former has less utility variance.*

If two lotteries have identical expected utilities, then the decider must be indifferent between them. Otherwise the behavior is inconsistent. The expected utility of L is $\frac{1}{2}u(B) + \frac{1}{2}u(C)$. The expected utility of L′ is $\frac{1}{2}u(A) + \frac{1}{2}u(D)$, which equals the expected utility of L by assumption. Then the decider must be indifferent between these two lotteries. Utility functions are calculated to represent reactions to risky situations in the expected utilities. Recall that we chose the utility of an outcome based on the risk the decider was willing to accept in the lottery between the best and worst outcome that was equivalent to that outcome. The other assumptions of the expected utility theorem showed that these values represented all preferences of the lotteries if those preferences are consistent. Then all reactions to risk in lotteries are included in the expected utilities. The variance of the expected utilities has no meaning in utility theory.

> *Misconception 3: Let* A, B, C, *and* D *be outcomes with* APBPCPD. *Suppose that* $u(A) - u(B) > u(C) - u(D)$. *Then the change from* B *to* A *is more preferred than the change from* D *to* C.

Utility functions only represent preferences between gambles. They cannot be used to draw conclusions about the net desirability of moving from one outcome to another. If we could express this change in terms of lotteries, then we could judge which lottery would be preferred. But utilities do not give the net desirability of changes between outcomes. Rather, they specify what risks a decider prefers among a set of available actions.

> *Misconception 4: If* A *and* B *are outcomes,* i *and* j *are actors, and* $u_i(A) - u_i(B) > u_j(A) - u_j(B)$, *then* i *prefers the change from* B *to* A *more than* j *does.*

Utility theory does not allow interpersonal comparisons of utilities. A utility function simply represents an individual's preferences over gambles, not his or her "true" intensity of preference. For example, let $u_i(A) = 1$, $u_i(B) = 0$, $u_j(A) = \frac{1}{2}$, and $u_j(B) = 0$. At first glance, it may appear that i prefers the shift from B to A more than j does. But if you recall the construction of the utilities from the equivalent lotteries, the utility of the best and worst outcomes were chosen arbitrarily to be 1 and 0, respectively. If we chose the endpoints of j's utility scale to be 10 and 0 instead, then all of j's utilities would be multiplied by a factor of 10, yielding $u_j(A) = 5$ and $u_j(B) = 0$. But then j appears to prefer the change from B to A more strongly than i does.

Interpersonal comparisons are more difficult than they appear from this simple illustration. Even if we give all actors identical utility for their best and worst outcomes, we cannot know that each actor's best outcome produces the same amount of "bliss" for him or her. Imagine a set of outcomes where all agree the best outcome is receiving a shot of heroin. A heroin addict would find this outcome far more desirable than would those of us with no curiosity about the effects of heroin and a substantial fear of addiction. Assigning the same utility to this outcome for all actors does not imply that all see the outcome as equally desirable.

Utility Functions and Types of Preferences*

Having covered the technical details of how utility is calculated and some common misunderstandings of the concept, I will now demonstrate the use of utility functions. Often we treat outcomes as a continuous set and define a

*This section requires differential calculus.

utility function across them. Altho
money cannot be divided into sm
poses), it is often easier to treat a
For example, there are many, but
that a candidate can adopt on an is
a hundred or more, it is easier to r
a continuum from the furthest left
issue. A function allows us to spe
utility of a position is the value of
For example, let x be a numbe
with $a > 0$, so more x is better—
preferable to L = (.5, x = 1; .5, x

$$u(2) = a$$

and

$$u(L) = .5[a(1) + b] +$$
$$= 2a + b.$$

This decider is indifferent betwee
L; u(2) = u(L).
Now let L′ = (.5, x = −1; .5,
Which is preferred?

$$u(L') = .5[a(-1)$$

and

$$u(L*) = .3[a(-10)$$

so L*PL′.
Neither of these answers depe
Utility functions are determined o
any number to a utility function o
resulting function represents the s
determines in part why interperso
If I can always multiply my utilit
its ability to represent my prefe
between any two outcomes is me
as I desire simply by multiplying
But then how do utility func
risks? In their shapes. The utility

tcomes with APBPCPD.
)). Then the change from
from D *to* C.

between gambles. They cannot
esirability of moving from one
nange in terms of lotteries, then
ed. But utilities do not give the
Rather, they specify what risks
ons.

es, i *and* j *are actors, and*
efers the change from B *to*

comparisons of utilities. A util-
's preferences over gambles, not
example, let $u_i(A) = 1, u_i(B) =$
, it may appear that i prefers the
ou recall the construction of the
ty of the best and worst outcomes
ctively. If we chose the endpoints
n all of j's utilities would be mul-
and $u_j(B) = 0$. But then j appears
gly than i does.
ult than they appear from this sim-
identical utility for their best and
actor's best outcome produces the
agine a set of outcomes where all
of heroin. A heroin addict would
would those of us with no curios-
antial fear of addiction. Assigning
ors does not imply that all see the

eferences*

ow utility is calculated and some
pt, I will now demonstrate the use
es as a continuous set and define a

utility function across them. Although outcomes are rarely continuous (even money cannot be divided into smaller units than pennies for practical purposes), it is often easier to treat a large set of outcomes as a continuous set. For example, there are many, but probably only a finite number of, positions that a candidate can adopt on an issue. If the number of positions is large, say a hundred or more, it is easier to represent the set of positions on the issue as a continuum from the furthest left to the furthest right position possible on the issue. A function allows us to specify the utility of each position easily. The utility of a position is the value of the utility function for that position.

For example, let x be a number between -10 and 10. Let $u(x) = ax + b$, with $a > 0$, so more x is better—utility increases with x. Is $x = 2$ for certain preferable to $L = (.5, x = 1; .5, x = 3)$?

$$u(2) = a(2) + b = 2a + b,$$

and

$$u(L) = .5[a(1) + b] + .5[a(3) + b] = .5a + .5b + 1.5a + .5b$$
$$= 2a + b.$$

This decider is indifferent between an outcome of 2 for certain and the lottery L; $u(2) = u(L)$.

Now let $L' = (.5, x = -1; .5, x = 5)$ and $L^* = (.3, x = -10; .7, x = 8)$. Which is preferred?

$$u(L') = .5[a(-1) + b] + .5[a(5) + b] = 2a + b,$$

and

$$u(L*) = .3[a(-10) + b] + .7[a(8) + b] = 2.6a + b,$$

so L^*PL'.

Neither of these answers depends upon the values of a or b (given $a > 0$). Utility functions are determined only up to a linear transformation. You can add any number to a utility function or multiply it by any positive number, and the resulting function represents the same preferences over lotteries. This property determines in part why interpersonal comparisons of utilities are not permitted. If I can always multiply my utility function by any positive number and retain its ability to represent my preferences, the magnitude of utility differences between any two outcomes is meaningless. I can make that difference as large as I desire simply by multiplying my utility function by a large number.

But then how do utility functions capture different willingness to take risks? In their shapes. The utility function above, $ax + b$, evaluates gambles according to their mathematical expectation (see Appendix One for a definition of a mathematical expectation). The first pair of gambles had an identical expected value, 2. The first gamble in the second pair had an expected value of 2; the second, 2.6. We can show that the expected utility of any gamble using this utility function is a(expectation of the gamble) +b. **Risk-neutral** actors have no premium for or against risky gambles; they evaluate gambles solely on the basis of the expectation of those gambles. Risk-neutral utility functions are linear.

> Exercise 2.5: Show that the expected utility of a gamble, G, with $G = (p_1x_1, p_2x_2, \ldots p_nx_n)$, is $aE(G) + b$ for $u(x) = ax + b$, where $E(G) = \sum_{i=1}^{n} p_ix_i$ is the expectation of G.

Preferences for and against risks are represented in the shape of utility functions. Deciders unwilling to take risks have utility functions that bow upwards from a line, like an upside-down bowl. The mathematical term is *concave downward;* the second derivative of the utility function is less than zero, $u''(x) < 0$. Such actors are called **risk averse**. Among gambles with an equal expectation, risk-averse actors prefer the gamble with a smaller variance in outcome. Further, risk-averse actors prefer some gambles that have both less risk and a lower expectation than other gambles. **Risk-acceptant** preferences are represented by utility functions that bow below a line. The mathematical term is *concave upward;* the second derivative of the utility function is greater than zero, $u''(x) > 0$. Risk-acceptant actors prefer risky gambles among a set with the same expectation. They also prefer some gambles with both greater risk and lower expectation than other gambles. Neither risk acceptance nor risk aversion implies that such actors always prefer gambles with greater or lower risks, respectively. An actor that is more risk acceptant than another accepts all the gambles that the latter will and some that the latter will not.

Figure 2.1 shows three utility functions: one risk neutral, u_{rn}; one risk averse, u_{rav}; and one risk acceptant, u_{rac}. The outcomes are graphed on the horizontal axis, and utility on the vertical axis. Consider a choice between a median outcome, M, for certain and a gamble that will result in either the best outcome, B, with a probability of p, or the worst outcome, W, with a probability of $1 - p$. All three deciders have the same utility for the best and for the worst outcome. Let M be such that the risk-neutral decider is indifferent between the lottery and the median outcome for certain, $u_{rn}(M) = pu(B) + (1 - p)u(W) = EU$. The risk-neutral utility function is linear as noted above. The risk-averse utility function arches above the risk-neutral function. The risk-averse utility function prefers the median

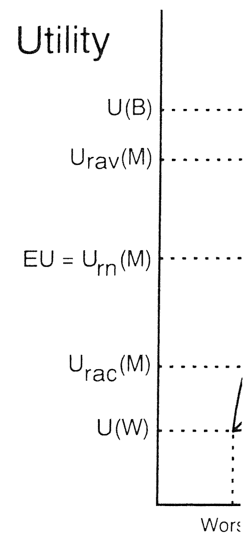

Figure 2.1
with Dif

outcome for certain over the lot
come is greater than the expect
risk-acceptant utility function cu
prefers the lottery to the median
exceeds its utility for the median

Exercise 2.6: Calculate the
ing gambles for each of th
nine calculations):

Gamble
$L_1 = (.5, x = 0; .$
$L_2 = (1, x = 1)$
$L_3 = (.8, x = 0; .$

For each utility function,
tractive to least attractive.

see Appendix One for a defini-
pair of gambles had an identical
nd pair had an expected value of
cted utility of any gamble using
mble) +b. **Risk-neutral** actors
s; they evaluate gambles solely
es. Risk-neutral utility functions

tility of a gamble, G, with
for $u(x) = ax + b$, where
G.

esented in the shape of utility
ave utility functions that bow
owl. The mathematical term is
the utility function is less than
verse. Among gambles with an
gamble with a smaller variance
ome gambles that have both less
es. **Risk-acceptant** preferences
below a line. The mathematical
of the utility function is greater
efer risky gambles among a set
ome gambles with both greater
s. Neither risk acceptance nor
s prefer gambles with greater
more risk acceptant than an-
ill and some that the latter will

one risk neutral, u_{rn}; one risk
The outcomes are graphed on
l axis. Consider a choice be-
d a gamble that will result in
y of p, or the worst outcome,
eciders have the same utility
M be such that the risk-neutral
l the median outcome for cer-
The risk-neutral utility func-
utility function arches above
ty function prefers the median

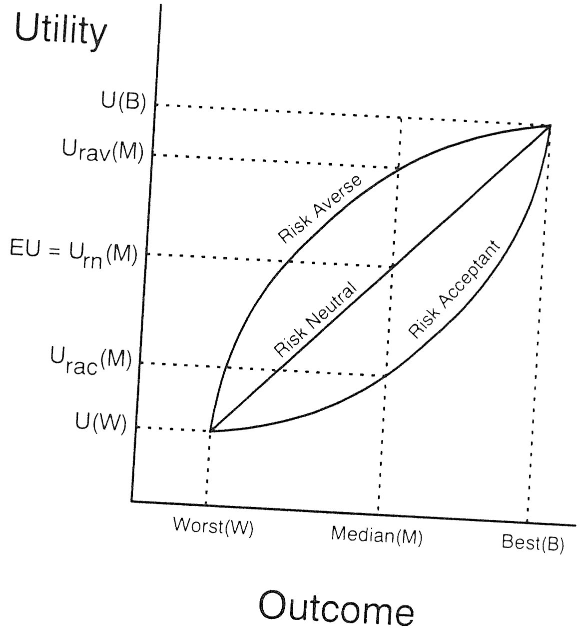

Figure 2.1 Three Utility Functions
with Different Risk Attitudes

outcome for certain over the lottery because the utility of the median outcome is greater than the expected utility of the lottery, $u_{rav}(M) > EU$. The risk-acceptant utility function curves below the risk-neutral utility function. It prefers the lottery to the median for certain; its expected utility for the lottery exceeds its utility for the median outcome for certain, $u_{rac}(M) < EU$.

Exercise 2.6: Calculate the expected utility of each of the following gambles for each of the following utility functions (a total of nine calculations):

Gamble	Utility Function
$L_1 = (.5, x = 0; .5, x = 2)$	$u_1(x) = x$
$L_2 = (1, x = 1)$	$u_2(x) = x^2$
$L_3 = (.8, x = 0; .2, x = 4)$	$u_3(x) = \sqrt{x}$

For each utility function, rank the three gambles from most attractive to least attractive. Which of the three is risk acceptant, which is risk averse, and which is risk neutral? (Hint: it may help to sketch each of the three functions.)

Utility functions need not have the same risk attitude across all ranges of values. They can be risk averse in some areas and risk acceptant or neutral in others.

Preferences over Time

Actors can also differ in their preference over time when they receive prizes over several time periods. Models often assume that actors make choices over several time periods of equal length, where the length of each period is fixed and known in advance. Desirable outcomes earlier are preferable. **Discount factors** represent the deciders' impatience for rewards. An actor's discount factor, usually written as δ, falls between 0 and 1. The smaller it is, the more the actor prefers prizes now over prizes later. If the utility of an outcome is x now, the discounted value of that outcome one time period from now is δx, its discounted value two time periods from now is $\delta^2 x$, and so on.

Exercise 2.7: Governments in early modern Europe, particularly the Dutch government, raised money by selling annuities. The buyer would pay a lump sum of money, say, a hundred guilders, and then once a year from then on, would receive a payment of G guilders. Assume buyers are risk neutral for money; $u(x) = x$.

a) If a decider's discount factor is .9 per year, calculate the annual payment that makes it indifferent between paying the lump sum and the annuity. (Hint: it helps to know that

$$\sum_{i=1}^{+\infty} x^i = x + x^2 + x^3 \ldots = \frac{x}{1-x};$$

see Appendix One.)

b) If a decider's discount factor is δ per year, calculate the annual payment that makes it indifferent between paying the lump sum and the annuity.

A Simple Example: The Calculus of Deterrence

The above examples are calculations detached from substantive interest. I turn to a simple example of the application of utility theory, the logic of deterrence in international relations. Deterrence, in the broadest sense, is the attempt by one nation, called the defender, to forestall anticipated actions by another nation, called the challenger, by a threat to impose costs on the second nation. For the purposes of this discussion, I will not consider the many variations

in international relations on this
model that reflects the logic of
cises in this section lead you thr
ics here is very simple; as Barry
lus of deterrence might have bei
(1991, 313).

The general logic of deterrenc
of economic sanctions, the use of
of nuclear war are a few of the 1
Deterrence also occurs in dome
ple, Congress uses its power of
agencies to try to control their po
criminals is the assumed deterre

In this model, I assume that tl
of the challenger's intentions by
absent the deterrent threat, the cl
also assume that each side treats
probabilities.

Consider the challenger's deci
threat. It faces two options: it car
down. If the challenger backs do
the challenger may suffer some c
dience, either other nations or do
for Back Down; the challenger h
tells us that it is the challenger's
challenger presses ahead with its
defender will carry out its deterre
of that retaliation. Call this outco
has utility $u_{CH}(TC)$ for it. There i
will not carry out its threat, in wh
action. Call this outcome CS for
point of view, it is a success); th
sume that the challenger prefers s
down to having the defender carry
$u_{CH}(TC)$.

Deterrence works when the ch
ahead with its intended action. Th

$$u_{CH}(BD) > p[u_{CH}$$

Solving for p, we obtain

$$p > \frac{u_{CH}}{u_{CH}}$$

Utility functions need not have the same risk attitude across all ranges of values. They can be risk averse in some areas and risk acceptant or neutral in others.

Preferences over Time

Actors can also differ in their preference over time when they receive prizes over several time periods. Models often assume that actors make choices over several time periods of equal length, where the length of each period is fixed and known in advance. Desirable outcomes earlier are preferable. **Discount factors** represent the deciders' impatience for rewards. An actor's discount factor, usually written as δ, falls between 0 and 1. The smaller it is, the more the actor prefers prizes now over prizes later. If the utility of an outcome is x now, the discounted value of that outcome one time period from now is δx, its discounted value two time periods from now is $\delta^2 x$, and so on.

Exercise 2.7: Governments in early modern Europe, particularly the Dutch government, raised money by selling annuities. The buyer would pay a lump sum of money, say, a hundred guilders, and then once a year from then on, would receive a payment of G guilders. Assume buyers are risk neutral for money; $u(x) = x$.

a) If a decider's discount factor is .9 per year, calculate the annual payment that makes it indifferent between paying the lump sum and the annuity. (Hint: it helps to know that

$$\sum_{i=1}^{+\infty} x^i = x + x^2 + x^3 \ldots = \frac{x}{1-x};$$

see Appendix One.)

b) If a decider's discount factor is δ per year, calculate the annual payment that makes it indifferent between paying the lump sum and the annuity.

A Simple Example: The Calculus of Deterrence

The above examples are calculations detached from substantive interest. I turn to a simple example of the application of utility theory, the logic of deterrence in international relations. Deterrence, in the broadest sense, is the attempt by one nation, called the defender, to forestall anticipated actions by another nation, called the challenger, by a threat to impose costs on the second nation. For the purposes of this discussion, I will not consider the many variations

... you through some alternatives. All of the mathematics here is very simple; as Barry Nalebuff has quipped, "[I]n truth, the calculus of deterrence might have better been called [the] algebra [of deterrence]" (1991, 313).

The general logic of deterrence covers many situations in politics. The threat of economic sanctions, the use of force to defend allies, and general deterrence of nuclear war are a few of the forms of deterrence in international relations. Deterrence also occurs in domestic politics. In the United States, for example, Congress uses its power of oversight as a threat against some executive agencies to try to control their policies. One justification for the punishment of criminals is the assumed deterrence of other crimes.

In this model, I assume that the situation is not purely a misunderstanding of the challenger's intentions by the defender; in other words, I assume that absent the deterrent threat, the challenger would take the anticipated action. I also assume that each side treats the other's response as a lottery with known probabilities.

Consider the challenger's decision after the defender has made a deterrent threat. It faces two options: it can press ahead with its intended action or back down. If the challenger backs down, the outcome is the existing situation, and the challenger may suffer some damage to its reputation in the eyes of its audience, either other nations or domestic constituencies. Call this outcome BD for Back Down; the challenger has utility $u_{CH}(BD)$ for it. (The subscript CH tells us that it is the challenger's utility for the back down outcome.) If the challenger presses ahead with its intended action, there is a chance, p, that the defender will carry out its deterrent threat. The challenger will suffer the costs of that retaliation. Call this outcome TC for Threat Carried out; the challenger has utility $u_{CH}(TC)$ for it. There is then a probability, $1 - p$, that the defender will not carry out its threat, in which case the challenger will take its intended action. Call this outcome CS for Challenger Succeeds (from the challenger's point of view, it is a success); the challenger has utility $u_{CH}(CS)$ for it. Assume that the challenger prefers success to backing down and prefers backing down to having the defender carry out its threat. Then $u_{CH}(CS) > u_{CH}(BD) > u_{CH}(TC)$.

Deterrence works when the challenger prefers backing down to pressing ahead with its intended action. The challenger is deterred when

$$u_{CH}(BD) > p[u_{CH}(TC)] + (1 - p)[u_{CH}(CS)].$$

Solving for p, we obtain

$$p > \frac{u_{CH}(CS) - u_{CH}(BD)}{u_{CH}(CS) - u_{CH}(TC)}.$$

We can think of p as the credibility of the defender's intention to carry out its threat. The greater p is, the stronger the challenger's belief that the defender will carry out its threat. The right-hand side of the inequality above is the challenger's critical risk. The challenger presses ahead if p, the credibility of the defender's intention, is less than its critical risk and is deterred when p is greater than that. The numerator of the challenger's critical risk gives the difference between success and backing down, the possible gain if the challenger takes its desired action. The denominator gives the difference between success and the threat's being carried out, the loss to the challenger if it presses ahead and then the defender carries out its threat.

We could determine when the challenger would press ahead with its threat if we knew the values of all the terms in the above inequality. However, these terms are difficult to measure because we cannot ask historical figures to give us their complete preferences over all possible lotteries. Instead, I derive hypotheses about when deterrence is more likely to work. If p is held constant, increases in the critical risk make deterrence less likely to succeed, and decreases make it more likely to succeed. If the challenger's critical risk is held constant, the greater the credibility of the defender's intention to carry out its threat (i.e., the larger p is), the more likely deterrence is to succeed.

How do changes in the challenger's utilities for the outcomes affect the probability that deterrence will succeed? The method of comparative statics can answer this question. Comparative statics analysis examines the effect of one variable on behavior by holding all other variables constant.[8] The greater the costs that the defender can impose on the challenger, the more likely deterrence is to succeed. As $u_{CH}(TC)$ decreases, the denominator grows, which lowers the critical risk. The greater $u_{CH}(BD)$ is, the more likely deterrence is to succeed. There are at least two ways the defender can influence $u_{CH}(BD)$. The more attractive the status quo is for the challenger, the greater is $u_{CH}(BD)$. The lower the audience costs of backing down, the more likely it is that the challenger will back down. Thus the prospects for deterrence can be improved by offering some concessions to the challenger if it backs down. Such concessions can reduce the audience costs of backing down, thereby raising the attractiveness of the resulting situation. The greater is $u_{CH}(CS)$, the less likely it is that deterrence will succeed. Increases in $u_{CH}(CS)$ raise both the numerator and denominator of the critical risk; the total effect causes the critical risk to rise. The greater the value of taking the desired action unopposed, the less likely it is that deterrence will succeed.

These results have two caveats. First, all three of the relevant terms are the challenger's utilities for the outcomes, not the outcomes themselves. The success of deterrence depends not on the objective damage that the defender can impose on the challenger, but on how the challenger reacts to the possibility of that damage. Concessions that the defender offers to encourage the challenger to back down may not be successful if the challenger does not see those

concessions as valuable. Second,
deterrence will fail if the challen
carry out its threat is sufficiently lc
does not believe the defender ma
costs are high and the benefits of

Exercise 2.8: Some argue th
to prevent action by the cha
fender not to attack the chall
cases, the challenger wants
the defender's will attack i
problem above so there is a
impose costs on the challeng
can you derive from this revi
if q > p?

The defender faces a prior choi
that threat always works, every d
never works, only defenders who
will make such a threat. But what
how the challenger will respond to
that the challenger will take some
with punishment. Call the resultin
outcome entails the challenger's
has utility $u_D(NT)$ for this outcom
the challenger could back down c
there be a probability of r that the
this probability as the defender's p
greater r is, the more convinced th
proceed regardless of a deterrent
the challenger backs down in the
before, for Back Down. The statu
down, and the defender has utility
presses ahead, the defender either
out its threat, we have outcome C
outcome TC results. For now, I ig
if the challenger presses ahead ar
lenger Presses ahead. The defend
come, where $u_D(CP)$ is the greater
fender prefers BD to NT and prefe
The defender makes a deterrent

$u_D(NT) < r[u_D($

concessions as valuable. Second, assuming that infinite utilities do not exist, deterrence will fail if the challenger's estimate of the defender's intention to carry out its threat is sufficiently low. Deterrence cannot work if the challenger does not believe the defender may carry out the threat, even if the threatened costs are high and the benefits of carrying out the action are small.

> Exercise 2.8: Some argue that deterrence involves not only a threat to prevent action by the challenger, but also a promise by the defender not to attack the challenger if the latter backs down. In some cases, the challenger wants to take the action because it fears that the defender's will attack if it backs down. Recast the decision problem above so there is a probability, q, that the defender will impose costs on the challenger if it backs down. What hypotheses can you derive from this revised model? Can deterrence ever work if $q > p$?

The defender faces a prior choice of whether to make a deterrent threat. If that threat always works, every defender wishes to make such a threat. If it never works, only defenders who gain by imposing costs on the challenger will make such a threat. But what should the defender do when it is uncertain how the challenger will respond to a threat? Assume that the defender believes that the challenger will take some action if it does not threaten the challenger with punishment. Call the resulting outcome NT for No Threat. The No Threat outcome entails the challenger's proceeding with its action, and the defender has utility $u_D(NT)$ for this outcome. If the defender makes a deterrent threat, the challenger could back down or press ahead with its intended action. Let there be a probability of r that the challenger presses ahead. We can think of this probability as the defender's perception of the challenger's intentions. The greater r is, the more convinced the defender is that the challenger intends to proceed regardless of a deterrent threat. There is a probability of $1 - r$ that the challenger backs down in the face of the threat. Call this outcome BD, as before, for Back Down. The status quo is preserved if the challenger backs down, and the defender has utility $u_D(BD)$ for this outcome. If the challenger presses ahead, the defender either carries out its threat or not. If it fails to carry out its threat, we have outcome CS, as before. If it does carry out its threat, outcome TC results. For now, I ignore the question of what the defender does if the challenger presses ahead and call the resulting outcome CP for Challenger Presses ahead. The defender has utility $u_D(CP)$ for the resulting outcome, where $u_D(CP)$ is the greater of $u_D(CS)$ or $u_D(TC)$. We assume that the defender prefers BD to NT and prefers NT to CP; $u_D(BD) > u_D(NT) > u_D(CP)$.

The defender makes a deterrent threat when

$$u_D(NT) < r[u_D(CP)] + (1 - r)[u_D(BD)].$$

Solving for r, we obtain

$$r < \frac{u_D(BD) - u_D(NT)}{u_D(BD) - u_D(CP)}.$$

As with the challenger's decision whether to press ahead, we can use comparative statics to determine when the defender is more likely to make a deterrent threat. The defender's critical risk is given by the right-hand side of the inequality above. If the defender's critical risk is held constant, deterrence is more likely to be tried as r decreases—as the defender becomes more convinced that the challenger can be deterred.

Increasing values of the defender's critical risk increase the chance that the defender will make a deterrent threat. Increases in the defender's utility for the No Threat outcome make it less likely that the defender will try deterrence. The less threatening the challenger's intended action if unchecked, that is, as $u_D(NT)$ increases, the less likely it is that the defender will run the risk of trying to deter it. The less attractive the outcome if the challenger presses forward, the less likely it is that the defender will attempt deterrence. As $u_D(CP)$ decreases, the denominator grows, and deterrence is less likely to be tried. Whether $u_D(CP)$ decreases with changes in $u_D(CS)$ and $u_D(TC)$ depends upon what the defender will do if deterrence is challenged. If the defender will not carry out the threat, that is, if $u_D(CP) = u_D(CS)$, then increases in the costs of carrying out the threat, that is, in $u_D(TC)$, have no effect on whether or not the defender tries deterrence. Similarly, changes in the audience costs of not carrying out a deterrent threat, that is, changes in $u_D(CS)$, do not matter if the defender intends to carry it out, that is, if $u_D(CP) = u_D(TC)$. Finally, increases in the defender's utility for the challenger's backing down increase the chance that deterrence will be tried. The critical risk rises.

> Exercise 2.9: Some argue that the challenger has no real intention of proceeding with the action in many situations where deterrent threats are made. Recast the model to capture this possibility. If no threat is made, there is a probability, s, that the challenger does not take the feared action, and consider this outcome the same as BD. When are deterrent threats made in this model? Discuss.

> Exercise 2.10: The assumed preferences play a large role in the analysis above. Discuss when deterrence is tried if $u_D(CP) > u_D(NT)$ and when deterrence succeeds if $u_{CH}(TC) > u_{CH}(BD)$. What situations do these alternate assumptions model?

It may seem odd to you that the two decisions appear unrelated. Deterrence is commonly thought of as an interaction between the two sides. Deterrence theorists often recommend increasing the credibility of the threat. But p, which measures the credibility of the defender's intention to carry out

its threat, is fixed in this model.
the defender. This limitation of th
modeling deterrence with decision
sides change their perceptions of
tion. The credibility of threats ca
may change its judgment of how
The defender should choose its ac
judgments about the defender's i
intends to take actions that it has
intentions of the defender. If the
may expand its challenge to other
these interactive decisions requir
following chapters.

Another Simple Example

Consider an individual deciding
which candidate. We assume that
names of Candidate 1 and Candid
date 1, C1, to Candidate 2, C2, ar
is elected over Candidate 2 is B.
0. The net benefit could include
didate 1 will enact, judgments of
material benefits that Candidate 1
from the questions of how the vote
produce preferable outcomes for
tions in decision theory in the ev
on the question of when the voter

The act of voting imposes som
from the possible benefits of votin
polls and registering to vote. Let p
i will win given that the voter take
i, voting for Candidate j, or abstai
her chance of winning, so

p(Ci wins|vote Ci) > p(C

Calculate expected utilities for al
1, voting for Candidate 2, not vot

u(vote C1) = p(C1 wins|vc
u(vote C2) = p(C1 wins|vc
u(abstain) = p(C1 wins|ab

Solving for r, we obtain

$$r < \frac{u_D(BD) - u_D(NT)}{u_D(BD) - u_D(CP)}.$$

As with the challenger's decision whether to press ahead, we can use comparative statics to determine when the defender is more likely to make a deterrent threat. The defender's critical risk is given by the right-hand side of the inequality above. If the defender's critical risk is held constant, deterrence is more likely to be tried as r decreases—as the defender becomes more convinced that the challenger can be deterred.

Increasing values of the defender's critical risk increase the chance that the defender will make a deterrent threat. Increases in the defender's utility for the No Threat outcome make it less likely that the defender will try deterrence. The less threatening the challenger's intended action if unchecked, that is, as $u_D(NT)$ increases, the less likely it is that the defender will run the risk of trying to deter it. The less attractive the outcome if the challenger presses forward, the less likely it is that the defender will attempt deterrence. As $u_D(CP)$ decreases, the denominator grows, and deterrence is less likely to be tried. Whether $u_D(CP)$ decreases with changes in $u_D(CS)$ and $u_D(TC)$ depends upon what the defender will do if deterrence is challenged. If the defender will not carry out the threat, that is, if $u_D(CP) = u_D(CS)$, then increases in the costs of carrying out the threat, that is, in $u_D(TC)$, have no effect on whether or not the defender tries deterrence. Similarly, changes in the audience costs of not carrying out a deterrent threat, that is, changes in $u_D(CS)$, do not matter if the defender intends to carry it out, that is, if $u_D(CP) = u_D(TC)$. Finally, increases in the defender's utility for the challenger's backing down increase the chance that deterrence will be tried. The critical risk rises.

Exercise 2.9: Some argue that the challenger has no real intention of proceeding with the action in many situations where deterrent threats are made. Recast the model to capture this possibility. If no threat is made, there is a probability, s, that the challenger does not take the feared action, and consider this outcome the same as BD. When are deterrent threats made in this model? Discuss.

Exercise 2.10: The assumed preferences play a large role in the analysis above. Discuss when deterrence is tried if $u_D(CP) > u_D(NT)$ and when deterrence succeeds if $u_{CH}(TC) > u_{CH}(BD)$. What situations do these alternate assumptions model?

It may seem odd to you that the two decisions appear unrelated. Deterrence is commonly thought of as an interaction between the two sides. Deterrence theorists often recommend increasing the credibility of the threat. But p, which measures the credibility of the defender's intention to carry out its threat, is fixed in this model. Actions to try to raise p are not available to the defender. This limitation of the calculus of deterrence is a consequence of modeling deterrence with decision theory. Deterrence theorists believe that the sides change their perceptions of each other's intentions through their interaction. The credibility of threats can be influenced by actions. The challenger may change its judgment of how likely the defender is to carry out the threat. The defender should choose its acts in part for their effect on the challenger's judgments about the defender's intentions. The challenger may imply that it intends to take actions that it has no interest in pursuing, simply to probe the intentions of the defender. If the defender offers no resistance, the challenger may expand its challenge to other, more valuable areas of conflict. Modeling these interactive decisions requires a game theory model, the subject of the following chapters.

Another Simple Example: The Decision to Vote

Consider an individual deciding whether to vote in an election, and if so, for which candidate. We assume that there are two candidates, with the colorful names of Candidate 1 and Candidate 2. Assume that the voter prefers Candidate 1, C1, to Candidate 2, C2, and that the voter's net benefit if Candidate 1 is elected over Candidate 2 is B. Let u(C1 elected) = B and u(C2 elected) = 0. The net benefit could include the policies that the voter believes that Candidate 1 will enact, judgments of the candidates' competence to hold office, or material benefits that Candidate 1 has promised to the voter. I simplify away from the questions of how the voter reaches the judgment that Candidate 1 will produce preferable outcomes for the voter. There are many interesting questions in decision theory in the evaluation of candidates, but I choose to focus on the question of when the voter should vote.

The act of voting imposes some costs, C, with $C > 0$, which are subtracted from the possible benefits of voting. These costs include time spent going to the polls and registering to vote. Let p(Ci wins|A) be the probability that Candidate i will win given that the voter takes action A, the action of voting for Candidate i, voting for Candidate j, or abstaining. Voting for a candidate increases his or her chance of winning, so

$$p(\text{Ci wins}|\text{vote Ci}) > p(\text{Ci wins}|\text{abstain}) > p(\text{Ci wins}|\text{vote Cj})$$

Calculate expected utilities for all three possible actions (voting for candidate 1, voting for Candidate 2, not voting):

$$\begin{aligned}
u(\text{vote C1}) &= p(\text{C1 wins}|\text{vote C1})B + p(\text{C2 wins}|\text{vote C1})0 - C,\\
u(\text{vote C2}) &= p(\text{C1 wins}|\text{vote C2})B + p(\text{C2 wins}|\text{vote C2})0 - C,\\
u(\text{abstain}) &= p(\text{C1 wins}|\text{abstain})B + p(\text{C2 wins}|\text{abstain})0.
\end{aligned}$$

Voting for Candidate 1 is always better than voting for Candidate 2 because $B > 0$. When will voters bother to vote for their preferred candidate? Compare the expected utilities of the two actions. Voting is preferred when

$$u(\text{vote C1}) > u(\text{abstain}),$$

that is, when

$$[p(\text{C1 wins}|\text{vote C1}) - p(\text{C1 wins}|\text{abstain})]B > C.$$

Although the costs of voting are small, they are not zero. The difference between the two probabilities is the marginal effect of a vote on a candidate's chance of winning. The magnitude of this difference depends on the number of other voters in the election and very roughly equals 1/[2(number of other voters)]. In a presidential election in the United States, p is vanishingly small (in the neighborhood of .00000002). Even if B is relatively large, it seems very likely that the costs of voting are greater than the expected benefits. Hence the question, Why does anyone bother to vote?

This paradox of voting produces significant problems for the application of utility theory to politics; large numbers of people vote in every election, even for such officials as dog catcher and commissioner of the Mosquito Abatement District. I will discuss several solutions to this paradox. The first involves the observation that utility functions reflect behavior, as opposed to the other way around. The costs may not be as great as we think. For example, many individuals may derive benefits from the act of voting itself (referred to as the D term), independent of how they cast their vote. A sense of civic obligation may drive them to feel they must vote, or a desire to express support for their government in general, or simply a fascination with voting machines and an appreciation for the entertainment value of voting (I fall into the last category). Then, if $D > C$ (i.e., if the benefits for the act of voting in and of itself are greater than the costs of voting), voting is rational.

But does this explanation provide us with any empirical handle on why people vote? Both C and D, the costs and benefits of voting apart from its value in electing candidates, are unobservable. Any turnout is consistent with this explanation, then. Anyone who votes has $D > C$, and anyone who does not vote has $C > D$, and there is no obvious way to check these values. Later, I discuss another possible solution to the paradox of voting.

Why Might Utility Theory Not Work?

Although utility theory is very flexible and allows us to represent consistency in behavior, psychological experiments demonstrate violations of utility theory in people's behavior. I begin with the best-known example, the Allais paradox.

Example: Choose one of the
two pairs of gambles (I sugge
choose in each case):

Choice 1:

L_1 = \$500,000 1

L_2 = \$2,500,00(

\$500,000 ·

\$0 with pr

Choice 2:

L_3 = \$500,000 \

\$0 with pr

L_4 = \$2,500,00(

\$0 with pr

If you chose one even-numb
gamble, your choices are inc
they violate the Sure Thing P
culate expected utilities for e

$u(L_1)$ = u(\$500,000);

$u(L_2)$ = .1[u(\$2,500,000)]

$u(L_3)$ = .11[u(\$500,000)]

$u(L_4)$ = .1[u(\$2,500,000)]

If L_1 is preferred to L_2, then

.11[u(\$500,000)] > .1

Add .89[u(\$0)] to each side a

.11[u(\$500,000)] + .89[u(\$C

Substituting back gives us u(l
preferred to L_4.

:r than voting for Candidate 2 because
or their preferred candidate? Compare
Voting is preferred when

u(abstain),

C1 wins|abstain)]B > C.

they are not zero. The difference be-
ginal effect of a vote on a candidate's
his difference depends on the number
roughly equals 1/[2(number of other
: United States, p is vanishingly small
en if B is relatively large, it seems very
: than the expected benefits. Hence the
ote?
nificant problems for the application of
of people vote in every election, even
nmissioner of the Mosquito Abatement
to this paradox. The first involves the
behavior, as opposed to the other way
; we think. For example, many individ-
voting itself (referred to as the D term),
. A sense of civic obligation may drive
o express support for their government
n voting machines and an appreciation
(I fall into the last category). Then, if
f voting in and of itself are greater than

with any empirical handle on why peo-
enefits of voting apart from its value in
Any turnout is consistent with this ex-
D > C, and anyone who does not vote
y to check these values. Later, I discuss
x of voting.

Work?

and allows us to represent consistency
demonstrate violations of utility theory
est-known example, the Allais paradox.

Example: Choose one of the two gambles in each of the following two pairs of gambles (I suggest you note which gambles you would choose in each case):

Choice 1:

L_1 = \$500,000 for certain, or

L_2 = \$2,500,000 with probability .1;
\$500,000 with probability .89; and
\$0 with probability .01

Choice 2:

L_3 = \$500,000 with probability .11 and
\$0 with probability .89 or

L_4 = \$2,500,000 with probability .1 and
\$0 with probability .9

If you chose one even-numbered gamble and one odd-numbered gamble, your choices are inconsistent with utility theory because they violate the Sure Thing Principle. To see the inconsistency, calculate expected utilities for each of the four gambles:

$$\begin{aligned}
u(L_1) &= u(\$500{,}000);\\
u(L_2) &= .1[u(\$2{,}500{,}000)] + .89[u(\$500{,}000)] + .01[u(\$0)];\\
u(L_3) &= .11[u(\$500{,}000)] + .89[u(\$0)];\\
u(L_4) &= .1[u(\$2{,}500{,}000)] + .9[u(\$0)].
\end{aligned}$$

If L_1 is preferred to L_2, then

$$.11[u(\$500{,}000)] > .1[u(\$2{,}500{,}000)] + .01[u(\$0)].$$

Add .89[u(\$0)] to each side and we have the following:

$$.11[u(\$500{,}000)] + .89[u(\$0)] > .1[u(\$2{,}500{,}000)] + .9[u(\$0)].$$

Substituting back gives us $u(L_3) > u(L_4)$, which implies that L_3 is preferred to L_4.

Choices of one even and one odd gamble in the pairs of choices are quite common in psychological experiments based on this paradox. In the Allais paradox, people often prefer certainties over chances in the first pair of gambles; in the second pair, they go for the big bucks. Another example, the Ellsberg paradox, uses subjective probabilities.[9] In experiments based on this paradox, people tend to shy away from an unknown probability in favor of an known probability. However, these are tendencies in behavior, not ironclad regularities. Many people do make consistent choices in these paradoxes.

What does this mean for utility theory? First of all, people do not always react to probabilities as the theory demands. When faced with very high-probablity or very low-probability events (i.e., p close to 1 or 0, respectively) some people respond in strange ways. They inflate the probabilities of very low-probability events or else ignore the probability of those events. When faced with very probable but not certain events, people tend to subjectively lower those probabilities. Consequently, they are willing to pay excessively to obtain a certain outcome.

These observations may explain why people vote. The probability of determining the outcome of an election is vanishingly small, but it may not be so in the minds of voters. If people treat this probability as, say, .01 instead of .00000002, voting could be rational for reasonable values of the costs and benefits of voting. However, such an explanation is no more satisfying than the addition of the benefits of voting argument was. Like the evocation of the D term, this argument cannot distinguish voters from nonvoters except by the observation that they voted. Hypotheses that address something other than patterns of turnout are needed to determine which hypotheses is supported by the evidence. It is possible that no evidence could separate the two explanations.

The use of reference points in decisions also violates utility theory. Preferences should not be influenced by the description of outcomes. Unfortunately, preferences can be shaped by the reference points people use to judge the outcomes. The following example, taken from Quattrone and Tversky 1988, demonstrates the idea of reference points:

> Example: There are two candidates in an election, Brown and Green, who both propose economic policies. Two highly regarded economists then make projections about the standard of living in the country over the next two years if each candidate's set of policies is adopted. The economists are both impartial and equally skilled. Both economists also provide a prediction for the standard of living in comparable nations in the next two years. Table 2.2 gives their predictions of living standards converted to dollars. Which candidate's policy do you prefer?

Table 2.2

	Other Nations	Brown's Policy	Green's Policy
Economist 1	$43,000	$65,000	$51,000
Economist 2	$45,000	$43,000	$53,000

Economic Forecasts for Brown's and Green's Policies

> In another election, Brown and Green again face each other. Their proposed policies are again evaluated by two eminent economists, who also provide predictions for a comparable set of other nations. Table 2.3 presents the new predictions. Which candidate's policy do you prefer?

It is not unusual to find that people prefer Green in the first race and Brown in the second. But the only changes from the first to the second comparison are the estimates of the other nations' standard of living. The outcomes remain the same in both comparisons. What is going on here? Of the two candidates, Brown's policy offers a higher expectation than does Green's policy, assuming each economist's prediction is equally likely to be true. But Green's policies are less risky in the economists' eyes. How people evaluate risks depends whether they see those risks as gains or losses. In the first election, Green is preferred because he offers a certain gain over the reference point of the other nations, while Brown offers a riskier gain. In the second election, Brown is preferred because she offers a chance of avoiding a loss relative to the now-higher reference point of the other nations, while Green offers the certainty of a loss. The generality lurking here is that people are risk averse when choosing among gambles that offer increases from the reference point and risk acceptant when choosing among gambles that offer losses from the reference point.

Table 2.3

	Other Nations	Brown's Policy	Green's Policy
Economist 1	$63,000	$65,000	$51,000
Economist 2	$65,000	$43,000	$53,000

More Economic Forecasts for Brown's and Green's Policies

ıble in the pairs of choices are quite
ɔased on this paradox. In the Allais
s over chances in the first pair of
ɔr the big bucks. Another example,
ɔbabilities.[9] In experiments based on
m an unknown probability in favor of
e tendencies in behavior, not ironclad
istent choices in these paradoxes.
First of all, people do not always react
Vhen faced with very high-probablity
se to 1 or 0, respectively) some people
e probabilities of very low-probability
those events. When faced with very
tend to subjectively lower those prob-
; to pay excessively to obtain a certain

y people vote. The probability of de-
; vanishingly small, but it may not be
at this probability as, say, .01 instead
for reasonable values of the costs and
:xplanation is no more satisfying than
argument was. Like the evocation of
inguish voters from nonvoters except
potheses that address something other
) determine which hypotheses is sup-
ıat no evidence could separate the two

ions also violates utility theory. Prefer-
lescription of outcomes. Unfortunately,
erence points people use to judge the
:en from Quattrone and Tversky 1988,
nts:

dates in an election, Brown and
nic policies. Two highly regarded
ns about the standard of living in
ars if each candidate's set of poli-
s are both impartial and equally
ovide a prediction for the standard
; in the next two years. Table 2.2
g standards converted to dollars.
u prefer?

Table 2.2

	Other Nations	Brown's Policy	Green's Policy
Economist 1	\$43,000	\$65,000	\$51,000
Economist 2	\$45,000	\$43,000	\$53,000

Economic Forecasts for
Brown's and Green's Policies

In another election, Brown and Green again face each other. Their proposed policies are again evaluated by two eminent economists, who also provide predictions for a comparable set of other nations. Table 2.3 presents the new predictions. Which candidate's policy do you prefer?

It is not unusual to find that people prefer Green in the first race and Brown in the second. But the only changes from the first to the second comparison are the estimates of the other nations' standard of living. The outcomes remain the same in both comparisons. What is going on here? Of the two candidates, Brown's policy offers a higher expectation than does Green's policy, assuming each economist's prediction is equally likely to be true. But Green's policies are less risky in the economists' eyes. How people evaluate risks depends whether they see those risks as gains or losses. In the first election, Green is preferred because he offers a certain gain over the reference point of the other nations, while Brown offers a riskier gain. In the second election, Brown is preferred because she offers a chance of avoiding a loss relative to the now-higher reference point of the other nations, while Green offers the certainty of a loss. The generality lurking here is that people are risk averse when choosing among gambles that offer increases from the reference point and risk acceptant when choosing among gambles that offer losses from the reference point.

Table 2.3

	Other Nations	Brown's Policy	Green's Policy
Economist 1	\$63,000	\$65,000	\$51,000
Economist 2	\$65,000	\$43,000	\$53,000

More Economic Forecasts for
Brown's and Green's Policies

Utility functions can be risk acceptant over some outcomes and risk averse over other ranges. But here those ranges are defined by a reference point that can change. If the reference point changes, then the evaluation of outcomes is inconsistent. Reference points can be altered by how the gambles are offered—by what is called the framing of the question.

The observation of framing effects is important, but it begs the question of how reference points are defined and when they change. Psychologists can choose reference points in their experiments through the wording of their questions, altering those points when they choose. It is more difficult to see what reference points should be in political situations. For example, one can argue that the status quo in international politics is the reference point. Do all actors see the status quo as the reference point, or do some actors see their reference point elsewhere? When does the reference point change during a crisis—at at every intermediate step, or only at the end? These questions are not insurmountable, but they will require answers before prospect theory increases our understanding of politics.

These questions bring up the critical consideration that leads me to use utility and game theory even in the face of these problems. Is there a better alternative? Prospect theory, the psychologists' alternative to utility theory, is substantially more complicated than utility theory. It requires information about individuals' reference points and reactions to probabilities as well as their preferences over gambles. Although a substantial proportion of experimental subjects exhibit these inconsistencies, there are no results to my knowledge that tell us which individuals exhibit which inconsistencies. There are also theories of rational choice that do not use expected utilities (Machina 1987, 1989).

Selection pressures could explain why political leaders maximize expected utility. Inconsistent choices are common but not predominant in these experiments. Typically, one-quarter to one-third of the subjects exhibit inconsistent behavior. But inconsistent behavior is inefficient. Political leaders are selected as a consequence of their own decisions, through a competitive process. Inefficient choosers are less likely to be selected and advance than efficient choosers. Even if these inconsistencies are common in a population, they may be rare in that population's leaders. A model that demonstrated that inefficient choosers were denied advancement would strengthen this argument.

Further, the experimental results all address isolated decisions, rather than interactive decisions. Social and political settings involve interactive decisions; we need a theory that allows us to predict decisions in those settings. Strategic logic is quite complicated, even given the simplified representation of choices in utility theory. Game theory is built on utility theory. Someday, it may be replaced by a theory of strategic interaction based on some theory of individual choice other than utility theory. But for now, game theory is the tool we have to think about strategic interaction.

If these inconsistencies occur c
utility theory will not be a useful f
ence theory. However, if such viol
acceptable framework. The test of
dation lies in how it can explicate
lead to novel hypotheses that can
by using game theory to formalize
actual political decisions. The inco
phenomena, not the choices of po
alternatives requires showing actu
predict and the alternative predict
Although such approaches are
powerful, their proponents have no
to the construction of social theory
problems compelling, I encourage
explain political phenomena. You
abandon when you choose the alter
to a block of coherent and sophisti
led to testable hypotheses that are
decisions. Explicating political th
stands; strategic logic is complex.
politics with game theory before d

Review

This chapter presents utility theor
actions can be represented by a
action is the product of the utilit
that outcome given the action sum
actions, the one with greater expe
Rational preferences are assum
ences over consequences and lotte
lotteries can be reduced to simple
comes and are substitutable for th
there exists a utility function that
utility function captures the willin
in time preferences are captured i

Further Reading

All of the textbooks on classical
ing section of Chapter One have

over some outcomes and risk averse
are defined by a reference point that
s, then the evaluation of outcomes is
ed by how the gambles are offered—
ion.
important, but it begs the question
/hen they change. Psychologists can
its through the wording of their ques-
pose. It is more difficult to see what
uations. For example, one can argue
s is the reference point. Do all actors
or do some actors see their reference
ce point change during a crisis—at
end? These questions are not insur-
before prospect theory increases our

nsideration that leads me to use utility
e problems. Is there a better alterna-
alternative to utility theory, is sub-
heory. It requires information about
s to probabilities as well as their pref-
ntial proportion of experimental sub-
are no results to my knowledge that
consistencies. There are also theories
l utilities (Machina 1987, 1989).
political leaders maximize expected
but not predominant in these experi-
l of the subjects exhibit inconsistent
ficient. Political leaders are selected
hrough a competitive process. Ineffi-
and advance than efficient choosers.
in a population, they may be rare in
monstrated that inefficient choosers
en this argument.
dress isolated decisions, rather than
l settings involve interactive deci-
predict decisions in those settings.
given the simplified representation
s built on utility theory. Someday, it
interaction based on some theory of
But for now, game theory is the tool
n.

If these inconsistencies occur often in many phenomena of interest, then utility theory will not be a useful foundation on which to develop political science theory. However, if such violations are rare, then utility theory may be an acceptable framework. The test of whether utility theory is an acceptable foundation lies in how it can explicate useful theories of politics. Do such theories lead to novel hypotheses that can be tested? Do we gain insight into politics by using game theory to formalize our intuition? That test requires examining actual political decisions. The inconsistencies discussed here are experimental phenomena, not the choices of political actors. Establishing the case for the alternatives requires showing actual political choices that utility theory fails to predict and the alternative predicts successfully.

Although such approaches are quite interesting and may prove to be quite powerful, their proponents have not yet demonstrated their general importance to the construction of social theory based on individual choice. If you find these problems compelling, I encourage you to read further and try to apply them to explain political phenomena. You should also have an appreciation of what you abandon when you choose the alternative to game theory. Choice theory has led to a block of coherent and sophisticated theories of politics. Its application has led to testable hypotheses that are supported by the evidence of actual political decisions. Explicating political theory by using choice theory is difficult as it stands; strategic logic is complex. We should know what we can learn about politics with game theory before discarding it for an uncertain alternative.

Review

This chapter presents utility theory. The essential idea is that preferences over actions can be represented by a utility function. The expected utility for an action is the product of the utility of each outcome times the probability of that outcome given the action summed across all outcomes. Between a pair of actions, the one with greater expected utility is preferred.

Rational preferences are assumed to be connected and transitive. If preferences over consequences and lotteries are connected and transitive, compound lotteries can be reduced to simple lotteries, equivalent lotteries exist for all outcomes and are substitutable for them, and the Sure Thing Principle holds, then there exists a utility function that represents those preferences over lotteries. A utility function captures the willingness to take risks in its values. Differences in time preferences are captured in discount factors.

Further Reading

All of the textbooks on classical game theory discussed in the Further Reading section of Chapter One have chapters on utility theory at different levels of technical difficulty. I recommend Chapter Three of Kreps 1991 as the best source for further reading. Chapters Two and Thirteen in Luce and Raiffa 1957 are also a good source, although they include material that is quite dated. I learned from Savage 1972, one of the original sources. Savage 1972 presents the mathematics demonstrating the recovery of both a utility function and a subjective probability distribution from a set of preferences over actions. Savage also provides a good discussion of the motivation of utility theory and of many misinterpretations of the theory. *The New Palgrave: Utility and Probability* (Eatwell, Milgate, and Newman 1990) is a useful place to begin further reading.

Jackman 1993, Riker 1990, and Zagare 1990 are recent justifications of the assumption of rational choice in political science. The deterrence example traces originally back to Ellsberg 1960; my presentation here is different and more general. The voting example is quite well known in political science. Original sources are Downs 1957 and Riker and Ordeshook 1968. Aldrich 1993 is a recent survey on why people vote.

The work on violations of utility theory is found in psychology. Three initial sources are Quattrone and Tversky 1988, Kahneman and Tversky 1979, and Tversky and Kahneman 1981. The first uses political science examples, the second is the most technical and detailed, and the third is the shortest. Machina 1987 and Machina 1989 are accessible introductions to non–expected utility theories of rational choice.

Game theory examines decisionma
several actors produce the final outc
depends upon the other actors' deci
tween game theory and utility theor
ory correspond to the states of the w
establish the likely moves of the ot
sponses. Before we can analyze gan
the players' interrelated decisions.

This chapter presents three diffe
go from an informal understanding
a game? The first section discuss
presents a game that captures som
example introduces the basic eleme
extensive form of a game details t
quences of their choices, how they
they know when they choose. Exten
form of a game through the use of
is a complete plan to play the game
taneously select a strategy for the
determines the outcome.

Formalizing a Situation: De
Missile Crisis

Political situations can present acto
conditions. Before we can analyze th
what choices the players have. Consi
analysis of the Crisis focuses on the
39) identifies: "Why did the Soviet
in Cuba? Why did the United State
emplacement with a blockade of Cu
withdraw the missiles?"

But these questions are just the s
tions that arise in an explanation of a
a contest between two actors, a cha
lenger attempt to change the status q