Alternative Preferences

Kerry E. Back

DOI:10.1093/acprof:oso/9780190241148.003.0025

Abstract and Keywords
The Allais and Ellsberg paradoxes are presented. Various generalizations of expected utility motivated by these and other paradoxes are discussed, including betweenness preferences, rank-dependent preferences, multiple prior max-min preferences, and prospect theory. For betweenness preferences, which include weighted utility and disappointment aversion, an investor’s marginal utility is proportional to a stochastic discount factor. Disappointment averse utility and rank-dependent utility have first-order risk aversion. Multiple prior max-min utility is one way to accommodate the Ellsberg paradox (ambiguity aversion or Knightian uncertainty). The dynamic consistency of updating multiple priors is discussed.

Keywords: Allais paradox, Ellsberg paradox, betweenness preferences, rank-dependent preferences, multiple priors, prospect theory, disappointment aversion, first-order risk aversion, ambiguity aversion, Knightian uncertainty

Contents
25.1 Experimental Paradoxes 652
25.2 Betweenness Preferences 658
Alternative Preferences

25.3 Rank-Dependent Preference 663
25.4 First-Order Risk Aversion 665
25.5 Ambiguity Aversion 666
25.6 Notes and References 673
There is considerable experimental evidence that individuals make choices that depart systematically from the predictions of expected utility theory. This chapter reviews some of the evidence and some of the models of decision making that have been developed to accommodate the evidence. The chapter is necessarily a brief introduction to this large literature. Applications of the models in finance constitute part of what is called behavioral finance. The first section describes the Allais and Ellsberg paradoxes and some of the experimental evidence of Kahneman and Tversky (1979) about loss aversion and prospect theory.

Most of the discussion in this chapter is atemporal, in the sense that there is no definite amount of time elapsing between a choice and its consequence. Parts of the chapter consider a single-period portfolio choice problem. While some of the theory discussed in this chapter pertains to general outcomes, the theory will be presented for outcomes that are monetary (or in units of the consumption good). Thus, we will assume outcomes are real numbers, and more is preferred to less.

Much of the discussion in this chapter concerns preferences over gambles that define gains or losses. This is a departure from most of the book, which considers preferences over consumption or wealth. Under conventional assumptions, these are equivalent concepts. Suppose we are given an initial (or reference) wealth \( w_0 \) which could itself be random, and a preference relation \( \succeq \) over random terminal wealth \( \hat{w} \). Denoting the gain or loss by \( \hat{x} = \hat{w} - w_0 \), the preference relation \( \succeq \) over terminal wealth is equivalent to a preference relation \( \succeq^* \) over gains and losses defined as

\[
\hat{x} \succeq^* \hat{x}' \iff \hat{x} + w_0 \succeq \hat{x}' + w_0.
\]

Thus, under conventional assumptions, it is simply a matter of convenience whether we discuss preferences over gains/losses or preferences over terminal wealth. In most of the chapter, we will move between the two without comment. The qualification “under conventional assumptions” is motivated by the prospect theory of Kahneman and Tversky (1979), who present evidence that the framing of a decision problem can affect the decomposition of terminal wealth \( \hat{w} \) into reference wealth \( w_0 \) and the gain/loss \( \hat{x} \). In their theory, preferences over
gains/losses are the fundamental objects, and preferences over terminal wealth are induced by preferences over gains/losses in conjunction with a coding process that determines \( w_0 \).

25.1 Experimental Paradoxes
We first discuss evidence against objective expected utility as axiomatized by von Neumann and Morgenstern (1947). We then discuss evidence against subjective expected utility as axiomatized by Savage (1954). Finally, we discuss evidence of Kahneman and Tversky (1979) that framing matters and individuals exhibit loss aversion.

The Allais Paradox and the Independence Axiom
In the theory of von Neumann and Morgenstern (1947), a person compares different gambles with known probabilities of outcomes. The defining characteristic of expected utility is that it is a utility function over gambles that is linear in probabilities. Specifically, given a set of possible outcomes \({x_0, \ldots, x_n}\), the expected utility of a gamble is

\[
\sum_{i=1}^{n} p_i u(x_i)
\]

for some utility function \( u \), where the \( p_i \) denote the probabilities. Thus, the utilities \( u(x_i) \) are the linear coefficients on the probabilities \( p_i \). Evidence that preferences are not linear in probabilities is evidence against the expected utility hypothesis.

(p.653) Consider the following pairs of gambles:

\[
A: 100\% \text{ chance of$1,000,000} \quad \text{versus} \quad B: 10\% \text{ chance of$5,000,000}
\]

\[
C: 11\% \text{ chance of$1,000,000} \quad \text{versus} \quad D: 99\% \text{ chance of$0}
\]

Various researchers, beginning with Allais (1953), have found a propensity for people to prefer \( A \) to \( B \) and \( D \) to \( C \). Apparently, when comparing \( B \) to \( A \), the \( 1\% \) chance of getting 0 seems large compared to the \( 10\% \) chance of a larger gain. However, when comparing \( D \) to \( C \), the extra \( 1\% \) chance of getting 0 seems small compared to the \( 10\% \) chance of a larger gain. These preferences are inconsistent with expected utility maximization, and this is known as the Allais paradox.
To see that the preferences are inconsistent with expected utility maximization, let \( x_1 = 0, \ x_2 = 1,000,000, \) and \( x_3 = 5,000,000. \)

If the preferences were consistent with expected utility maximization, we would have \(^1\)

\[
A > B = \quad u(x_2) > 0.01u(x_2) + 0.89u(x_3) + 0.10u(x_3), \\
D > C = \quad 0.90u(x_3) + 0.10u(x_3) > 0.89u(x_3) + 0.11u(x_3).
\]

However, adding \( 0.89[u(x_2) - u(x_3)] \) to both sides of the bottom inequality gives the reverse of the top inequality. Thus, there is no utility function \( u \) for which these preferences are consistent with expected utility maximization.

The preferences \( A > B \) and \( D > C \) violate the independence axiom used to derive objective expected utility (Herstein and Milnor, 1953). The independence axiom says that preferences regarding two gambles should be independent of whether they are mixed in the same way with a third gamble. For \( 0 < \alpha < 1 \) and two gambles \( P = (p_1, ..., p_n) \) and \( Q = (q_1, ..., q_n) \) on an outcome space \( \{x_1, ..., x_n\} \), the mixture \( \alpha P + (1 - \alpha)Q \) is the gamble that assigns probability \( \alpha p_i + (1 - \alpha)q_i \) to outcome \( i \), for \( i = 1, ..., n \). A mixture is a compound gamble (gamble over gambles): with probability \( \alpha \) we get the gamble \( P \) and with probability \( 1 - \alpha \) we get the gamble \( Q \).

The gambles in the Allais paradox can be represented as

\[
A = \alpha P + (1 - \alpha)Q \quad \text{versus} \quad B = \alpha P^* + (1 - \alpha)Q \\
C = \alpha P + (1 - \alpha)Q^* \quad \text{versus} \quad D = \alpha P^* + (1 - \alpha)Q^*,
\]

(\textbf{p.654}) where \( \alpha = 0.11, \ P = (0, 1, 0), \ P^* = (0.01/0.11, 0, 0.10/0.11), \) and \( Q^* = (1, 0, 0). \) According to the independence axiom, preferences regarding \( A \) and \( B \) should depend on preferences regarding the gambles \( P \) and \( P^* \) that occur with probability \( \alpha \), because with probability \( 1 - \alpha \) we get the same gamble \( Q \) in both cases. Thus, the preference \( A > B \) should imply \( P > P^* \). However, \( C \) and \( D \) are also mixtures of \( P \) and \( P^* \) with a gamble \( Q^* \), so the preference \( P > P^* \) should imply \( C > D \), contrary to the preferences expressed by the experimental subjects.

The Allais paradox is a special case of a more general phenomenon that researchers have observed, which is known as the common consequence effect. Suppose \( P \) is a gamble with a sure outcome, as in the Allais paradox (\$1,000,000 for sure). Thus, the choice between \( P \) and \( P^* \) is a choice between a
sure outcome and a risky gamble. The common consequence effect is that people seem to be more risk averse in comparing $P$ and $P^*$ when the gamble with which they are mixed is attractive (like $Q$ in the Allais paradox, which is $1,000,000$ for sure). They are less risk averse in comparing $P$ and $P^*$ when both are mixed with a less desirable gamble (like $Q^*$ in the Allais paradox, which is $0$ for sure). This produces the preference for $P$ over $P^*$ when $A$ and $B$ are compared but preference for $P^*$ over $P$ when $C$ and $D$ are compared.

The fact that the preferences $A > B$ and $D > C$ are nonlinear in probabilities can be seen in Figure 25.1. The triangular region (simplex) is the set of probability distributions over the outcomes $(0, 1,000,000, 5,000,000)$ represented as $\{(p_1, p_2) | p_1 \geq 0, p_2 \geq 0, p_1 + p_2 \leq 1\}$. Linearity in probabilities means there are some $u_i$ such that

\[
U(P) = u_1 p_1 + u_2 p_2 + u_3 p_3 \\
= u_1 p_1 + u_2 (1 - p_1 - p_2) + u_3 p_2 \\
= u_1 + (u_3 - u_2)p_3 + (u_1 - u_2)p_3
\]

Thus, a linear utility function is affine in $(p_1, p_2)$. The indifference curves are the parallel lines: $\{(p_1, p_2) | b_1 p_1 + b_2 p_2 = \text{constant}\}$, where $b_i = u_i - u_2$. The direction of increasing utility is up (higher $p_3$) and to the left (lower $p_1$). As the figure illustrates, the preferences $A > B$ and $D > C$ imply that at least some of the indifference curves must “fan out” from the origin rather than being parallel.\(^2\)
The Ellsberg Paradox and the Sure Thing Principle

In the Savage (1954) theory of decision making, subjective probabilities are deduced from preferences. Consider two mutually exclusive events $A$ and $B$. Suppose we offer someone a choice between two gambles: a gamble that pays $1 if $A$ occurs and 0 if $B$ occurs, or a gamble that pays $1 if $B$ occurs and 0 if $A$ occurs. If the person chooses the first gamble, then we can infer that she regards $A$ as more likely than $B$. This is a simple example of deriving a probability distribution such that expected utility represents preferences.

**Figure 25.1 Nonlinearity of utility in the Allais paradox.** This figure illustrates the Allais paradox in the probability simplex $(p_1, p_3)|p_1 \geq 0, p_3 \geq 0, p_1 + p_3 \leq 1$. The dotted lines are parallel, the one on the left passing through gambles $A$ and $B$—that is, $(0, 0)$ and $(0.01, 0.10)$—and the one on the right passing through gambles $C$ and $D$—that is, $(0.89, 0)$ and $(0.90, 0.10)$. The solid lines are indifference curves consistent with strict preference for $A$ over $B$ and $D$ over $C$. The indifference curve containing $A$ must lie above and to the left of the dotted line connecting $A$ and $B$, because $A > B$. The indifference curve containing $C$ must lie below and to the right of the dotted line connecting $C$ and $D$, because $C < D$. Therefore, the two indifference curves cannot be parallel. Hence, the preferences are not linear in probabilities.
Savage shows that preferences for gambles over events can be represented by expected utility whenever the preferences satisfy certain axioms. The most important axiom is the sure thing principle. For any event $A$, let $A^c$ denote the complement of $A$, and, as usual, let $1_A$ denote the random variable that is 1 on $A$ and 0 on $A^c$. The sure thing principle states that, for any four gambles $\bar{x}$, $\bar{y}$, $\bar{w}$, and $\bar{z}$ and any event $A$,

$$\bar{x}1_A + \bar{w}1_A^c \succeq \bar{y}1_A + \bar{w}1_A^c \Rightarrow \bar{x}1_A + \bar{z}1_A^c \succeq \bar{y}1_A + \bar{z}1_A^c.$$  
(25.1)

This is quite similar to the independence axiom. The difference is that here we are mixing over events instead of probabilities. Preference for the gamble $\bar{x}1_A + \bar{w}1_A^c$ over the gamble $\bar{y}1_A + \bar{w}1_A^c$ indicates that $\bar{x}$ is preferred to $\bar{y}$ conditional on $A$ occurring. The sure thing principle states that this preference for $\bar{x}$ conditional on $A$ occurring should imply that $\bar{x}1_A + \bar{z}1_A^c$ is preferred to $\bar{y}1_A + \bar{z}1_A^c$. The sure thing principle seems reasonable. Indeed, Savage states, “Except possibly for the assumption of simple ordering, I know of no other extralogical principle governing decisions that finds such ready acceptance.”

The sure thing principle is contradicted by the following example, due to Ellsberg (1961). Consider an urn in which there are 30 red balls and 60 balls that are either black or yellow. The relative proportion of black and yellow balls is unknown. Consider a gamble that pays a certain amount of money if a ball of a particular color is drawn. The set of states of the world is the set $\{R, B, Y\}$ of colors that can be drawn. Call the amount of money one unit, so, for example, the gamble that pays if red is drawn is $1_{\{R\}}$. The probability of winning when betting on the red ball is $1/3$, but the odds when betting on either black or yellow are unknown. It is common for people to state a strict preference for gambling on the red ball versus either the black or yellow, which we denote by $1_{\{R\}} > 1_{\{B\}}$ and $1_{\{R\}} > 1_{\{Y\}}$. Thus, people seem to prefer known odds.

Now, consider a gamble that pays the unit of money if either red or black is chosen ($1_{\{R,B\}}$) and a gamble that pays if either yellow or black is chosen ($1_{\{Y,B\}}$). The probability of winning with the latter gamble is $2/3$, but the odds for the former are unknown. It is common for people to be consistent in
preferring known odds and to prefer the yellow and black gamble here \((1_{(Y,B)} > 1_{(R,B)})\).

The preferences \(1_{(R)} > 1_{(Y)}\) and \(1_{(Y,B)} > 1_{(R,B)}\) are inconsistent with the sure thing principle and inconsistent with expected utility maximization. First, consider expected utility maximization and denote a person’s subjective probabilities by \(P\). The expected utility of \(A\) for any event \(A\) is \(u(I)P(A)\). Therefore, expected utility maximization and the additivity of probabilities imply that

\[
1_{(R)} > 1_{(Y)} \iff P(R) > P(Y) \iff P(R,B) > P(Y,B) \iff 1_{(R,B)} > 1_{(Y,B)}.
\]

(p.657) Thus, expected utility maximization (and additivity of probabilities) is inconsistent with the preferences \(1_{(R)} > 1_{(Y)}\) and \(1_{(Y,B)} > 1_{(R,B)}\) expressed in the Ellsberg example.

To see that the preferences contradict the sure thing principle, set \(A = \{R, Y\}, \tilde{x} = 1_{(R)}, \tilde{y} = 1_{(Y)}, \tilde{w} = 0\), and \(\tilde{z} = 1\). Then,

\[
\tilde{x}1_{A} + \tilde{w}1_{A'} = 1_{(R)} > 1_{(Y)} = \tilde{y}1_{A} + \tilde{w}1_{A'},
\]

but

\[
\tilde{x}1_{A} + \tilde{z}1_{A'} = 1_{(R,B)} < 1_{(Y,B)} = \tilde{y}1_{A} + \tilde{z}1_{A'}.
\]

Framing, Loss Aversion, and Prospect Theory

Kahneman and Tversky (1979) present evidence that people may sometimes not maximize any utility function over outcomes. Instead, they argue that people look at gains and losses rather than final outcomes. Furthermore, they argue that how people identify gains and losses can depend on how gambles are framed. They report responses of experimental subjects to the following scenarios:

(i) In addition to whatever you own, you have been given $1,000. You are now asked to choose between

\[
A: \{50\% \text{ chance of $1,000} \} \quad \text{versus} \quad B: \{100\% \text{ chance of $500} \}
\]

(ii) In addition to whatever you own, you have been given $2,000. You are now asked to choose between

\[
C: \{50\% \text{ chance of $0} \} \quad \text{versus} \quad D: \{100\% \text{ chance of $-500} \}.
\]
A majority of the subjects chose B in case (i) and C in case (ii). This is inconsistent with any preference relation over terminal wealth gambles. The preference for B in case (i) means a preference for $1,500 with certainty over a 50-50 gamble with outcomes $2,000 and $1,000. Of course, this is consistent with risk aversion. However, the preference for C in case (ii) means that the gamble with outcomes $2,000 and $1,000 is preferred to the certain outcome $1,500. Similar results have been obtained by other experimenters. The difference between (i) and (ii) is obviously that the outcomes are framed as gains relative to a reference wealth level in (i) but as losses relative to a reference wealth level in (ii).

Kahneman and Tversky (1979) suggest that choices are determined by preferences defined over gains and losses. A coding process determines the translation (p.658) of an outcome into a gain/loss, and this coding process is affected by framing.³ Hence, preferences over terminal wealth, which are induced by preferences over gains/losses and the coding process, are affected by framing. They suggest further that the utility function over gains/losses is concave over gains (reflecting risk aversion, as in the preference for B over A in the example of the preceding section) and convex over losses (reflecting risk seeking, as in the preference for C over D). In addition, they hypothesize that the utility function is steeper in the realm of losses than in the realm of gains, arguing that aversion to losses is greater than the desire for gains. An example of such a utility function is

\[
u(x) = \begin{cases} \frac{x^{1-p}}{1-p} & \text{if } x > 0, \\ -\frac{\gamma x^{1-p}}{1-p} & \text{if } x < 0, \end{cases}\]

(25.4)

for 0 < p < 1 and \( \gamma > 1 \), where x denotes the gain or loss rather than the resulting wealth level. The Kahneman-Tversky theory is called prospect theory.

25.2 Betweenness Preferences
There are two main generalizations of expected utility theory accommodating the Allais paradox that have been applied in the finance literature. One ("betweenness") retains the linearity of indifference curves but does not require them to be parallel. The other ("rank-dependent preferences") allows
nonlinear indifference curves. Rank-dependent preferences are discussed in Section 25.3.

The betweenness axiom is that

\[ P > Q = P > \alpha P + (1-\alpha)Q > Q \quad \text{and} \quad P - Q = P - \alpha P + (1-\alpha)Q - Q \]

for all gambles \( P \) and \( Q \) and \( 0 < \alpha < 1 \). This implies that indifference curves are linear;\(^4\) however, unlike the stronger independence axiom, it does not imply that indifference curves are parallel. Indifference curves can fan out or fan in or do both in different regions of the simplex in Figure 25.1.

A preference relation over gambles on a finite outcome space satisfies the betweenness axiom (and monotonicity and continuity axioms) if and only if (p.659) there is a utility function \( U \) over gambles with \( U \) taking values between 0 and 1 and a function \( u: X \times [0, 1] \rightarrow \mathbb{R} \) that is strictly monotone in \( x \) such that

(a) For all gambles \( P \) and \( Q \), \( P \geq Q \) if and only if \( U(P) \geq U(Q) \),

and

(b) For all gambles \( P \),

\[
U(P) = \mathbb{E}[u(x, U(P))].
\]

(25.5)

In (25.5), the expectation is with respect to the distribution \( P \) of the outcome \( x \). The utility \( U(P) \) of a gamble \( P \) is defined implicitly as the solution of (25.5).

Of course, if the function \( u \) depends only on the outcome \( x \) and not on \( U(P) \), then this is standard expected utility. Along an indifference curve—that is, fixing \( U(P) \)—the utility function \( x \mapsto u(x, U(P)) \) represents preferences just as in standard expected utility theory. Consequently, each indifference curve is linear. However, the utility function \( x \mapsto u(x, U(P)) \) changes when we change indifference curves (because \( U(P) \) changes), so the indifference curves need not be parallel.

The utility function \( U \) is monotone in the sense that \( U(P) \geq U(Q) \) if \( P \) first-order stochastically dominates \( Q \). If \( u \) is concave in \( x \), then the preferences are risk averse, in the sense of aversion to mean-preserving spreads.

**Weighted Utility**
Weighted utility is a special case of betweenness preferences. Weighted utility $U(p)$ is defined implicitly by

$$v(U(p)) = \frac{E^p[\lambda(x)|x]}{E^p[\lambda(x)]}$$

(25.6)

for a strictly monotone function $v$ and a positive function $\lambda$. For each outcome $x$, let $\delta_x$ denote the gamble that produces outcome $x$ with probability 1. From (25.6),

$$v(U(\delta_x)) = \frac{\lambda(x)}{\lambda(\delta_x)} = v(x),$$

which implies $U(\delta_x) = x$. Thus, we have normalized utility so that the utility of a sure outcome is the outcome itself.\(^5\) In fact, this normalization means that the utility of any gamble $p$ is the certainty equivalent of $p$. To see this, note that the certainty equivalent of a gamble $p$ is the outcome $x$ such that $U(\delta_x) = U(p)$. From $x = U(\delta_x)$, we obtain $U(p) = x$.

(p.660) If there is a finite number $n$ of outcomes, then we can write (25.6) as

$$v(U(p)) = \sum_{i=1}^{n} \hat{p}_i v(x_i),$$

where

$$\hat{p}_i = \frac{x_i \lambda(x_i)}{\sum_{i=1}^{n} x_i \lambda(x_i)}.$$  

Note that the $\hat{p}_i$ are nonnegative and sum to 1, so we can regard them as transformed probabilities.\(^6\) For example, if $\lambda$ is a decreasing function, then the $\hat{p}_i$ overweight low values of $x_i$ compared to the objective probabilities $p_i$. Weighted utility can be written in the general betweenness form (25.5) by defining\(^7\)

$$u(x, a) = a + \lambda(x)v(x) - v(a).$$

(25.7)

To see that weighted utility with this definition of $u$ satisfies (25.5), note that

$$E^p[u(x, U(p))] = U(p) + E^p[\lambda(x)|x] - v(U(p))E^p[\lambda(x)] = U(p),$$

the last equality following from (25.6).

Disappointment Aversion

Disappointment aversion is another special case of betweenness. Disappointment-averse utility $U(p)$ is defined implicitly by
\( v(U(P)) = \frac{E[\lambda(x, U(P), x)]}{E[\lambda(x, U(P))]} \).

(25.8a)

This is similar to weighted utility, except that the weighting function \( \lambda \) depends on \( U(P) \) here in a specific manner specified in (25.8b) below. As for weighted utility, the property (25.8a) implies that \( U(P) \) is the certainty equivalent of \( P \). Thus, the second argument of \( \lambda \) in (25.8a) is the certainty equivalent. The function \( \lambda \) is defined as

\[
\lambda(x, a) = \begin{cases} 
1 + \beta & \text{if } x < a, \\
1 & \text{if } x \geq a,
\end{cases}
\]

(25.8b)

**p.661** for some \( \beta > 0 \). Thus, these preferences overweight outcomes worse than the certainty equivalent and overweight all such outcomes by the same factor \( \beta \).

As for weighted utility, disappointment-averse utility can be written in the general betweenness form (25.5) by defining

\[
u(x, a) = a + \lambda(x, a)[v(x) - v(a)].
\]

Another special case of betweenness is generalized disappointment aversion, in which the conditions \( x < a \) and \( x \geq a \) in (25.8b) are replaced by \( x < \delta a \) and \( x \geq \delta a \) for a constant \( \delta \leq 1 \).

Constant Relative Risk Aversion

CRRA betweenness preferences can be constructed as follows:

Suppose there is a utility function \( U \) and strictly monotone function \( g \) with \( g(1) = 0 \) such that

\[
E\left[g\left(\frac{X}{U(P)}\right)\right] = 0
\]

(25.9)

for each gamble \( P \). Then, as is shown below, the preferences satisfy betweenness, and the utility \( U(P) \) of any gamble \( P \) is its certainty equivalent. The preferences represented by the utility function \( U \) exhibit constant relative risk aversion. A general definition of constant relative risk aversion is that scaling the outcomes of a gamble by a positive constant scales the certainty equivalent by the same constant. This is obvious from (25.9), because scaling the numerator and denominator in the ratio in (25.9.) by the same number leaves the equality unchanged. This linear homogeneity of the certainty equivalent is equivalent to the following property: The proportion of initial wealth an individual would pay to avoid a gamble proportional to initial wealth is independent of initial
wealth (compare to Exercise 1.5). This means that if $w(1 - \eta)$ is the certainty equivalent of $w(1 + \epsilon)$ for any constant $w_0 > 0$ and a random $\epsilon$, then $\hat{w}(1 - \eta)$ is the certainty equivalent of $\hat{w}(1 + \epsilon)$ for every $\hat{w}_0 > 0$. Clearly, this is the same as the certainty equivalent being linearly homogeneous.

A CRRA weighted utility is obtained by taking

$$v(x) = \frac{1}{1 - \rho} x^{1+\rho} \text{ and } \lambda(x) = x^\gamma$$

(25.10)

for constants $\rho$ and $\gamma$ such that $\gamma \leq 0$ and $\rho \leq 1$ with at least one of these being a strict inequality. The proof that this is a CRRA utility is given below. The relative risk aversion of CRRA weighted utility should be regarded as $\rho - 2\gamma$ (p.662) (Exercises 25.3 and 25.4). A CRRA disappointment-averse utility is obtained by taking

$$v(w) = \frac{1}{1 - \rho} w^{1+\rho}$$

(25.11)

for any $\rho > 0$. Again, the proof that this is a CRRA utility is given below.

To see that (25.9) implies betweenness, define

$$u(x, a) = a + g\left(\frac{x}{a}\right).$$

(25.12)

Then,

$$E[U(x, U(P))] = U(P) + E[g\left(\frac{x}{U(P)}\right)] = U(P)$$

by virtue of (25.9). To see that $U(P)$ is the certainty equivalent of $P$, note that, for any outcome $x$, (25.9) implies

$$g\left(\frac{x}{U(P)}\right) = 0.$$

Hence, the assumption that $g$ is strictly monotone with $g(1) = 0$ implies $x/U(\delta) = 1$. This implies that the utility of any gamble is its certainty equivalent, as discussed for weighted utility.

To see that the weighted utility (25.10) has constant relative risk aversion, define

$$g(\gamma) = \lambda(\gamma^{1+\rho} - 1).$$

(25.13)
The proof that the restrictions on the parameters $\rho$ and $\gamma$ imply $g$ is strictly monotone is left for the exercises. We have

$$
E^I[\chi(U(P))] = E^I[\chi(U(P))^{1+\gamma}] - E^I[\chi(U(P))]
$$

$$
= U(P)^{1+\gamma}\frac{\partial^2 \Upsilon_{(p)}}{\partial \chi \partial \chi} - U(P)^\gamma E^I[\chi(\chi)]
$$

$$
= 0,
$$

using $\chi(U(P)) = \chi(U(P))^{1+\gamma}$ for the second equality and the definition (25.6) of weighted utility $U(P)$ for the third. Thus, (25.9) holds.

To see that (25.11) is a CRRA disappointment-averse utility, set

$$
\delta(y) = \begin{cases} 
(1+\beta)(1-\rho)(y) - 1 - \beta & \text{if } y < 1, \\
(1-\rho)(y) - 1 & \text{if } y \geq 1.
\end{cases}
$$

A calculation similar to that given for weighted utility shows that $g$ satisfies (25.9).

(p.663) Portfolio Choice and SDFs

Consider a single-period portfolio choice problem. Let $\mathbf{R}$ denote the vector of returns, and let $\mathbf{1}$ denote a vector of 1’s.

The investor’s final wealth is

$$
\tilde{w} = w_0 R + w_1 (\mathbf{R} - R_0).
$$

To emphasize the dependence of the end-of-period wealth on the portfolio, write $\mathbf{w}_n$ for $\tilde{w}$. To map this to the discussion in this chapter, the probability distribution $P$ of outcomes (final wealth) is determined by the portfolio $\pi$. It is convenient to write $U(\pi)$ instead of $U(P)$ for the utility corresponding to a portfolio $\pi$. For betweenness preferences, we have

$$
U(\pi) = \mathbb{E}[u(\mathbf{w}_n, U(\pi))].
$$

Denote partial derivatives by subscripts. The first-order condition for optimization in $\pi$ is

$$
0 = U(\pi) = \mathbb{E}[u_1(\mathbf{w}_n, U(\pi))(\mathbf{R} - R_0)] + \mathbb{E}[u_2(\mathbf{w}_n, U(\pi))U(\pi)]
$$

$$
= \mathbb{E}[u_1(\mathbf{w}_n, U(\pi))(\mathbf{R} - R_0)]
$$

(25.14)

This means that marginal utility $u_1(\mathbf{w}_n, U(\pi))$ is proportional to an SDF. Consequently, many of the asset pricing results obtained from the expected utility hypothesis can be straightforwardly generalized to betweenness preferences. This is due to the fact that, for each fixed utility level (indifference curve), betweenness preferences are the same as expected utility preferences; that is, each indifference curve is linear.
25.3 Rank-Dependent Preferences

Rank-dependent preferences do not satisfy either the independence axiom or the betweenness axiom. Rank-dependent preferences are defined by a strictly monotone function \( v \) and a strictly monotone function \( f : [0, 1] \rightarrow [0, 1] \) satisfying \( f(0) = 0 \) and \( f(1) = 1 \). Given a finite number of possible outcomes \( x_1 \prec x_2 \prec \cdots \prec x_n \), the utility of a gamble \( p \) is defined as the solution \( U(p) \) of

\[
\nu(U(p)) = \sum_{i=1}^{n} \nu(x_i) \left[ f \left( \sum_{j=1}^{i} p_j \right) - f \left( \sum_{j=1}^{i-1} p_j \right) \right].
\]

(25.15)

This equation implies \( U(\delta_i) = x_i \), so we are measuring utility in certainty equivalent terms again. If \( f(a) = a \) for all \( a \), then this is standard expected utility \( \sum p_i \nu(x_i) \). In general, rank-dependent utility is expected utility with respect to (p.664) transformed probabilities

\[
\hat{p}_i = f \left( \sum_{j=1}^{i} p_j \right) - f \left( \sum_{j=1}^{i-1} p_j \right).
\]

Note that the \( \hat{p}_i \) are nonnegative and sum to \( f(1) - f(0) = 1 \) (we adopt the usual convention that \( \sum_{i=1}^{n} p_i = 1 \)).

The transformed probabilities depend on the cumulative objective probabilities.\(^8\) If we fix \( P \) and vary the outcomes, then the transformed probability attached to any outcome depends only on the rank of the outcome within the set of outcomes. This is in contrast to weighted or disappointment-averse utility, where the transformed probability depends on the value of the weighting function \( \lambda \) at the outcome. Which outcomes are underweighted or overweighted, relative to the objective probabilities, depends on the curvature of \( f \). For example, if \( f \) is convex with \( f(0) = 0 \) and \( f(1) = 1 \), then \( f(a) \leq a \) for all \( a \).\(^9\)

Consequently, if there are two outcomes with \( p_1 = p_2 = 0.5 \), then

\[
\hat{p}_1 = f(p_1 + p_2) - f(p_2) = f(1) - f(0.5) = 1 - f(0.5) \geq 0.5
\]

and \( \hat{p}_2 = f(0.5) \leq 0.5 \). Hence, the worse outcome is overweighted.

There is experimental evidence that \( f \) should be assumed concave on \( [0, a] \) for some \( a \) and convex on \( [a, 1] \). See Starmer (2000) for references. This implies that both very low and very high outcomes are overweighted.\(^10\)
CRRA rank-dependent preferences are obtained by taking \( \nu \) to be power or log utility. Scaling outcomes by a constant does not affect any of the \( \hat{p}_i \) (because they depend only on ranks, given \( p \)), so the certainty equivalent is scaled by the same factor, just as for power or log expected utility.

(p.665) 25.4 First-Order Risk Aversion
As discussed in Section 1.2, expected utility implies approximate risk neutrality with respect to small gambles. Of course, we can obtain moderate aversion to small gambles by assuming a high degree of risk aversion. However, this implies extreme aversion to large gambles. Rabin (2000) shows the following. Suppose a person whose preferences are represented by concave expected utility will turn down a 50-50 gamble in which she loses $100 or gains $110 (that is, a gamble in which a loss of $100 and a gain of $110 occur with probability \( 1/2 \) each) given any initial wealth level. Then, the person will, for any initial wealth level, turn down a 50-50 gamble in which she loses $1,000 or gains any amount of money. This example assumes an aversion to a small gamble for all wealth levels, which may be unreasonable. Perhaps more striking is another fact shown by Rabin (2000). Suppose a person with concave expected utility will turn down a 50-50 gamble in which she loses $100 or gains $105 at any initial wealth level less than $350,000. Then, from an initial wealth of $340,000, the person will turn down a 50-50 gamble in which she loses $4,000 or gains $635,670. For similar examples in the special case of CRRA utility, see Exercise 1.2.

The tight link between aversion to small gambles and aversion to large gambles in the expected utility framework stems from the differentiability of expected utility. See, for example, the proof in Section 1.2 that the risk premium for a small gamble is proportional to the variance of the gamble when preferences are represented by expected utility. This proportionality to the variance is called second-order risk aversion. Like expected utility, weighted utility has second-order risk aversion (Exercises 25.3 and 25.4).

On the other hand, disappointment-averse utility has first-order risk aversion, due to the discontinuity in the weighting function \( \lambda \) at the certainty equivalent. For example, consider an individual with CRRA disappointment-averse utility who has initial wealth \( w \). Let \( w - \pi \) be the certainty equivalent of the
gamble paying $w + \varepsilon$ or $w - \varepsilon$ with equal probabilities, for a constant $\varepsilon$. From the definition (25.8a) of disappointment-averse utility, we have

$$\frac{1}{T^p} (w - \pi)^{1-p} = \frac{1}{T^p} (U(P))^{1-p} = \frac{1}{T^p} \frac{0.5(w + \varepsilon)^{1-p} + 0.5(w - \varepsilon)^{1-p}}{1 + \beta}.$$ 

Solving for $\pi$, a straightforward calculation gives

$$\pi(0) = \frac{\beta \varepsilon}{2 + \beta},$$

so the risk premium for a small gamble is approximately $\beta \varepsilon/(2 + \beta)$. Therefore, the risk premium is approximately proportional to the standard deviation $\varepsilon$ as the gamble becomes small. See Exercise 25.5 for another example.

**25.5 Ambiguity Aversion**

The Ellsberg experiment (Section 25.1) indicates that people may dislike bets with unknown odds. This is called ambiguity aversion. Ambiguity is also known as Knightian uncertainty (see the end-of-chapter notes). There are two closely related generalizations of Savage’s theory that accommodate ambiguity aversion. One generalization replaces the subjective probability with a subjective nonadditive set function. The other replaces it with a set of subjective probabilities (multiple priors) and represents preferences by the worst-case expected utility. The optimal gamble is the one that maximizes this minimum expected utility. We first describe the multiple-priors
Alternative Preferences

approach, which is the model that has been more often used in finance. Nonadditive set functions are discussed at the end of the section.

To see how multiple priors resolve the Ellsberg paradox, let $P$ denote the class of probabilities $P$ that satisfy $P([R]) = 1/3$. Denote expectation with respect to $P$ by $E^P$. Define the utility $U$ of a gamble to be the minimum expected value of the gamble, where the minimum is taken over $P \in P$. The interpretation is that the decision maker is unsure of the probabilities of black and yellow balls, knowing only that they sum to $2/3$, and evaluates any gamble according to the worst-case distribution. For example,

$$U(1_{[B]}) = \min_{P \in P} E^P[1_{[B]}] = \min_{P \in P} P([R]) = 1/3.$$

Likewise, $U(1_{[Y]} = 0$, and $U(1_{[Y]}) = 0$, so the minimum expected utility $U$ is consistent with the preference ordering $1_{[B]} > 1_{[Y]}$ and $1_{[Y]} > 1_{[Y]}$. Also, $U(1_{[Y,R]}) = 2/3$, $U(1_{[Y,Y]}) = 1/3$, and $U(1_{[R,Y]}) = 1/3$. Thus, the minimum expected utility $U$ is also consistent with the preferences $1_{[Y,R]} > 1_{[R,Y]}$ and $1_{[Y,B]} > 1_{[R,B]}$. 

(p.667)
Nonparticipation

A simple but important consequence of ambiguity aversion is that investors may decline to participate in security markets. Suppose there is a risk-free asset with return \( R_f \) and a risky asset with return \( \tilde{R} \). Let \( \mathbf{P} \) be a class of probability distributions for the return \( \tilde{R} \) such that \( E[\tilde{R}] < R_f < E^P[\tilde{R}] \) for some \( P_1, P_2 \in \mathbf{P} \). This is a model for an investor who is unsure whether the risk premium is positive. Suppose the utility function \( u \) is concave.

For a long position \( \pi > 0 \), concavity implies that

\[
\min_{\mathbf{P}} E[u(w_R R_f + w_{\tilde{R}}(\tilde{R} - R_f))] \leq E^P[u(w_R R_f + w_{\tilde{R}}(\tilde{R} - R_f))] \leq u(w_R R_f + w_{\pi} E(\tilde{R} - R_f)) \leq u(w_R R_f).
\]

Therefore, \( \pi = 0 \) is preferred to any \( \pi > 0 \). Likewise, for a short position,

\[
\min_{\mathbf{P}} E[u(w_R R_f + w_{\tilde{R}}(\tilde{R} - R_f))] \leq E^P[u(w_R R_f + w_{\tilde{R}}(\tilde{R} - R_f))] < u(w_R R_f).
\]

Thus, \( \pi = 0 \) is optimal. Hence, ambiguity aversion can potentially explain why there are many investors who do not participate in the stock market and why many others participate only selectively, holding undiversified positions. Recall that it is rare for \( \pi = 0 \) to be optimal in the standard model, because in a single-asset model a long (short) position is optimal if the expected return is greater than (less than) the risk-free return (Section 2.2). Thus, \( \pi = 0 \) is optimal in the standard model only if the risk premium is exactly zero.

As an example, suppose there is a single risky asset and \( \mathbf{P} \) is a class of normal distributions for its return \( \tilde{R} \). Suppose there is a risk-free asset and the investor has a CARA utility function. Let \( \phi = w_{\tilde{R}} \) denote the investment in the risky asset. The investor chooses \( \phi \) to maximize the worst-case certainty equivalent:

\[
\min_{\mathbf{P}} \phi \left( E[\tilde{R}] - R_f \right) - \frac{1}{2} \phi \sigma^2(\tilde{R}).
\]

Suppose \( \mathbf{P} \) is the class of all normal distributions for which \( \mu_\phi \leq E[\tilde{R}] \leq \mu_\phi \) and \( \sigma^2_\phi \leq \sigma^2(\tilde{R}) \leq \sigma^2_\phi \) for constants \( \mu_\phi < \mu_\phi \) and \( \sigma_\phi < \sigma_\phi \).

Then, the worst-case distribution for a long position is the lowest mean \( \mu_\phi \) and the maximum variance \( \sigma^2_\phi \) and the worst-case distribution for a short position is the highest mean \( \mu_\phi \) and
the maximum variance \( \sigma_a^2 \). A long position is optimal if \( \mu_a > R_f \), a short position is optimal if \( \mu_b < R_f \), and \( \phi = 0 \) is optimal if \( \mu_a < R_f < \mu_b \).

**Euler Inequalities**

To see how the usual relationship between marginal utility and SDFs is affected by ambiguity aversion, let \( \phi^* \) be an optimal portfolio and let \( Q(\phi^*) \subset P \) be the class of worst-case distributions for \( \phi^* \). In the CARA–normal example in the preceding subsection, if \( \phi^* \neq 0 \) (the investor chooses to go long or short), then \( Q(\phi^*) \) contains only a single distribution \( P_{\phi^*} \) (the lowest mean and highest variance if long and the highest mean and highest variance if short), and the first-order condition must hold relative to that distribution—that is,

\[
E^{\phi^*}[u(\tilde{w}^* (R - R_f))] = 0,
\]

where \( \tilde{w}^* = w_0 R_f + \phi^*(\tilde{R} - R_f) \). Thus, pricing is as in a single-prior model with prior \( P_{\phi^*} \). On the other hand, if \( \phi^* = 0 \), then \( Q(\phi^*) = P \) and

\[
\min_{\phi \in Q(\phi^*)} E^{\phi}[u(\tilde{w}^* (R - R_f))] = u(w_0 R_f \mu_a - R_f) < 0 < u(w_0 R_f \mu_b - R_f) = \min_{\phi \in Q(\phi^*)} E^{\phi}[u(\tilde{w}^* (R - R_f))].
\]

Thus, in this example, the asset is not correctly priced by marginal utility relative to all of the worst-case distributions (though it is correctly priced relative to the distributions with \( E[\tilde{R}] = R_f \)). Whether \( \phi^* \) is zero or not, we have

\[
\min_{\phi \in Q(\phi^*)} E^{\phi}[u(\tilde{w}^* (R - R_f))] \leq 0 \leq \min_{\phi \in Q(\phi^*)} E^{\phi}[u(\tilde{w}^* (R - R_f))].
\]

(25.17)

Epstein and Wang (1994) call (25.17) the Euler inequalities.
Updating Multiple Priors

The issue of learning from information when there are multiple priors is complex. The fundamental object is preferences, so the basic question is how preferences are updated when new information is obtained. Consider the Ellsberg experiment, in which the numbers of black and yellow balls are unknown, and suppose an ambiguity-averse person is informed that the ball drawn is either red or yellow. How will the person now evaluate the gambles \( l_{(r)} \) and \( l_{(y)} \)? It is impossible to give a general answer to this question, because the violation of the sure thing principle means that conditional preferences—here, conditional on the event \( A = \{R, Y\} \)—are not uniquely determined. The preferences depend on what would have happened on the complementary event (when a black ball is drawn). If someone with the Ellsberg preferences gets 0 when a black ball is drawn, then she prefers the bet \( l_{(r)} \) to \( l_{(y)} \) conditional on \( A \). But if the same person gets 1 when a black ball is drawn, then she prefers \( l_{(y)} \) to \( l_{(r)} \) conditional on \( A \). See (25.1).

\(^{(p.669)}\) Gilboa and Schmeidler (1993) suggest two different updating rules. One is to preserve the ranking the gambles had when the worst possible outcome is paid on the complement of \( A \). This is called optimistic, because \( A \) is good news in this circumstance. The other is to preserve the ranking when the best possible outcome is paid on the complement of \( A \), which is called pessimistic. In the Ellsberg experiment, 0 is the worst outcome and 1 is the best, so the optimistic ranking of the gambles conditional on \( A = \{R, Y\} \) is \( l_{(r)} > l_{(y)} \) as shown in (25.3a). On the other hand, the pessimistic ranking is the reverse, as shown in (25.3b).

The different updating rules for preferences correspond to different updating rules for multiple priors (and also for nonadditive set functions). The optimistic ranking corresponds to applying Bayes’ rule to each prior and then evaluating the gambles according to the worst-case conditional distribution. The pessimistic ranking corresponds to using maximum likelihood to update the set of priors.

As an example, consider the Ellsberg experiment, taking \( P \) as before to be the set of probabilities for which \( P((R)) = 1/3 \).
Applying Bayes’ rule to each probability produces the set of conditional probabilities

\[
P(R|R, Y) = \frac{1/3}{1/3 + p},
\]

\[
P(Y|R, Y) = \frac{p}{1/3 + p},
\]

where the prior probability \( p \) of yellow ranges between 0 and 2/3. Thus, the conditional distributions assign probability between 1/3 and 1 to red and the complementary probability to yellow. The implied conditional preferences are the optimistic preferences:

\[
U(1_{[R]}|R, Y) = \min_{p \neq 0} P(R|R, Y) = \min_{0 < p < 2/3} \frac{1/3}{1/3 + p} = 1/3
\]

\[
> 0 = \min_{0 < p < 2/3} \frac{p}{1/3 + p} = \min_{p \neq 0} P(Y|R, Y) = U(1_{[Y]}|R, Y).
\]

Updating by maximum likelihood in this example implies that the set of priors shrinks to a single conditional distribution. For each prior, the likelihood of observing the event \((R, Y)\) is \(1/3 + p\), where \( p \) is the prior probability of yellow, so the likelihood is maximized at \( p = 2/3 \). For this prior, the conditional probability of red is 1/3 and the conditional probability of yellow is 2/3. The implied conditional preferences are the pessimistic preferences: \(1_{[Y]} > 1_{[R]}\).

For another example, consider a risky asset with payoff \( \bar{x} \).
Suppose there is no ambiguity about the marginal distribution of \( \bar{x} \). In particular, suppose it is normal with mean \( \mu \) and variance \( \sigma^2 \). Assume that a signal \( \bar{z} = \bar{x} + \bar{\varepsilon} \) is observed. Assume \((p.670)\) that \( \bar{\varepsilon} \) is known to be normally distributed, independent of \( \bar{x} \), and to have a zero mean, but there is ambiguity about its standard deviation \( \phi \), with any standard deviation in an interval \([\phi_a, \phi_b]\) being possible. Thus, there is ambiguity about the quality of the signal \( \bar{z} \). Applying Bayes’ rule to each prior produces the class of normal conditional distributions for \( \bar{x} \) with mean \( \mu + \beta(\bar{z} - \mu) \) and variance \((1 - \beta)\sigma^2\), where

\[
\beta = \frac{\sigma^2}{\sigma^2 + \phi^2}.
\]

(25.18)

On the other hand, the maximum likelihood estimate is

\[
\hat{\phi}^2 = \max\left(0, (\bar{z} - p)^2 - \sigma^2\right).
\]
Again, updating by maximum likelihood shrinks the set of priors to a single conditional distribution. It seems a bit unreasonable that a single realization of the signal $\hat{s}$ should completely resolve the ambiguity about its distribution. The model that has been used in finance is Bayesian updating of the set of priors.

### Dynamic Consistency

One desired characteristic of a model is that preferences be dynamically consistent, meaning that plans formulated at one date for implementation at a later date remain optimal when the later date is reached. To see the implications of ambiguity aversion for dynamic consistency, consider a three-date ($t = 0, 1, 2$) version of the model in the previous subsection.

Assume the asset pays $\bar{x}$ at date 2 and pays no dividends at prior dates. Assume $\bar{x}$ is known to be normally distributed with mean $\mu$ and variance $\sigma^2$. Assume the investor consumes only at date 2. Assume a signal $\hat{s} = \bar{x} + \tilde{z}$ is observed before trade at date 1, with the ambiguity about the signal quality being the same as in the preceding subsection. This means that $\hat{s}$ is conditionally normal with mean $\mu + \beta (\bar{s} - \mu)$ and variance $(1 - \beta)\sigma^2$, where $\beta$ is defined in (25.18). Because $\phi$ is in the interval $[\phi_a, \phi_b]$, $\beta$ is in the interval $[\beta_a, \beta_b]$, where

$$
\beta_a = \frac{\sigma^2}{\sigma^2 + \phi_a^2} \quad \text{and} \quad \beta_b = \frac{\sigma^2}{\sigma^2 + \phi_b^2}.
$$

Assume the risk-free return is $R_t$ in each period. Let $\theta_t$ denote the number of shares the investor chooses to hold at date $t$.

The investor’s wealth $w$ evolves as

$$W_{t+1} = W_t R_t + \theta_t (P_{t+1} - P_t R_t),$$

(p.671) where $P_2 = \bar{x}$ and $P_1$ depends on $\hat{s}$. This model is not dynamically consistent. To see the difficulty, note that $\theta_t$ will be chosen to maximize

$$\min_{\theta_t} \mathbb{E}^g [u(W_t, \hat{s})],$$

where $\mathbb{E}^g$ denotes expectation with regard to $\hat{x}$ being normal with mean $\mu + \beta (\bar{s} - \mu)$ and variance $(1 - \beta)\sigma^2$. The worst-case $\beta$ will in general depend on $\hat{s}$, and this dependence will in general affect the choice of $\theta_t$. However, when the investor chooses $\theta_0$ and formulates a plan $s \mapsto \theta_s(s)$ at date 0, she does so to maximize $\min_{\theta} \mathbb{E}^g [u(W_0, \hat{s})]$, where now the expectation is over the
joint distribution of $\tilde{s}$ and $\tilde{x}$, with $\tilde{s}$ being normal with mean $\mu$ and variance $\sigma^2/\beta$. This minimization does not allow $\beta$ to depend on $\tilde{s}$. In particular, the worst-case $\beta$ at date 0 need not be the worst-case $\beta$ at date 1. Hence, the plan that is selected at date 0 will not in general be optimal when date 1 arrives. Furthermore, the date-0 portfolio that is optimal in conjunction with the inconsistent plan will not generally be the same as the portfolio that would be chosen if the actual choice of $\theta_i$ were anticipated correctly. In a nutshell, dynamic inconsistency is a failure of Bellman’s principle of optimality: Backward induction produces different decisions than forward planning. This phenomenon is illustrated in Exercise 25.6, which analyzes this model with a risk-neutral representative investor (a model due to Epstein and Schneider, 2008).

We must either give up dynamic consistency or modify the model. To modify the model, we can start with the dynamic programming solution and change the forward-planning model to match it. Thus, in the example, we can take the marginal distribution of $\tilde{s}$ to be normal with mean $\mu$ and variance $\sigma^2/\beta_0$ for $\beta_0 < \beta < \beta_0$, and take the conditional distribution of $\tilde{x}$ given $\tilde{s}$ to be normal with mean $\mu + \beta_0 (\tilde{s} - \mu)$ and variance $(1 - \beta_0) \sigma^2$, with the investor believing that all (measurable) functions $\beta_0 : R \to [\beta_0, \beta_0]$ are possible. In this modified model, when the investor formulates a plan at date 0, she maximizes the worst case over all constants $\beta$ and functions $\beta$. The modified model possesses the property of rectangularity defined by Epstein and Schneider (2003), who show that rectangularity implies dynamic consistency when the set of priors is updated by Bayes’ rule. The modified model may seem artificial, but a multiple-priors model (like expected utility) is always an “as if” construction—solving a max-min problem produces the same decisions an Ellsberg agent would make, but it is not necessary that the agent literally believe in the multiple priors. In this example, we can take the view that the backward induction with ambiguity represents the investor’s preferences. If so, then the modified forward-planning model is simply another way to represent them.

(p.672) Nonadditive Set Functions
To describe the other generalization of Savage’s theory, let \( \phi \) be a function of events \( A \subseteq \Omega \) with the properties that \( \phi(\emptyset) = 0 \), \( \phi(\Omega) = 1 \), and \( \phi(A) \leq \phi(B) \) if \( A \subseteq B \). Additivity is the property that \( \phi(A \cup B) = \phi(A) + \phi(B) \) when \( A \) and \( B \) are disjoint sets. Probabilities are additive (in fact, they are additive for countable unions of disjoint sets). However, here we do not require additivity. A nonadditive set function \( \phi \) consistent with the Ellsberg preferences—in the sense that, for all sets \( A_n \), \( A_1 \subseteq \{R, B, Y\} \), \( 1_{A_n} \geq 1_{A_1} \) if and only if \( \phi(A_n) \geq \phi(A_1) \)—is the set function \( \phi(\emptyset) = 0 \), \( \phi(\{R\}) = 1/3 \), \( \phi(\{B\}) = \phi(\{Y\}) = 0 \), \( \phi(\{R, B\}) = \phi(\{R, Y\}) = 1/3 \), \( \phi(\{Y, B\}) = 2/3 \), \( \phi(\{R, B, Y\}) = 1 \). Additivity fails because \( \phi(\{Y, B\}) = 2/3 > 0 = \phi(\{Y\}) + \phi(\{B\}) \), and this breaks the chain of implications (with \( P = \phi \)) in (25.2).

A set function \( \phi \) is said to be ambiguity averse if

\[
\phi(A \cup B) + \phi(A \cap B) \geq \phi(A) + \phi(B)
\]

for all events \( A \) and \( B \). Given \( \phi \), let \( P \) be the set of additive set functions \( P \) satisfying \( P(A) \geq \phi(A) \) for all events \( A \). The set \( P \) is called the core of \( \phi \). Then, for all events \( A \),

\[
\phi(A) = \min_{P \in P} P(A).
\]

For example, with \( \phi \) as in the preceding paragraph for the Ellsberg experiment, \( P \) is the class of probabilities \( P \) satisfying \( P(\{R\}) = 1/3 \), as discussed earlier in the multiple-priors resolution of the Ellsberg paradox. The Ellsberg preferences are represented both by the nonadditive set function \( \phi \) and by its core \( P \).

If \( \phi \) is an ambiguity-averse set function and \( P \) is the core of \( \phi \), then, for any measurable function \( u \) and gamble \( \bar{x} \),

\[
E^\phi_u(\bar{x}) = \min_{P \in P} E^P_u(\bar{x}),
\]

where the superscript denotes the set function with respect to which the “expectation” is taken. Thus, a utility function \( u \) defines the same preferences over gambles, whether we use it with the nonadditive set function \( \phi \) or with the core \( P \) of \( \phi \). The “expectation” of a function \( u \) with respect to a nonadditive set function \( \phi \) is defined as

\[
E^\phi_u(\bar{x}) = \int_0^\infty \phi(\omega | u(\bar{x}(\omega)) \geq a) \, da - \int_0^\infty \left[ 1 - \phi(\omega | u(\bar{x}(\omega)) \geq a) \right] \, da.
\]

(p.673) This is called the Choquet integral.\(^\text{13}\)
25.6 Notes and References

Another example of framing given by Kahneman and Tversky (1979) concerns the example in Exercise 25.1. They posed the same example to subjects as a two-stage (compound) gamble. In the first stage, there is a 75% chance of the game ending with no payment. With 25% probability, the game continues and the subjects get the choice between A and B. However, the choice must be made before the game begins, so the actual gambles faced are C and D. A majority of subjects chose C over D when the choice was described as this compound gamble, in contrast to the dominant preference for D over C when described as a single gamble.

Other examples of loss aversion given by Kahneman and Tversky (1979) concern the examples in Exercises 25.1 and 25.2. When the gains in those examples (e.g., $3,000) were changed to losses (e.g., −$3,000) the opposite pattern of preferences was obtained (still violating the independence axiom). For example, the preferences A > B and D > C in Exercise 25.1 indicate that, when gains are probable, the sure outcome is preferred, but when gains are less probable, the riskier gamble (the gamble with the larger possible outcome) is preferred. However, when cast as losses, subjects generally preferred the riskier gamble when losses are probable (preferring to avoid a sure loss) but the safer outcome when losses are less likely.

Tversky and Kahneman (1992) extend prospect theory to incorporate ambiguity aversion and rank-dependent preferences. They recommend a utility function of the form (25.4) with a nonadditive set function φ replacing the subjective probability and with a transformation of the "probabilities" φ(x) as in rank-dependent preferences.

Benartzi and Thaler (1995) argue that prospect theory can explain the equity premium puzzle, because loss aversion combined with frequent portfolio evaluation makes people unwilling to hold equities even when the equity premium is large. The importance of frequent portfolio evaluation is that the probability of loss with equities is small over long horizons when the equity premium is substantial; hence, loss aversion would have little effect if gains/losses over long horizons were the issue. However, Benartzi and Thaler argue that even investors with long horizons are likely to evaluate...
their portfolios frequently, adjusting the reference wealth level to the current value each time, and suffer disutility each time a loss occurs. Barberis and Huang (2001) and Barberis, Huang, and Santos (2001) develop more formal dynamic models of prospect theory.

Betweenness preferences are often called Chew-Dekel preferences, in recognition of Chew (1983, 1989) and Dekel (1986). The utility representation (25.5) for betweenness preferences is due to Dekel (1986). Weighted utility is axiomatized by Chew (1983). A generalization of weighted utility that also satisfies betweenness, called semiweighted utility, is axiomatized by Chew (1989). Disappointment aversion is axiomatized by Gul (1991). Generalized disappointment aversion is due to Routledge and Zin (2010). They show that it produces countercyclical risk aversion, a large equity premium and a small and slowly varying risk-free rate. Thus, it can explain the major aggregate puzzles.

Chew and Epstein (1989) and Skiadas (1998) axiomatize recursive utility (Section 11.6) with non-expected-utility certainty equivalents, extending Kreps and Porteus (1978). Epstein and Zin (1989) develop EZW utility for betweenness preferences, replacing the certainty equivalent (11.27) with a certainty equivalent defined from (25.5). They also describe portfolio choice with betweenness preferences.

Rank-dependent preferences are axiomatized by Yaari (1987) and Segal (1990). The distinction between first- and second-order risk aversion is made by Segal and Spivak (1990). For surveys of non-expected-utility theory that are more extensive on some dimensions and for additional references, see Machina (1987), Starmer (2000), and Backus, Routledge, and Zin (2005). The last of these focuses on applications to finance and macroeconomics and is the source for Exercises 25.4(c) and 25.5.

It is common to cite Knight (1921) as making a distinction between risk and uncertainty, defining risk as a circumstance in which probabilities can be attached to events and uncertainty as a situation that is so unfamiliar that probabilities cannot be assessed. In the economics and finance literature, “Knightian uncertainty” and “ambiguity” are synonymous. However, LeRoy and Singell (1987) argue that
the distinction between risk and uncertainty is made by Keynes (1921) and that Knight intended no such distinction.

Expected utility with respect to a nonadditive set function is axiomatized by Gilboa (1987) and Schmeidler (1989). Max-min utility with multiple priors is axiomatized by Gilboa and Schmeidler (1989). The connection between nonadditive set functions and multiple priors described in Section 25.5 is due to Schmeidler (1986). The result on nonparticipation with ambiguity aversion is due to Dow and Werlang (1992), though they use the nonadditive set function approach instead of multiple priors. The Euler inequalities are derived in a more (p.675) general setting by Epstein and Wang (1994). Chen and Epstein (2002) study ambiguity aversion in continuous time. They extend stochastic differential utility (Duffie and Epstein, 1992b) to accommodate multiple priors.

The three-date asset pricing model with ambiguity aversion described in Section 25.5 is due to Epstein and Schneider (2008). The primary conclusion of that model is that there is a price discount (an ambiguity premium) even when investors are risk neutral and have no ambiguity about the marginal distribution of the asset payoff—see Exercise 25.6(c) for the precise formula. Thus, ambiguity aversion is another possible explanation of the equity premium puzzle. Illeditsch (2011) extends the Epstein-Schneider model by allowing the representative investor to be risk averse. He shows that the model exhibits portfolio inertia and excess volatility. Furthermore, he shows that the price function \( P_\delta(\xi) \) has a discontinuity. Thus, a small change in the information \( \xi \) can have a large effect on the price.

Exercises

**25.1.** Consider the following pairs of gambles:

- \( A : \frac{100\%}{25\%} \) chance of \$3,000 versus \( B : \frac{80\%}{20\%} \) chance of \$4,000 versus \( C : \frac{75\%}{25\%} \) chance of \$0 versus \( D : \frac{80\%}{20\%} \) chance of \$0.

(a) Show that an expected utility maximizer who prefers \( A \) to \( B \) must also prefer \( C \) to \( D \).

(b) Show that the preferences \( A > B \) and \( D > C \) violate the independence axiom by showing that \( C = aA + (1 - a)Q \) and \( D = aB + (1 - a)Q \) for some \( 0 < a < 1 \) and some gamble \( Q \).
(c) Plot the gambles $A$, $B$, $C$, and $D$ in the probability simplex of Figure 20.1, taking $p_1$ to be the probability of $0$ and $p_3$ to be the probability of $4,000$. Show that the line connecting $A$ with $B$ and the line connecting $C$ with $D$ are parallel.

Note: The preferences $A > B$ and $D > C$ are common. This example is due to Allais (1953) and is a special case of the common ratio effect. See, for example, Starmer (2000).

**25.2.** Consider the following pairs of gambles:

$$
A: \begin{cases} 50\% \text{ chance of } 3,000 \\ 10\% \text{ chance of } 0 \end{cases} \quad \text{versus} \quad B: \begin{cases} 45\% \text{ chance of } 6,000 \\ 55\% \text{ chance of } 0 \end{cases}
$$

$$
C: \begin{cases} 0.21\% \text{ chance of } 3,000 \\ 99.79\% \text{ chance of } 0 \end{cases} \quad \text{versus} \quad D: \begin{cases} 0.1\% \text{ chance of } 6,000 \\ 99.9\% \text{ chance of } 0 \end{cases}.
$$

Show that an expected utility maximizer who prefers $A$ to $B$ must also prefer $C$ to $D$. Note: The preferences $A > B$ and $D > C$ are common. This example is due to Kahneman and Tversky (1979).

**25.3.** Consider weighted utility. Let $\bar{\epsilon}$ have zero mean and unit variance. For a constant $\sigma$, denote the certainty equivalent of $w + \sigma \bar{\epsilon}$ by $w - \pi(\sigma)$. Assume $\pi(\cdot)$ is twice continuously differentiable. By differentiating

$$
\nu(w - \pi(\sigma)) = \frac{\nu(w + \sigma \bar{\epsilon}) - \nu(w)}{\nu(w)} = \mu(w + \sigma \bar{\epsilon} + \sigma \bar{\epsilon}),
$$

assuming differentiation and expectation can be interchanged, show successively that $\pi(0) = 0$ and

$$
\pi'(0) = \frac{\nu'(w)}{\nu(w)} - \frac{2\nu(\sigma)}{\nu(w)}.
$$

Note: This implies that for CRRA weighted utility and small $\sigma$,

$$
\pi(\sigma) \approx (\rho - 2\gamma) \text{var}(\sigma \bar{\epsilon} / w) / 2.
$$

**25.4.** Consider CRRA weighted utility.

(a) Show that $g$ in (25.13) is strictly monotone in $y > 0$—so the preferences are monotone with regard to stochastic dominance—if and only if $y \leq 0$ and $\rho \leq y + 1$ with at least one of these being a strict inequality.

(b) Show that $g$ in (25.13) is strictly monotone and concave if and only if $y \leq 0$ and $y \leq \rho \leq y + 1$ with either $y < 0$ or $\rho < y + 1$.  

Page 30 of 34

PRINTED FROM OXFORD SCHOLARSHIP ONLINE (www.oxfordscholarship.com). (c) Copyright Oxford University Press, 2018. All Rights Reserved. Under the terms of the licence agreement, an individual user may print out a PDF of a single chapter of a monograph in OSO for personal use (for details see www.oxfordscholarship.com/page/privacy-policy). Subscriber: HINARI; date: 29 November 2018
(c) Consider a lognormal gamble: \( \tilde{w} = w(1 + \tilde{\varepsilon}) \) where 
\( \log(1 + \tilde{\varepsilon}) \) is normally distributed with variance \( \sigma^2 \) and 
mean \(-\sigma^2/2\) (implying \( \mathbb{E}[\tilde{\varepsilon}] = 0 \)). Show that the certainty 
equivalent is \( w(1 - \pi) \), where 
\[\pi = 1 - e\left(\rho - 2\rho\sigma^2/2\right).\]
Note: This implies that \( \pi = (\rho - 2\rho\sigma^2/2 \) for small \( \sigma \). Compare 
Exercise 1.5.

25.5. Consider CRRA disappointment-averse utility and a random wealth \( \tilde{w} = e^{\tilde{x}} \), where \( \tilde{x} \) is normally distributed with mean \( \mu \) and variance \( \sigma^2 \). Let \( \xi \) denote the certainty equivalent of \( \tilde{w} \), and set \( \theta = \log \xi \).

\( \textbf{(p.677)} \) (a) Show that \( \theta \) satisfies the equation 
\[\theta = \mu + \frac{\sigma^2}{2} + \frac{\sigma^2}{1+\rho}\log\left[\frac{1+\rho}{1+\rho}\frac{\Phi\left(\frac{\mu+\sigma^2}{\rho}\right)}{\Phi\left(\frac{\sigma^2}{\rho}\right)}\right],\]
where \( \Phi \) denotes the standard normal distribution function.
Hint: See the calculation of \( \mathbb{E}[e^{\tilde{x}} | \tilde{x} = x] \) for a normal random 
variable \( \tilde{x} \) in Section 7.5.

(b) Let \( \mu = -\sigma^2/2 + \log w \) for a constant \( w \). (Defining the 
standard normal \( \tilde{x} = (\tilde{x} - \mu)/\sigma \), we then have \( \tilde{w} = w(1 + \tilde{\varepsilon}) \), 
where \( \tilde{\varepsilon} \) is \( \mathcal{N}(0, \sigma^2) \) has mean zero.) Define \( \pi = (w - \xi)/w \). 
(Then, \( w(1 - \pi) \) is the certainty equivalent of \( w(1 + \tilde{\varepsilon}) \) \.) Show 
numerically that \( \pi / \sigma^2 \) appears to increase without bound 
and \( \pi / \sigma \) appears to converge to a positive constant as 
\( \sigma \to 0 \).

25.6. Consider the portfolio choice model with a single risky 
asset described in Section 25.5, in which there is no ambiguity 
about the marginal distribution about the asset payoff but 
there is ambiguity about the informativeness of a signal.
Assume the investor is a representative investor; there is a 
single unit of the risky asset, and the risk-free asset is in zero 
net supply. Assume Bayesian updating of the set of priors, and 
assume the representative investor is risk neutral (but 
ambiguity averse). Let \( P_t \) denote the price of the risky asset at 
date \( t \), with \( P_2 = \tilde{x} \). Take \( R_t = 1 \). The intertemporal budget 
equation is 
\[W_{t+1} = W_t + \theta(P_{t+1} - P_t),\]
where \( \theta_t \) denote the number of shares of the risky asset chosen at date \( t \). The distribution of \( x \) conditional on \( \bar{s} \) is normal with mean \( \mu + \beta(\bar{s} - \mu) \) and variance \((1 - \beta)\sigma^2\), and the marginal distribution of \( \bar{s} \) is normal with mean \( \mu \) and variance \( \sigma^2/\beta \), for \( \beta_0 < \beta < \beta_b \). Take the backward induction (dynamic programming) approach.

(a) Suppose \( \bar{s} < \mu \). Show that \( \theta_t = 1 \) maximizes

\[
\min_{\beta} \mathbb{E}_t [W|\bar{s}] 
\]

(p.678) if and only if

\[ P_t = \mu + \beta(\bar{s} - \mu). \]

(b) Suppose \( \bar{s} > \mu \). Show that \( \theta_t = 1 \) maximizes

\[
\min_{\beta} \mathbb{E}_t [W|\bar{s}] 
\]

if and only if

\[ P_t = \mu + \beta(\bar{s} - \mu). \]

(c) Suppose that \( P_t \) depends on \( \bar{s} \) as described in the previous parts. Show that \( \theta_0 = 1 \) maximizes

\[
\min_{\beta} \mathbb{E}_0 [W_0] 
\]

if and only if

\[ P_0 = \mu - \frac{\bar{\theta}_0 \bar{x}}{\sqrt{\sigma^2 \bar{s}}} \cdot \]

Hint: The function \( P_t \) is concave in \( \bar{s} \). Hence, \( \mathbb{E}_0 [P_t(\bar{s})] < P_t(\mu) \) and the difference \( P_t(\mu) - \mathbb{E}_0 [P_t(\bar{s})] \) is maximized at the maximum variance for \( \bar{s} \).

(d) Now take the forward-planning approach. Let \( P_0 \) and \( P_t(\bar{s}) \) be as described in the previous parts. The investor chooses \( \theta_0 \) and a plan \( s \mapsto \theta_t(s) \) at date 0 to maximize

\[
\min_{\beta} \mathbb{E}_0 [W_0] = \min_{\beta} \mathbb{E}_0 [W_0 + \theta_0(P_t(\bar{s}) - P_0) + \theta_t(s)(\bar{x} - P_t(s))] 
\]

Show that the investor can achieve unbounded worst-case expected wealth. In particular, choosing \( \theta_0 = 1 \) and \( \theta_t(s) = 1 \) for all \( s \) is not optimal.

Notes:
(1.) We use the standard notation \( > \) for strict preference, \( \sim \) for indifference, and \( \geq \) for weak preference (strict preference or indifference).

(2.) The slope of an indifference curve in the probability simplex reflects risk aversion. Moving up and to the right in the simplex creates a riskier gamble (higher probabilities of the extreme outcomes \( x_i \) and \( x_j \)), so a steeper slope means that the probability of the best outcome must be increased more in order to maintain indifference when risk is increased. Thus, “fanning out” implies greater risk aversion when comparing more attractive gambles (gambles in the left portion of the simplex).

(3.) The coding that translates outcomes into gains/losses is said to be part of a more general editing process that includes, for example, rounding probabilities.

(4.) Technically, indifference curves are convex in the mixture space. In the probability simplex of Figure 20.1, indifference curves are line segments.

(5.) We can equivalently define weighted utility by replacing \( \nu(U(P)) \) on the left-hand side of (25.6) by \( U(P) \). With this definition, \( \nu(x) \) is the utility of a sure outcome \( x \).

(6.) In general, \( \lambda(x)/\mathbb{E}[^{\lambda(x)}] \) is the Radon-Nikodym derivative \( d\hat{P}/dP \) of a probability \( \hat{P} \) with respect to \( P \), so \( \nu(U(P)) = \mathbb{E}[^{\lambda(x)}] \).

(7.) There are many different ways to do this: We can take \( U(x,a) = a + f(a)\lambda(x)\nu(x) - \nu(a) \) for any function \( f \).

(8.) We can define rank-dependent utility equivalently as

\[
\nu(U(P)) = \sum_{i=1}^{n} \nu(x)[g(\sum_{j=1}^{i} p_j) - g(\sum_{j=1}^{i+1} p_j)]
\]

((25.16))

by setting \( g(a) = 1 - f(1 - a) \).

(9.) We have \( f(a) = f((1 - a) \times 0 + a \times 1) \leq (1 - a) \times f(0) + a \times f(1) - a \).

(10.) For example, if there are four possible outcomes with equal objective probabilities and \( f \) is strictly concave on \([0, 1/2]\) and strictly convex on \([1/2, 1]\), then \( \hat{p}_1 > \hat{p}_2 \) and \( \hat{p}_4 > \hat{p}_3 \).
(11.) The variance of $\bar{x}$ is $\sigma^2 + \phi^2$ and $\beta = \sigma^2 / (\sigma^2 + \phi^2)$, so the variance of $\bar{x}$ is $\sigma^2 / \beta$.

(12.) If $\phi$ were additive, we would have
$$\phi(A \cup B) + \phi(A \cap B) = \phi(A) + \phi(B).$$

(13.) To interpret the Choquet integral, consider the example of $u$ being nonnegative and monotone and $x$ being uniformly distributed on $[0, 1]$ relative to a probability $\phi$ (i.e., $\phi(\omega; (\omega) \leq x)) = x$ for $x \in [0, 1]$). If we graph the function $u$ with $x$ on the horizontal axis and $u(x)$ on the vertical, then the Choquet integral is the area between the utility function and 0, computed by integrating over the vertical axis, instead of by integrating over the horizontal axis, as we would normally do.