Chapter 1 — Utility Theory: An Introduction

In much of Finance and Economics, utility functions are taken as primitives. This can lead to confusion when the discussion expands beyond the classical models to areas such as "Behavioral" Finance because it is not clear exactly where the differences arise. Virtually all of Finance (and Economics) is behavioral in the sense that it is about behavior. What distinguishes the "Behavioral" subfields is that the agents are not rational in the sense used in the classical models. Their choices might violate the Independence Axiom of choice or they may not update beliefs in a Bayesian manner, for example. To see where that irrationality arises, we must understand what lies behind utility theory — and that is the theory of choice. This is an enormous field of study. What is provided here is merely an introduction to that large subject.

Preferences and Ordinal Utility

The two primitives in the theory of choice are a set, $\mathcal{X}$, of goods, attributes, or other features from which a selection is to be made and a preference relation on this set, denoted by $\succcurlyeq$, The relation $\mathbf{x} \succcurlyeq \mathbf{y}$ is read as **x** is (weakly) preferred to **y** and means that the vector of goods **x** is at least as good as **y** so that **y** would never be strictly chosen over **x**. From these two primitives, choice theory derives a utility function which simplifies how choices can be described. A utility function is a real valued function $u(\mathbf{x})$ such that

$$u(\mathbf{x}) \ge u(\mathbf{y}) \quad \Leftrightarrow \quad \mathbf{x} \succcurlyeq \mathbf{y}\,. \tag{1}$$

This is an ordinal utility function; the only issue is whether $u(\mathbf{x})$ is greater or less that $u(\mathbf{y})$. The exact numerical values and difference between them are completely irrelevant. For example, $u(x) = x$ and $u(x) = x^2$ are equivalent provided $x > 0$. For cardinal utility functions, introduced later, the numerical values do have some meaning. Ordinal utility functions describe choices amongst certain prospects and cardinal utility describes choices amongst uncertain prospects.

The following two axioms are assumed to describe the preference relation

A1) **Completeness**: $\forall \mathbf{x}, \mathbf{y} \in \mathcal{X},\ \ \mathbf{x} \succcurlyeq \mathbf{y}$ or $\mathbf{y} \succcurlyeq \mathbf{x}$. That is, among all pairs of the choices, either the first is weakly preferred to the second or the second is weakly preferred to the first, or both.

A2) **Transitivity**: $\forall \mathbf{x}, \mathbf{y}, \mathbf{z} \in \mathcal{X},\ \ \mathbf{x} \succcurlyeq \mathbf{y}$ and $\mathbf{y} \succcurlyeq \mathbf{z} \Rightarrow \mathbf{x} \succcurlyeq \mathbf{z}$. That is, if **x** is weakly preferred to **y**, and **y** is weakly preferred to **z,** then **x** must be weakly preferred to **z**.

Axioms are supposed to be intuitively obvious truths. These two axioms possess that property. The first only insists that all comparisons can be made. Even an "I don't know" answer is valid if it is interpreted as both $\mathbf{x} \succcurlyeq \mathbf{y}$ and $\mathbf{y} \succcurlyeq \mathbf{x}$. This preference is perfectly consistent with the agent sometimes choosing **x** and sometimes **y**. The transitivity axiom would seem equally obvious; nevertheless, it is possible to get people to express choices that are not transitive. This could be due to complicated choice sets, preferences that change over time, or a number of other considerations. Preferences that change over time are not, strictly speaking a violation of the axiom which is atemporal. It does, however, complicate verifying its truth. One of the most famous violations of transitivity, though it applies to group not individual choice, is the Arrow Impossibility Theorem about voting. We will ignore any such complications. This seems reasonable in Finance where our choices will most often be monetary.

A good set of axioms should have two properties beyond being obvious. They should be consistent; that is, they must not contradict each other. And they should be independent; that is, none should be redundant and derivable from the others. Consistency is an obviously important property without which the entire structure of any theory would probably collapse. Independence is more of an esthetic property of parsimony. The two axioms given above are consistent and independent.

An example of a redundant axiom for a preference relation, which is nevertheless often separately stated, is *reflexivity*: $\forall \mathbf{x} \in \mathcal{X}, \mathbf{x} \succcurlyeq \mathbf{x};$ that is, any choice is at least as good as itself. This axiom is certainly obvious, but it is embedded in the completeness axiom which insists that we be able to compare **x** to itself. By completeness at least one of $\mathbf{x}_1 \succcurlyeq \mathbf{x}_2$ or $\mathbf{x}_2 \succcurlyeq \mathbf{x}_1$ must be true. So if $\mathbf{x}_1 = \mathbf{x}_2 \equiv \mathbf{x}$, we must have $\mathbf{x} \succcurlyeq \mathbf{x}$.

The weak preference relation induces several related concepts. First $\mathbf{x} \preccurlyeq \mathbf{y}$ has the obvious meaning that $\mathbf{y} \succcurlyeq \mathbf{x}$. Second, $\mathbf{x} \not\succcurlyeq \mathbf{y}$ means that **x** is not weakly preferred to **y**. This is more commonly written as $\mathbf{y} \succ \mathbf{x}$, read as **y** is strictly preferred to **x**. Other negated relations have similar interpretations. Third both $\mathbf{x} \succcurlyeq \mathbf{y}$ and $\mathbf{y} \succcurlyeq \mathbf{x}$ together mean that neither **x** nor **y** is strictly preferred; this indifference is usually written $\mathbf{x} \sim \mathbf{y}$. Given these derived relations, it can be shown that together $\succ$ and ~ comprise a complete ordering such that exactly one of $\mathbf{x} \succ \mathbf{y}$, $\mathbf{y} \succ \mathbf{x}$, or $\mathbf{x} \sim \mathbf{y}$ is true. Also $\succ$ and ~ are transitive relations, and together with $\succcurlyeq$ the three relations form a transitive set with $\succ$ holding between the first and last if it appears anywhere in the comparison; otherwise $\succcurlyeq$ holds if just $\succcurlyeq$ and ~ appear.

The two axioms given above are sufficient to prove the existence of a real-valued ordinal utility as characterized in equation (1) provided the choice set, $\mathcal{X}$, is finite. This point should be obvious. With a finite set, there are only a finite number of comparisons that can be made. Therefore, each element **x** can be ranked (including possible ties) and a real number can be assigned for that rank. The same is true if $\mathcal{X}$ is countable or even uncountably infinite if it is one dimensional. The latter should again be obvious as the preference relation $\succcurlyeq$ among the choices has the same properties as the relation $\geq$ among the real utility values.

If $\mathcal{X}$ has two or more dimensions and is uncountable, a third axiom is required to guarantee the existence of a real valued utility function satisfying (1), and, unfortunately, it does not have quite the same intuitive appeal of the previous two.

(A3o) **Continuity**: For every $\mathbf{x} \in \mathcal{X}$, the subsets of strictly preferred and strictly worse choices are both open.

The necessity of this continuity axiom can best be illustrated by the counter-example of lexicographic preferences, $(x_1, x_2) \succ (y_1, y_2)$ if $x_1 > y_1$ or if $x_1 = y_1$ and $x_2 > y_2$. Under lexicographic preferences both components are valued, but the first is of primary importance and no increment of the second can compensate for a shortfall in the first.[1] Here the subset preferred to (x_1^*, x_2^*) is not open as it includes the boundary subset with $x_1 = x_1^*, x_2 > x_2^*$.

Can a utility function describe this set of preferences? Any possible utility function certainly cannot be a continuous one. Consider the two-dimensional case. Note that for any point (x, y) with utility $u(x, y) = u_0$, there can be no other point with the same utility because all other points differ in one or both components. Now consider three points with

$$(x, y+\delta) \succ (x, y) \succ (x-\delta, y)\,. \tag{2}$$

For any path connecting the two outer points utility changes from a value above u_0 to a value

[1] "Lexicographic" means like a lexicon or dictionary. In a dictionary, all the words beginning with A come first. Among the words beginning with A, those beginning AA come before those beginning AB, etc. So the first letter is of paramount importance, then the second letter, etc.

below u_0, but unless the path passes through (x, y) utility can never be u_0. This means that the utility function must be discontinuous. Moreover, this is true at every point so the utility function must be discontinuous everywhere. But an increasing real-valued function cannot be discontinuous everywhere so no utility function exists for this set of preferences.

With the continuity axiom, we have the following theorem even with multi-dimensional uncountable sets.

Theorem 1.1: Ordinal Utility. For any preference relation satisfying axioms (A1) through (A3o) defined over a closed, convex[2] set of choices, $\mathcal{X}$, there exists a continuous ordinal utility function u mapping $\mathcal{X}$ to the reals satisfying

$$\begin{aligned} u(\mathbf{x}) > u(\mathbf{y}) \quad &\Leftrightarrow \quad \mathbf{x} \succ \mathbf{y} \\ u(\mathbf{x}) = u(\mathbf{y}) \quad &\Leftrightarrow \quad \mathbf{x} \sim \mathbf{y}\,. \end{aligned} \tag{3}$$

Furthermore, the utility function is unique up to a continuous increasing transformation. That is, if both $u(\mathbf{x})$ and $v(\mathbf{x})$ satisfy (3), then $v(\mathbf{x}) = f(u(\mathbf{x}))$, where f is an increasing continuous function also satisfies (3). ■

This theorem is obvious with a discrete set as that can simply be put into order based on the preference and then utilities of 1, 2, … assigned. The proof for a continuous set is slightly more involved, but the intuition should be clear because the preference ordering $\succcurlyeq$ has the same properties as the $\geq$ relation among the real values of the utility function. A formal proof can be found in many standard economics texts.

Properties of Ordinal Utility

The utility function is an ordinal one and, apart from continuity guaranteed by axiom A3o, it contains no more information than the ordering relation as indicated in (3). No meaning can be attached to the utility level other than that inherent in the "greater than" relation in arithmetic. It is not correct to say $\mathbf{x}$ is twice as good as $\mathbf{y}$ if $u(\mathbf{x}) = 2u(\mathbf{y})$. Likewise, the conclusion that $\mathbf{x}$ is more of an improvement over $\mathbf{y}$ than $\mathbf{y}$ is over $\mathbf{z}$ because $u(\mathbf{x}) - u(\mathbf{y}) > u(\mathbf{y}) - u(\mathbf{z})$ is also faulty. If particular utility function $u(\mathbf{x})$ is a valid representation of some preference ordering, then so is $v(\mathbf{x}) \equiv f\big(u(\mathbf{x})\big)$ where $f(\cdot)$ is any strictly increasing function. This is not true for cardinal utility functions introduced later.

To proceed further we now assume that $\mathcal{X}$ is a continuous set and that the utility function chosen to represent it is twice continuously differentiable. This assumption is one of technical convenience, but it admits to the use of marginal utility, a very important concept in Finance.

Using marginal utility, a utility function can be characterized by its indifference or isoutility surfaces. These capture all that is relevant in a given preference ordering but are invariant to any strictly increasing transformation. An indifference surface is the set of all $\mathbf{x}$ of equal utility; that is, $\{\mathbf{x} \in \mathcal{X} \mid \mathbf{x} \sim \mathbf{x}_0\}$ or, equivalently, $\{\mathbf{x} \in \mathcal{X} \mid u(\mathbf{x}) = u(\mathbf{x}_0)\}$. The directional slopes of the indifference surface determine the marginal rates of substitution. The marginal rate of substitution between x_i and x_j at any point $\mathbf{x}$ is the increase in x_i needed to offset a decrease in x_j. This is a movement confined to an indifference surface. Using the implicit function theorem gives

[2] There is no real need that the set be convex. The choices $\mathcal{X}$ can be embedded into a larger set which is convex with u defined on the larger set. The choice problem can be restricted to the original set, but this does not invalidate the measuring of u on unavailable choices.

$$\text{MRS}(\mathbf{x}_0) \equiv -\left.\frac{dx_i}{dx_j}\right|_{u=u(\mathbf{x}_0)} = \left.\frac{\partial u/\partial x_j}{\partial u/\partial x_j}\right|_{\mathbf{x}=\mathbf{x}_0} . \tag{4}$$

An idea closely related to the marginal rate of substitution is the elasticity of substitution. This is defined as the percentage change in relative proportions in the two goods per a given percentage change in the marginal rate of substitution,

$$\eta_{ij} \equiv \frac{(dx_i/x_i)\big/(dx_j/x_j)}{d\text{MRS}\big/\text{MRS}} \tag{5}$$

Denoting the partial derivative of u with respect to x_i by subscripts, the change in the marginal rate of substitution is

$$d(-dx_i/dx_j) = d(u_j/u_i) = \frac{\partial(u_j/u_i)}{\partial x_i}dx_i + \frac{\partial(u_j/u_i)}{\partial x_j}dx_j = \left[\frac{u_{ij}}{u_i} - \frac{u_j u_{ii}}{u_i^2}\right]dx_i + \left[\frac{u_{jj}}{u_i} - \frac{u_j u_{ii}}{u_i^2}\right]dx_j . \tag{6}$$

The change in the relative proportions is $d(x_i/x_j) = dx_i x_j - x_i dx_j/x_j^2$, and the isoquant is $dx_i = -(u_j/u_i)dx_j$. Combining all these gives the elasticity of substitution

$$\eta_{ij} = \frac{u_i u_j (u_i x_i + u_j x_j)}{x_i x_j (2u_{ij}u_i u_j - u_i^2 u_{jj} - u_j^2 u_{ii})} \tag{7}$$

If the elasticity is less than 1, the demand is said to be inelastic, while if the elasticity is greater than one it is said to be elastic. If the elasticity is ∞ or 0, the function is said to be perfectly elastic or perfectly inelastic, respectively.

As promised, the MRSs or elasticities capture what is important ignoring monotonic transformations because for the equivalent utility function $v(\mathbf{x}) = f(u(\mathbf{x}))$, the marginal rate of substitution is the same

$$-\left.\frac{dx_i}{dx_j}\right|_{v=v(\mathbf{x}_0)} = \left.\frac{\partial v/\partial x_j}{\partial v/\partial x_j}\right|_{\mathbf{x}=\mathbf{x}_0} = \left.\frac{f'(u)\partial u/\partial x_j}{f'(u)\partial u/\partial x_j}\right|_{\mathbf{x}=\mathbf{x}_0} = \left.\frac{\partial u/\partial x_j}{\partial u/\partial x_j}\right|_{\mathbf{x}=\mathbf{x}_0} \tag{8}$$

This can also be verified for the elasticity though it is simpler to note from the definition (5) that at a particular $\mathbf{x}_0$, the elasticity depends only on the marginal rate of substitution, which is an ordinal property. So elasticity, too, is an ordinal property.

For additive utility $u(x_i, x_j) \equiv u_i(c_i) + u_j(c_j)$, the elasticity can be simplified to

$$\eta_{ij} = \frac{u_i' u_j' (u_i' c_i + u_j' c_j)}{-c_i c_j (u_i'^2 u_j'' + u_j'^2 u_i'')} \quad \Rightarrow \quad \frac{1}{\eta_{ij}} = \frac{-c_i c_j (u_i'^2 u_j'' + u_j'^2 u_i'')}{u_i' u_j' (u_i' c_i + u_j' c_j)} = \frac{c_i u_i' \dfrac{-c_j u_j''}{u_j'} + c_j u_j' \dfrac{-c_i u_i''}{u_i'}}{c_i u_i' + c_j u_j'} . \tag{9}$$

The ratio $-cu''/u'$ is the Arrow-Pratt measure of relative risk aversion. It is discussed in detail below. Now it is merely noted that the elasticity of substation is equal to a weighted average of the c_i and c_j relative risk aversions. This is true even though the elasticity of substitution is defined even when there is no risk. The relation between the elasticity and risk aversion is simply a property of the utility functions. The distinction between elasticity and risk aversion will be important later and is one reason why care must be used when assuming additive utility functions.

Commonly Used Ordinal Utility Functions

Probably the most commonly used form of utility is the constant elasticity of substitution (CES) function[3]

$$u(\mathbf{x}) = \left(\sum \beta_i x_i^{(\eta-1)/\eta}\right)^{\eta/(\eta-1)} \quad \text{or} \quad v(\mathbf{x}) = \tfrac{\eta}{\eta-1}\sum \beta_i x_i^{(\eta-1)/\eta} \quad \text{with} \quad \eta > 0, \beta_i > 0, \sum \beta_i = 1. \qquad (10)$$

The two forms are equivalent using the monotonic transformation $v = f(u) = \rho^{-1}u^{\rho}$. The second form of CES utility is additively separable so that the marginal utility of any good depends only on the quantity of that good alone. That this is not true for the first form shows that marginal utility is not a sufficient description of utility. The marginal rates of substitution and elasticities do not depend on the form selected. They are

$$-\frac{dx_i}{dx_j}\bigg|_{v=v(\mathbf{x})} = \frac{\beta_j}{\beta_i}\left(\frac{x_j}{x_i}\right)^{-1/\eta} \quad \text{and} \quad \eta_{ij} = \frac{u_i u_j (u_i x_i + u_j x_j)}{x_i x_j (2u_{ij}u_i u_j - u_i^2 u_{jj} - u_j^2 u_{ii})} = \eta. \qquad (11)$$

The elasticity is the same between any two goods at any current levels. This utility function is also homothetic; that is, the marginal rates of substitution depend only on the relative allocation of the two goods.

There are several special case of CES utility, linear, Cobb-Douglas, and Leontief. For linear utility with infinite elasticity ($\eta = \infty$), the marginal rates of substitution are constant so the goods are perfect substitutes, and the consumer is willing to swap βj units of x_i for β_i units of x_j regardless of the quantities. For all other CES utilities, the marginal rate of substitution increases as the ratio x_j/x_i drops. The more x_i that the consumer has, the more he must receive per unit of x_j given up to maintain his level of utility.

As $\eta \to 0$, the marginal rate of substitution becomes infinite when $x_i > x_j$ or zero when $x_i < x_j$. This means the goods are strict complements; that is, utility does not increase at all when x_i increases if $x_i \geq x_j$. This is Leontief utility; the function can be expressed $u(\mathbf{x}) = \min_i\{\beta_i x_i\}$.[4]

For Cobb-Douglas utility, $\eta = 1$. The functional form can be determined using L'Hosptial rule for the limit of an ordinally modified version in (10)

$$\lim_{\eta\to 1}\sum \beta_i \tfrac{\eta}{\eta-1}(x_i^{(\eta-1)/\eta} - 1) = \sum \beta_i \ell\text{n}\, x_i \qquad \text{or} \qquad v(\mathbf{x}) = \prod x_i^{\beta_i}\,. \qquad (12)$$

Two extensions of Cobb-Douglas utility are Stone-Geary and translog utility. Stone-Geary or translated Cobb-Douglas has a minimum or subsistence requirement in each good

$$u(\mathbf{x}) = \sum \beta_i \ell\text{n}(x_i - \underline{x}_i) \qquad \text{or} \qquad v(\mathbf{x}) = \prod (x_i - \underline{x}_i)^{\beta_i} \quad \text{with} \quad \beta_i > 0, \sum \beta_i = 1. \qquad (13)$$

The marginal utility of good i becomes infinite and utility is zero at $\underline{x}_i$. CES utility can be similarly modified to the translated form of

$$u(\mathbf{x}) = \left(\sum \beta_i (x_i - \underline{x}_i)^{\eta/(\eta-1)}\right)^{(\eta-1)/\eta} \quad \text{or} \quad v(\mathbf{x}) = \tfrac{\eta-1}{\eta}\sum \beta_i (x_i - \underline{x}_i)^{\eta/(\eta-1)} \quad \beta_i, \eta > 0. \qquad (14)$$

---

[3] The usual constraint $\sum\beta_i = 1$ can ignored with no loss of generality as an allowed monotonic transformation. This is also true for Stone-Geary utility below. It is possible to have utility functions in which some elasticities are negative, $\eta_{ij} < 0$. However, not all cross good elasticities can be negative and, for CES utility all elasticities are the same, $\eta_{ij} = \eta$. So η must be positive.

[4] Leontief utility is only weakly monotonic and preferences are only weakly convex so some of the standard results do not apply or may only apply in a weakened sense.

The translog utility function is

$$u(\mathbf{x}) = \sum_i \beta_i \ell\text{n}\, x_i + \sum_i \sum_{j\neq i} \gamma_{ij} \ell\text{n}\, x_i \cdot \ell\text{n}\, x_j \quad \text{with} \quad \gamma_{ij} = \gamma_{ji} . \tag{15}$$

Cobb-Douglas is a special case with $\gamma_{ij} = 0$. The marginal rates of substitution and elasticities for the translated CES utility are

$$-\frac{dx_i}{dx_j}\bigg|_{v=v(\mathbf{x})} = \frac{\beta_j (x_j - \underline{x}_j)^{-1/\eta}}{\beta_i (x_i - \underline{x}_i)^{-1/\eta}} \quad \text{and} \quad \eta_{ij} = \eta \left[1 - \frac{\beta_i (x_i - \underline{x}_i)^{(\eta-1)/\eta} x_i \underline{x}_j + \beta_j (x_j - \underline{x}_j)^{(\eta-1)/\eta} \underline{x}_i x_j}{x_i x_j [\beta_i (x_i - \underline{x}_i)^{(\eta-1)/\eta} + \beta_j (x_j - \underline{x}_j)^{(\eta-1)/\eta}]} \right] . \tag{16}$$

For translog utility the marginal rates of substitution and elasticities are

$$-\frac{dx_i}{dx_j}\bigg|_{v=v(\mathbf{x})} = \frac{x_i \left(\beta_j + \sum_{k\neq j} \gamma_{jk} \ell\text{n}\, x_k \right)}{x_j \left(\beta_i + \sum_{k\neq i} \gamma_{ik} \ell\text{n}\, x_k \right)} \quad \text{and} \quad \eta_{ij} = \left[1 + \frac{2\gamma_{ij}}{\beta_i + \beta_j + \sum_{k\neq i} \gamma_{ik} \ell\text{n}\, x_k + \sum_{k\neq j} \gamma_{jk} \ell\text{n}\, x_k} \right]^{-1} . \tag{17}$$

The CES utility function is homothetic, the translated CES function and translog function are not. A homothetic function is any function that is a monotonic transformation of a function that is homogeneous of degree one. However, as ordinal utility functions are only defined up to a monotonic transformation, there is really no distinction. A preference based characterization of homothetic preferences is that $\mathbf{x}_1 \succcurlyeq \mathbf{x}_2 \Rightarrow k\mathbf{x}_1 \succcurlyeq k\mathbf{x}_2$ for all positive k. Homotheticity is a useful property that permits aggregation of consumption into a bundle that allows for a simple measurement of inflation or the cost of living. This is examined in Chapter 13.

The Consumer Demand Problem — A Brief Review of Price Theory

The standard problem of consumer choice is to choose the combination of goods that maximizes utility subject to a budget constraint. The existence and uniqueness of these solutions is one topic in introductory microeconomics. Here we will touch on only those features that are important in Finance.

If a consumer has wealth w to allocate between goods with prices $\mathbf{p}$, the problem and first-order conditions for solution are[5]

$$\begin{gathered} \max_{\mathbf{x}} \mathcal{L}(\mathbf{x},\lambda) \equiv u(\mathbf{x}) + \lambda(W - \mathbf{p}'\mathbf{x}) \\ \mathbf{0} = \frac{\partial \mathcal{L}}{\partial \mathbf{x}} = \frac{\partial u(\mathbf{x}^*)}{\partial \mathbf{x}} - \lambda \mathbf{p} \qquad 0 = \frac{\partial \mathcal{L}}{\partial \lambda} = W - \mathbf{p}'\mathbf{x} . \end{gathered} \tag{18}$$

For any two goods, the first-order conditions show that that marginal rate of substitution is equal to the ratio of the prices or that the marginal utility per unit cost is equated across all goods

$$\frac{\partial u/\partial x_i}{\partial u/\partial x_j} = \frac{p_i}{p_j} \quad \text{or} \quad \frac{\partial u/\partial x_i}{p_i} = \frac{\partial u/\partial x_j}{p_j} . \tag{19}$$

The individual demand curve for any good is determined by solving the first order conditions

$$\mathbf{x}^* = \mathbf{d}(\mathbf{p},W) \quad d_i = u_i^{-1}\left(p_i ; \lambda(W,\mathbf{p}) \right) \tag{20}$$

[5] Typically it is income which is allocated among the goods. It is more convenient in finance to think of the allocation of wealth. Of course, in both cases it is usually only a portion of wealth or income that is allocated. The remainder is saved.

here u_i^{-1} is the inverse of the marginal utility, $\partial u(\mathbf{x})/\partial x_i$. The demand function $\mathbf{x}(\mathbf{p}, w)$ is known as Marshallian demand. This is the demand function generally used in supply and demand analysis.

The Hicksian demand for the goods, $\mathbf{h} = \boldsymbol{\eta}(\mathbf{p}, \bar{u})$, is the least costly allocation that achieves a given level of utility. The Hicksian demand is also known as the compensated demand because to keep utility constant when prices change a compensating change in wealth must be made. It is the solution to

$$\min_{\mathbf{h}} \mathcal{L}(\mathbf{h}, \lambda) \equiv \mathbf{p}'\mathbf{h} + \lambda[\bar{u} - u(\mathbf{h})]$$
$$\mathbf{0} = \frac{\partial \mathcal{L}}{\partial \mathbf{h}} = \mathbf{p} - \lambda \frac{\partial u(\mathbf{h}^*)}{\partial \mathbf{h}} \qquad 0 = \frac{\partial \mathcal{L}}{\partial \lambda} = \bar{u} - u(\mathbf{h}). \tag{21}$$

The Hicksian demand can be used to define the expenditure function which defines the minimum level of income or wealth required to achieve a given level of utility. That is,

$$E(\bar{u}, \mathbf{p}) \equiv \min_{\mathbf{x}} \mathbf{p}'\mathbf{x} \quad \text{subject to} \quad u(\mathbf{x}) \geq \bar{u} \quad \Rightarrow \quad E(\bar{u}, \mathbf{p}) = \mathbf{p}'\mathbf{h}^*. \tag{22}$$

Optimal Hicksian and Marshallian demand and the expenditure function are related by

$$\mathbf{h}^* \equiv \boldsymbol{\eta}(\mathbf{p}, \bar{u}) = \mathbf{x}^* = \mathbf{d}\big(\mathbf{p}, E(\bar{u}, \mathbf{p})\big). \tag{23}$$

Indirect utility is a function of wealth and prices and measures the utility that a given level of wealth provides for a fixed set of prices when spent optimally; $I(W; \mathbf{p}) = \max_{\mathbf{x}} u(\mathbf{x})$, subject to $\mathbf{p}'\mathbf{x} = W$. The expenditure and indirect utility functions are all related by

$$I\big(E(\mathbf{p}, u); \mathbf{p}\big) = u \qquad \text{and} \qquad E\big(I(W; \mathbf{p}), \mathbf{p}\big) = W. \tag{24}$$

Assuming the Marshallian demand, $\mathbf{d}(W, \mathbf{p})$, is differentiable at the point (W°, $\mathbf{p}^\circ$), Hicksian demand is also differentiable at the corresponding point ($\mathbf{p}^\circ$, u°) where $u^\circ = I(W^\circ, \mathbf{p}^\circ)$, and

$$\begin{aligned} \frac{\partial \boldsymbol{\eta}(\mathbf{p}^\circ, u^\circ)}{\partial p_i} &= \frac{\partial \mathbf{d}(\mathbf{p}^\circ, W^\circ)}{\partial p_i} + \frac{\partial \mathbf{d}(\mathbf{p}^\circ, W^\circ)}{\partial W} \frac{\partial E(u^\circ, \mathbf{p}^\circ)}{\partial p_i} \\ &= \frac{\partial \mathbf{d}(\mathbf{p}^\circ, W^\circ)}{\partial p_i} + \frac{\partial \mathbf{d}(\mathbf{p}^\circ, W^\circ)}{\partial W} d_i(\mathbf{p}^\circ, W^\circ) \end{aligned} \tag{25}$$

where the last equality follows from (22). Rearranging terms and noting that the optimal Marshallian and Hicksian are equal gives the Slutsky equation that describes the change in the optimal Marshallian demand caused by the change of a price

$$\left.\frac{\partial \mathbf{x}^*}{\partial p_i}\right|_W = \left.\frac{\partial \mathbf{x}^*}{\partial p_i}\right|_u - x_i^* \left.\frac{\partial \mathbf{x}^*}{\partial W}\right|_{\mathbf{p}}. \tag{26}$$

The first term on the right hand side of (26), the change in Hicksian demand, is known as the substitution effect. It is always negative for the good whose price is increasing and positive for the other goods (unless they are strict complements). That is, an increase in price *i* causes demand to shift from good *i* into other goods. The second term (including the negative sign) is the wealth effect, more commonly called the income effect. It is the increase in the consumption when wealth is changed while prices are held fixed. An increase in wealth would typically increase consumption in all goods, but a price increase is an effective decrease in wealth so the wealth effect is subtracted. It reinforces the substitution effect for the good whose price is changing and offsets it for the other goods.[6]

The income and substitution effects are illustrated in the figure. An increase in the price of good 1 from p_1 to p_1' rotates the budget line clockwise around the level of wealth as measured in terms of good 2, the numeraire. The substitution effect is the change from point A to B. It looks at the compensated change in demand by increasing the wealth to keep the consumer on the same indifference curve. The consumption of good 1 falls, and that of good 2 rises. The wealth effect is the movement from point B to C, the actual new optimum at the changed price and unchanged wealth. The wealth effect further reduces consumption of good 1. The wealth effect on good 2 on good 2 is so strong that it more than offsets the substitution effect.

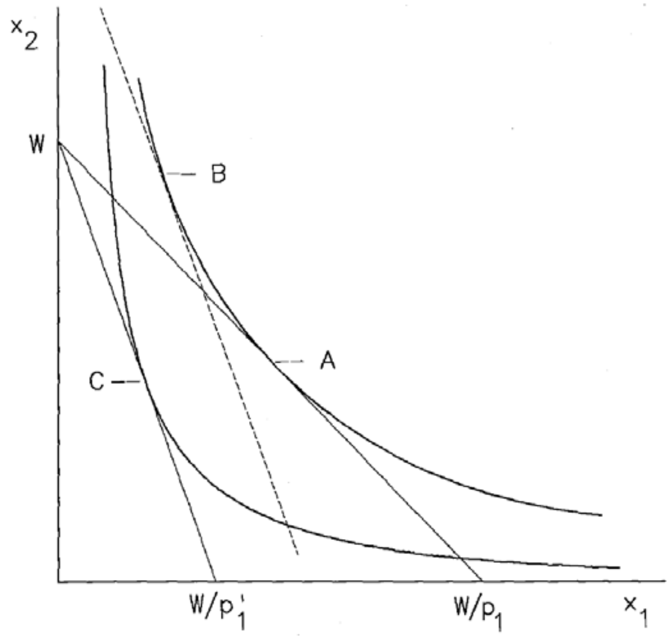


Figure 1.1: Income and Substitution Effects

This figure illustrates the income and substitution effects to so an increase in the price of good one from p_1 to p_1'. The optimal allocation changes from spot A to spot C. The change from A to B is the substitution effect. The price increase results in less consumption of good 1 and more of good 2. The change from B to C is the income (or wealth) effect. The price increase has the same effect as a decrease in income reducing expenditure of both goods.

As an example, consider CES utility defined in (10). The first-order conditions in (18) give demands of $x_i^* = k(p_i/\beta_i)^{-\eta}$ where k is a positive constant. Using the budget constraint, the demand functions are

$$x_i^* = d_i(\mathbf{p}, W) = k[\beta_i^{-1} p_i]^{-\eta} = \frac{\beta^{\eta} p_i^{-\eta}}{\sum_i \beta_i^{\eta} p_i^{1-\eta}} W\,. \tag{27}$$

The expenditure and indirect utility functions are[7]

$$E(u, \mathbf{p}) = \left(\sum p_i^{\rho}\right)^{1/\rho} \cdot u \qquad \text{and} \qquad I(W, \mathbf{p}) = \left(\sum p_i^{\rho}\right)^{-1/\rho} \cdot W\,. \tag{28}$$

For preferences that are homogeneous of degree one, like CES, optimal consumption and indirect utility are always proportional to wealth while the expenditure is proportional to utility. This can be verified directly from the first order conditions, $\partial u/dx_i = \lambda p_i$. Expenditure is

[6] If the wealth effect of price *i* for good *i* is negative, the good is called an *inferior* good. The wealth effect can be so negative that it dominates the substitution effect, so that an increase in the price of a good actually leads to an increase in its consumption. Such a good is called a *Giffen good*. Inferior and Giffen goods virtually never arise in Finance models outside of classroom examples.

[7] For Cobb-Douglas utility, the expenditure on each good is a constant fraction of wealth, $p_i x_i^* = \beta_i W$. The expenditure and indirect utility functions are $I(W, \mathbf{p}) = W \cdot \prod_i (\beta_i / p_i)^{\beta_i}$ and $E(u, \mathbf{p}) = u \cdot \prod_i (p_i / \beta_i)^{\beta_i}$

$$E\big(u(\mathbf{x}^*),\mathbf{p}\big)=\sum p_i x_i^* = \lambda(\mathbf{p})\sum \frac{\partial u}{\partial x_i}x_i^* = \lambda(\mathbf{p})u(\mathbf{x}^*)\,. \tag{29}$$

The second equality follows from Euler's theorem on homogeneous functions. As expenditure must equal wealth at the optimum and utility of the optimal consumption bundle is indirect utility, $W = E(u, \mathbf{p}) = \lambda(\mathbf{p})u = \lambda(\mathbf{p})I(W, \mathbf{p})$ so $I = W/\lambda(\mathbf{p})$. The function λ is homogeneous of degree one in the prices. The same is true for any homothetic function when the ordinally equivalent homogeneous representation is chosen.

If, instead, demand is linear in wealth, $x_i^* = \Gamma_i(\mathbf{p}) + \Lambda_i(\mathbf{p})W$, preferences are said to be quasi-homothetic. Stone-Geary preferences are an example with

$$x_i^* = \underline{x}_i + \frac{\beta_i}{p_i}\big(W - \textstyle\sum_j p_j \underline{x}_j\big)\ . \tag{30}$$

The expenditure and indirect utility functions are

$$E(u,\mathbf{p}) = \textstyle\sum_i p_i \underline{x}_i + u\cdot\prod_i (p_i/\beta_i)^{\beta_i} \qquad \text{and} \qquad I(W,\mathbf{p}) = \big(W - \sum_i p_i \underline{x}_i\big)\prod_i (p_i/\beta_i)^{-\beta_i}. \tag{31}$$

It is obvious that all goods are normal goods with CES utility, but the cross price effects on demand are ambiguous in sign

$$\left.\frac{\partial x_i}{\partial p_j}\right|_W = (\eta-1)\frac{W\beta_i^\eta p_i^{-\eta}}{\big[\sum_i \beta_i^\eta p_i^{\eta+1}\big]^2}\beta_j^\eta p_j^{-\eta} \gtrless 0 \text{ as } \eta \gtrless 1\,. \tag{32}$$

If elasticity is less than one, increasing the price of good j increases the consumption of the other good. When is less than one 0, the opposite is true. The figure illustrates this second case. Cobb-Douglas (or additive-log) utility is the dividing case where the income and cross substitution effects exactly cancel. We will see that it is generally true that log utility is very special under uncertainty as well.

Choice under Uncertainty, Cardinal Utility, and Expected Utility Theory

In Finance we are generally more concerned with choices under uncertainty. Our "goods" are portfolios or lifetime consumption plans. These are "goods" with uncertain attributes, their monetary payoffs. In the simplest case our choices are from a static set of probability distributions with known, objective probabilities. This is the von Neumann-Morgenstern case.

As with ordinal utility, there is a set, $\mathcal{X}$, of goods or attributes. In Finance, the vector of goods, $\mathbf{x}$, is often taken to be a scalar, like wealth or the amount of a single consumption good. Now the preference relation is defined not directly on $\mathcal{X}$ but over probability distribution of these attributes, $P(\mathbf{x}) \in \mathcal{P}$. That is, preferences are expressed as $\tilde{\mathbf{x}} \succcurlyeq \tilde{\mathbf{y}}$. This, of course, includes the previous ordinal preference because the random variable outcomes can be deterministic. If the lottery $\mathbb{1}_\mathbf{x}$ that pays $\mathbf{x}$ for sure is preferred to the lottery $\mathbb{1}_\mathbf{y}$ that pays $\mathbf{y}$ for sure, then we can say that $\mathbf{x}$ is preferred to $\mathbf{y}$. That is

$$\mathbb{1}_\mathbf{x} \succcurlyeq_{\text{cardinal}} \mathbb{1}_\mathbf{y} \quad \Leftrightarrow \quad \mathbf{x} \succcurlyeq_{\text{ordinal}} \mathbf{y} \tag{33}$$

If we wish to be precise when developing cardinal utility, the statement $\mathbf{x}$ is preferred to $\mathbf{y}$ should always be interpreted in this sense using lotteries rather than outcomes because the preference relation is defined on $\mathcal{P}$.

Utility here is cardinal rather than ordinal because preferences will be described with expectations

$$P(\mathbf{x}) \succsim Q(\mathbf{x}) \quad \Leftrightarrow \quad \sum_{\mathbf{x}\in\mathcal{X}} P(\mathbf{x})u(\mathbf{x}) \geq \sum_{\mathbf{x}\in\mathcal{X}} Q(\mathbf{x})u(\mathbf{x}) . \tag{34}$$

Monotonic transformations of u are no longer equivalent because their expectations will typically change differently, and the ordering may switch.

The math is simplest if we assume that the set of goods is discrete. In that case the choices are discrete probability distributions like lotteries, and we can denote them as vectors; $\mathbf{p} \equiv P(\mathbf{x})$. The i^{th} element of $\mathbf{p}$ is the probability of receiving the i^{th} prize. The intuition, however, is the same for continuous probability densities; we just need a few more axioms to eliminate pathologies.

Even if the set of prizes, $\mathcal{X}$, is finite, the set of (necessarily discrete) lotteries on it, $\mathcal{P}$, is uncountable as the probabilities can take on any real values between 0 and 1. So to derive the von Neumann-Morgenstern result we need an additional axiom like the continuity axiom (A3o). In fact, we require one like that plus two others. To describe these axioms, we must develop the properties of the set or space of distributions. For a finite $\mathcal{X}$ of size n, the space of all probability distributions on $\mathcal{X}$ is the subset of the n-dimensional vector space whose elements are nonnegative and sum to one. This is called the unit simplex. This space is closed and convex.[8] Because $\mathcal{P}$ is a convex set, if $P(\mathbf{x})$ and $Q(\mathbf{x})$ are any two probability distributions in $\mathcal{P}$ and $\pi\in[0,1]$. Then

$$R(\mathbf{x}) = \pi P(\mathbf{x}) + (1-\pi)Q(\mathbf{x}) \tag{35}$$

is also in $\mathcal{P}$. In this regard, R can be thought of as a compound lottery in which first a biased coin that lands heads with probability π is flipped. If heads comes up, then the lottery P is conducted. Otherwise Q is conducted.

The additional required axioms are:

A3c) **Archimedean**: $\forall P, Q, R \in \mathcal{P}$ if $P \succ Q \succ R$, then there exists π_1, π_2, $\in$ (0,1) and a unique $\pi^* \in (0, 1)$ such that[9]

$$\begin{gathered} \pi_1 P(\mathbf{x}) + (1-\pi_1)R(\mathbf{x}) \succ Q(\mathbf{x}) \succ \pi_2 P(\mathbf{x}) + (1-\pi_2)R(\mathbf{x}) \\ \text{and} \qquad \pi^* P(\mathbf{x}) + (1-\pi^*)R(\mathbf{x}) \sim Q(\mathbf{x}) . \end{gathered} \tag{36}$$

A4) **Independence:** $\forall P, Q, R \in \mathcal{P}$ and $\forall \pi \in [0,1)$,

$$\begin{aligned} P(\mathbf{x}) \succ Q(\mathbf{x}) \quad &\Leftrightarrow \quad \pi P(\mathbf{x}) + (1-\pi)R(\mathbf{x}) \succ \pi Q(\mathbf{x}) + (1-\pi)R(\mathbf{x}) \\ P(\mathbf{x}) \sim Q(\mathbf{x}) \quad &\Leftrightarrow \quad \pi P(\mathbf{x}) + (1-\pi)R(\mathbf{x}) \sim \pi Q(\mathbf{x}) + (1-\pi)R(\mathbf{x}) . \end{aligned} \tag{37}$$

A5) **State Independenc**e: If $P(\mathbf{x}) = Q(\mathbf{x})$, then $P \sim Q$.

The last axiom would appear to be trivial; of course there should be indifference between P and Q if they are the same gamble. However, that is not exactly what the axiom says. Is says

[8] It is important for the Theorem 1.2 that the choice set is convex. This is usually not an issue as any mixed lottery like $R(\mathbf{x})$ will typically be available. And even if it is not literally available as in some experimental situations, it can still be considered abstractly. However, there may be some situations in which the choice set cannot even be considered to be convex and the theorem is not applicable.

[9] The second relation in (36) can actually be proved from the first and the other axioms, but it is often stated separately or explicitly as part of the Archimedean Axiom as here. Similarly the second relation in (37) can be proved to hold, but it often stated explicitly as part of the Independence Axiom or as a separate axiom called the Calibration Axiom.

that two gambles that there must be indifference between to gambles whenever they have the same probability distribution. That is, the utility derived from a given outcome cannot depend on other circumstances outside of the gamble itself. When evaluating the gamble a $1000 payment has the same utility regardless of how much you might need the money. This axiom might seem overly restrictive; nevertheless, it need not be, because the other circumstances can be described by other elements of **x** when the payoff is a vector. Quite often in Finance, however, utility is assumed to depend only on a single consumption good or wealth. In models like that, the assumption that utility is state-independent and does not depend on other circumstances can limit the interpretation of utility theory.

The Archimedean Axiom is a generalization of the Continuity Axiom (A3o) under ordinal utility. (You may wish to prove that (A3o) is a consequence of (A3c) when applied to lotteries with no uncertainty.) It says there is no lottery P so good that all combinations of it with an inferior lottery R are better than a lottery Q that ranks between the two. Similarly, no lottery R is so bad that some combination with P cannot beat Q. In other words, like the Continuity Axiom for ordinal utility, it says there are always trade-offs possible.

As with ordinal utility, there are some preferences for which this may not be true. One might argue, for example, that any gamble involving death would never be selected in combination with another gamble, if there was a choice which precluded death. However, it should be noted that every day all of us make decisions like driving a car or crossing a street that do include the possibility of dying so we do make such tradeoffs implicitly. In any case, the types of choices we will be examining in Finance will usually be gambles just involving money for which the Archimedean Axiom seems quite reasonable. [10]

The Independence Axiom says that if some gamble is preferred to another, then mixtures of them with a third gamble cannot alter the ordering. This may seem counterintuitive at first. You probably already know that diversification is advocated to improve the performance of a portfolio of stocks. A risk averse individual might prefer a portfolio of corporate bonds to a portfolio of common stock, but they might also prefer a mixed portfolio of stocks and government bonds to the same mixture of corporate and government bonds. This is not a violation of the Independence Axiom, because this is not the type of mixing entailed.

Mixing in the Independence Axiom is usually styled with the compound lottery description used above. If the coin comes up tails, you move to the secondary lottery R so it does not matter which choice you made. If the coin comes up heads, you'll have either P or Q. You prefer P to Q, so you must prefer the mixture with P in it because receiving R has no differential effect on the two choices.

Another, and perhaps better, way to describe the Independence Axiom is by the elimination common consequences. Consider the two gambles $\tilde{y}$ and $\tilde{z}$ shown in the table. The row labeled common holds the smaller of the two probabilities. The rows labeled residual show the $\tilde{y}$ and $\tilde{z}$ probabilities after subtracting out the common probability. If $\tilde{y}$ is preferred to $\tilde{z}$ and expected utility as in (34) describes the choice, then both

	x_1	x_2	x_3	x_4	x_5
$\tilde{y}$	10%	30%	0%	40%	10%
$\tilde{z}$	0%	30%	10%	30%	30%
common	0%	30%	0%	30%	10%
$\tilde{y}$ residual	10%	0%	0%	10%	0%
$\tilde{z}$ residual	0%	0%	10%	0%	20%

[10] For continuous distributions, the Archimedean Axiom needs to be strengthened as follows. For all $p \in \mathcal{P}$, the sets $\{q \in \mathcal{P} \mid p \succcurlyeq q\}$ and $\{q \in \mathcal{P} \mid p \preccurlyeq q\}$ are closed in the topology of weak convergence.

$$
\begin{gathered}
0.1u(x_1)+0.3u(x_2)+0.4u(x_4)+0.1u(x_5) > 0.3u(x_2)+0.1u(x_3)+0.3u(x_4)+0.3u(x_5) \\
0.1u(x_1)+0.1u(x_4) > 0.1u(x_3)+0.2u(x_5)
\end{gathered} \tag{38}
$$

must be true. The utilities of the common consequences can be subtracted from both sides of the inequality. This does not require any notion of compound lotteries; it is a simple algebraic result. The Independence Axiom is obviously necessary if expected utility as in (34) describes choices.

The five axioms allow the derivation of Expected Utility Theory (EUT) as a criterion for choice. EUT is also known as Cardinal Utility Theory and von Neumann Morgenstern Utility.

Theorem 1.2: Expected Utility Theory. Let $\mathcal{X}$ be a finite set and let $\mathcal{P}$ be the set of all probability distributions on $\mathcal{X}$. Then if a preference relation $\succcurlyeq$ on $\mathcal{P}$ satisfies axioms (A1), (A2), (A3c), and (A4), there exists a real-valued utility function u defined on $\mathcal{X}$ such that

$$
P(\mathbf{x}) \succcurlyeq Q(\mathbf{x}) \quad \Leftrightarrow \quad \sum_{\mathbf{x}\in\mathcal{X}} P(\mathbf{x})u(\mathbf{x}) \ge \sum_{\mathbf{x}\in\mathcal{X}} Q(\mathbf{x})u(\mathbf{x}) . \tag{39}
$$

Moreover, u is unique up to a positive affine transformation.

Proof: (Sufficiency only) As $\mathcal{X}$ is a finite set it has both a most-preferred, $\bar{\mathbf{x}}$, and a least preferred, $\underline{\mathbf{x}}$, element. These elements need not be unique; if there are multiple equally best or worst elements, one can be chosen arbitrarily. For every other element, $\mathbf{x}_i$, find the equally preferred lottery between just the best and worst prizes by computing π_i such that $\pi_i \mathbb{1}_{\bar{\mathbf{x}}} + (1-\pi_i)\mathbb{1}_{\underline{\mathbf{x}}} \sim \mathbb{1}_{\mathbf{x}_i}$ This π_i exists and is unique by the Archimedean axiom.[11] By the independence axiom, each prize bundle $\mathbf{x}_i$ can be replaced by the lottery between the best and worst prize bundles, $\bar{\mathbf{x}}$ and $\underline{\mathbf{x}}$. By repeated application of the independence axiom

$$
\mathbf{p} \equiv P(\mathbf{x}) \sim \sum [\pi_i \mathbb{1}_{\bar{\mathbf{x}}} + (1-\pi_i)\mathbb{1}_{\underline{\mathbf{x}}}] P(\mathbf{x}_i) = \mathbb{1}_{\bar{\mathbf{x}}} \sum \pi_i P(\mathbf{x}_i) + \mathbb{1}_{\underline{\mathbf{x}}} \sum (1-\pi_i) P(\mathbf{x}_i) . \tag{40}
$$

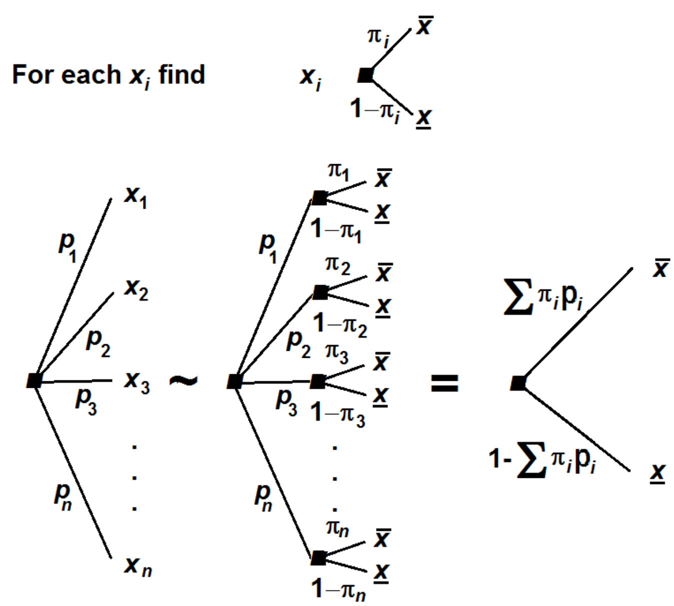


Figure 1.2: Expected Utility Theory

This figure illustrates the construction of the utility function using the individual binary lotteries.

So, for any lottery, we can find the equivalently ranked lottery that has as prizes just the best and

[11] If all elements of $\mathcal{X}$ are equal in preference then $\mathbb{1}_{\bar{\mathbf{x}}} \sim \mathbb{1}_{\underline{\mathbf{x}}}$, and π_i is not unique. But in this case any constant utility function describes choices which are always indifference.

the worst ones. This new lottery has only two outcomes, so it is completely defined by the probability of the best outcome, $\sum \pi_i P(\mathbf{x}_i) \equiv \bar{P}(\mathbf{p})$. By the Archimedean axiom $P \succcurlyeq Q$ if and only if $\bar{P}(\mathbf{p}) \ge \bar{P}(\mathbf{q})$. Now define the utility function as $u(\mathbf{x}_i) \equiv \pi_i$. Note that this makes $u(\bar{\mathbf{x}}) = 1$ and $u(\underline{\mathbf{x}}) = 0$. Then $\bar{P}(\mathbf{p}) = \sum u(\mathbf{x}_i)P(\mathbf{x}_i)$ and the utility function has been constructed.

Clearly $v(\mathbf{x}_i) \equiv a + bu(\mathbf{x}_i)$ with $b > 0$ also ranks lotteries in the same way. ■

This theorem only shows that the four axioms are sufficient conditions to get the cardinal utility representation. Completeness and Transitivity are clearly necessary as well because they are necessary for the ordinal representation which is embedded in EUT. The necessity of the Independence Axiom was discussed above.[12] We have not shown that the Archimedean is necessary, nor that positive affine transformations are the only ones that preserve the ordering, but these are also true.

This theorem remains valid when $\mathcal{X}$ is a bounded, countable or uncountable set provided the set of probability distributions considered, $\mathcal{P}$, is only all discrete distributions with no continuous portions. The theorem can be extended to unbounded continuous distributions with a few more axioms. Primarily the axioms are technical to ensure that the required expectations exist.

The most important of additional axioms is the *sure-thing principle*. This says if a lottery P is concentrated on some continuous set of prizes and every element in the set is as at least as good as some lottery Q, then P is at least as good as Q. The sure thing principle certainly satisfies the requirement that an axiom should seem intuitively obvious. What is perhaps surprising is that it is needed at all.

EUT together with Bayes law for manipulating probabilities form the fundamental underpinnings of essentially all of classical finance theory. Most of what is called *Behavioral Finance* assumes that agents do not choose to maximize expected utility or don't use Bayes law to update probabilities, or both. Some people also include in Behavioral Finance models in which utility is maximized, but the argument of the utility function is unusual. Models of this sort include realization utility, habit formation, and hyperbolic discounting.

Risk Aversion

The aversion to risk is a common assumption and one we will adopt unless otherwise explicitly noted. A von Neumann-Morgenstern utility function is locally *risk averse* at $\mathbf{x}_0$ if it is values every gamble $\tilde{\mathbf{x}}$ at less that its actuarially fair value at the status quo. That is, $\mathbb{E}[u(\mathbf{x}_0 + \tilde{\mathbf{x}})] \le u(\mathbf{x}_0 + \mathbb{E}[\tilde{\mathbf{x}}])$. If this holds at all relevant $\mathbf{x}_0$, the utility function is said to be globally risk averse. Note that local and global refer to the status quo and not the scope of the risky outcomes. In either case the risky outcomes range over the entire outcome space. If the inequality is a strict one (or strict everywhere), the utility function is strictly (globally) risk averse.

Theorem 1.3: Risk Aversion. A von Neumann Morgenstern utility function is (strictly) risk averse at $\mathbf{x}_0$ if and only if the function is (strictly) concave at $\mathbf{x}_0$. The utility function is (strictly) globally risk averse if and only if it is (strictly) concave for all $\mathbf{x} \in \mathcal{X}$.

Proof: Sufficiency follows directly from Jensen's inequality because utility is concave. To prove necessity, consider the set of simple gambles $\boldsymbol{\varepsilon} = a\mathbf{d},\ a > 0$ with probability $b/(a+b)$ and $\boldsymbol{\varepsilon} = -b\mathbf{d},\ b > 0$ with probability $a/(a+b)$ where $\mathbf{d}$ is an arbitrary vector describing the

[12] The Independence Axiom is not required for the preference ordering to be represented by a real-valued function. If Axioms A1, A2, and A3c hold then there is a real valued function V such that $\mathbf{p} \succcurlyeq \mathbf{q} \Leftrightarrow V(\mathbf{p}) \ge V(\mathbf{q})$. Independence is required only to obtain the specific expected utility representation in (39). The Further Notes at the end of this chapter discuss several other representations that do not satisfy Independence but have similar representations.

"direction" of the gamble.[13] This gamble is fair in every component so by assumption it is disliked, and

$$u(\mathbf{x}) \geq \frac{a}{a+b}u(\mathbf{x}-b\mathbf{d}) + \frac{b}{a+b}u(\mathbf{x}+a\mathbf{d}). \tag{41}$$

But this is the definition of a concave function. If the gambles are strictly disliked, then the inequality in (41) is strict and u is strictly concave. ■

If a utility function is twice differentiable, then it is concave and therefore risk averse at a point when the Hessian matrix, $\mathbf{H} \equiv (\partial^2 u/\partial\mathbf{x}\partial\mathbf{x}')$, is negative semi-definite at that point. For strict risk aversion, the Hessian matrix must be negative definite. For univariate utility, the utility function must have a nonpositive second derivative or negative second derivative for strict risk aversion. For a bivariate function, each second derivative must be nonpositive together with $(\partial^2 u/\partial x_1^2)(\partial^2 u/\partial x_2^2) - (\partial^2 u/\partial x_1 \partial x_2) \geq 0.$ For strict concavity, each second derivative must be negative and the expression just given must be positive.

Utility need not be differentiable to be risk averse, but it must be continuous as all increasing concave functions, except for a constant function, are continuous on any open interval. A discontinuity is possible only at the left boundary. Furthermore, utility can be nondifferentiable only at countable many points. These properties should be intuitive, but a formal proof is beyond the scope of this book.

Risk Premiums and Certainty Equivalents

In finance, utility is commonly defined on a single outcome variable, like consumption or wealth. For single-attributed utility like that, the certainty equivalent of any gamble can be defined. The certainty equivalent is the fixed amount that would give the same utility. If the expected utility maximizer currently has a certain x_0 along with the gamble $\tilde{\eta}$, the certainty equivalent, $c(x_0, \tilde{\eta})$, is defined by

$$\mathbb{E}[u(x_0 + \tilde{\eta})] \equiv u\left(x_0 + c(x_0, \tilde{\eta})\right). \tag{42}$$

Note that the certainty equivalent typically depends on the endowment, x_0, as well as the gamble itself. One utility function for which this is not true is the exponential

$$-\mathbb{E}[\exp(-a(x_0 + \tilde{\eta}))] \equiv -\exp(-a(x_0 + c)) \quad \Rightarrow \quad -\mathbb{E}[\exp(-a\tilde{\eta})] = -\exp(-ac). \tag{43}$$

So the certainty equivalent is independent of the endowment.

The certainty equivalent can be thought of as a price in some contexts. For example, if $\tilde{\eta}$ is the amount that is repaid on a loan, c might be the price at which the lender would be willing to sell the loan to a third party. If $\tilde{\eta}$ is the amount of a casualty loss, c would be the largest insurance premium that a buyer would be willing to pay to purchase full reimbursement.

The risk premium is the difference between the certainty equivalent and the expected payoff of the gamble

$$\begin{aligned} &\mathbb{E}[u(x+\tilde{\eta})] \equiv u(x + \mathbb{E}[\tilde{\eta}] - \pi(x, \tilde{\eta})). \\ &\Rightarrow \quad \pi(x, \tilde{\eta}) = \mathbb{E}[\tilde{\eta}] - c(x, \tilde{\eta}). \end{aligned} \tag{44}$$

This, too, depends on the current x in general. Note the different meaning of the word premium. An insurance premium is the certainty equivalent while the risk premium is difference between

[13] The only requirement on $\mathbf{d}$ is that both $\mathbf{x} + a\mathbf{d}$ and $\mathbf{x} - b\mathbf{d}$ be in $\mathcal{X}$.

the actuarially fair outcome and the certainty equivalent.

In Finance, we are more often interested in a slightly different interpretation, namely what extra compensation above the expected payoff to induce someone to take on a risk. An example would be how much extra return is required on stock to induce investors to purchase it rather than safe bonds? This premium, $\hat{\pi}$, is defined by

$$u(x) \equiv \mathbb{E}\left[u\left(x + \tilde{\eta} + \hat{\pi} - \mathbb{E}[\tilde{\eta}]\right)\right]. \tag{45}$$

The difference between (45) and (44) is that the premium is now included with the random outcome rather than the certain outcome. This risk premium is sometimes called a compensatory risk premium to distinguish the two.

The two risk premiums can be approximated using a Taylor expansion. Define $\tilde{\varepsilon} \equiv \tilde{\eta} - \bar{\eta}$, then the risk premium for taking on a new risk is

$$\begin{aligned} u(x) &\equiv \mathbb{E}[u(x + \tilde{\eta} + \hat{\pi} - \bar{\eta})] \approx \mathbb{E}[u(x) + u'(x)(\tilde{\varepsilon} + \hat{\pi}) + \tfrac{1}{2}u''(x)(\tilde{\varepsilon} + \hat{\pi})^2] \\ &= u(x) + u'(x)\hat{\pi} + \tfrac{1}{2}u''(x)\left(\hat{\pi}^2 + \text{var}[\tilde{\varepsilon}]\right) \\ &\Rightarrow \quad \hat{\pi} \approx \frac{1}{2}\frac{-u''(x)}{u'(x)}\text{var}[\tilde{\varepsilon}]. \end{aligned} \tag{46}$$

While the risk premium to remove a risk already present is the solution to

$$\mathbb{E}[u(x + \tilde{\eta})] \equiv u\left(x + \bar{\eta} - \pi(x, \tilde{\eta})\right). \tag{47}$$

Using Taylor expansions on both sides gives

$$\begin{aligned} \mathbb{E}[u(x+\bar{\eta}) + u'(x+\bar{\eta})\tilde{\varepsilon} + \tfrac{1}{2}u''(x+\bar{\eta})\tilde{\varepsilon}^2] &\approx u(x+\bar{\eta}) - \pi u'(x+\bar{\eta}) + \tfrac{1}{2}\pi^2 u''(x+\bar{\eta}) \\ \tfrac{1}{2}u''(x+\bar{\eta})\,\text{var}[\tilde{\varepsilon}] &\approx -\pi u'(x+\bar{\eta}) + \tfrac{1}{2}\pi^2 u''(x+\bar{\eta}) \\ \Rightarrow \quad \pi &\approx \frac{1}{2}\frac{-u''(x+\bar{\eta})}{u'(x+\bar{\eta})}\text{var}[\tilde{\varepsilon}]. \end{aligned} \tag{48}$$

These second-order approximations are valid when the variance and the utility ratio are both small. The utility ratio, $A(\cdot) \equiv -u''(\cdot)/u'(\cdot)$ in (46) and (48) is known as the Arrow-Pratt measure of absolute risk aversion. The difference in the two risk premiums arises primarily because the risk aversion is evaluated at different arguments.[14] For exponential utility, the two risk premiums approximations are equal as A is constant.[15] In general, they are only approximately equal.

The Arrow-Pratt measure is the same for any positive affine transformation of a cardinal utility function and, therefore, must completely describe it. In other words, the utility function can be completely recovered from its Arrow-Pratt measure up to the arbitrary constants of the affine transformation. Noting that $A(x) = -d[\ell\text{n}\,u'(x)]/dx$, [16]

[14] The next two terms in the expansions are $-\frac{1}{6}A^3(\cdot)\,\text{var}^2[\tilde{\varepsilon}] \pm \frac{1}{24}A^5(\cdot)\,\text{var}^3[\tilde{\varepsilon}]$, where the plus sign applies to (48) and the minus sign to (46).

[15] In fact the two risk premiums are exactly the same for exponential utility even if the risks are not small. From (45) $e^{-ax} = \mathbb{E}[e^{-a(x+\tilde{\eta}+\hat{\pi}-\bar{\eta})}]$. From (44), $\mathbb{E}[e^{-a(x+\tilde{\eta})}] = e^{-a(x+\bar{\eta}-\pi)}$. Both reduce to $e^{a\hat{\pi}} = \mathbb{E}[e^{-a(\tilde{\eta}-\bar{\eta})}]$.

[16] The inner integral is $\ell\text{n}\,u'(x) + b$. Exponentiating gives $e^b\ell\text{n}[u'(x)]$ with $B \equiv e^b > 0$.

$$\int^x \exp\left[-\int^z A(s)ds\right]dz = \int^x \exp\left[\int^z \frac{d\ell\text{n}\,u'(s)}{ds}ds\right]dz = \int^x \exp\left[\ell\text{n}\,u'(z)+b\right]dz$$
$$= e^b\int^x u'(z)dz = a + Bu(x) \quad \text{with } B \equiv e^b > 0\,. \tag{49}$$

Two other measures can be derived directly from the Arrow-Pratt measure. They are the absolute risk tolerance measure, $T(x) \equiv 1/A(x)$, and the relative risk aversion measure, $R(x) \equiv xA(x)$ (for $x > 0$). Risk tolerance is useful in describing the behavior of groups. In many circumstances a group of people who can share risks act like an individual with a risk tolerance equal to the average risk tolerance. Relative risk aversion is useful in describing investment allocations where the proportional allocations are important or when premiums are expressed in rate of return form.

Commonly Used Utility Functions

Some commonly used utility functions are given in the table below. The exponential utility function is also called constant absolute risk aversion (CARA) utility. Power and log utility functions are also called constant relative risk aversion (CRRA) utility. Log utility is the special case of power utility with $\gamma = 1$.[17]

Common Single-Attributed Utility Functions

name	form	domain	$A(x)$	$R(x)$
(negative) exponential or CARA	$-e^{-ax}$	$-\infty < x < \infty$	a	ax
power or isoelastic or CRRA	$x^{1-\gamma}/(1-\gamma)$	$x > 0$	α/x	α
generalized power	$(x-\underline{x})^{1-\gamma}/(1-\gamma)$	$x > \underline{x}$	$\alpha/(x-\underline{x})$	$\alpha x/(x-\underline{x})$
logarithmic	$\ell\text{n}\,x$	$x > 0$	$1/x$	1
generalized logarithmic	$\ell\text{n}(x-\underline{x})$	$x > \underline{x}$	$1/(x-\underline{x})$	$x/(x-\underline{x})$
quadratic	$x-\frac{1}{2}bx^2$	$x \le b$	$b/(1-bx)$	$bx/(1-bx)$

All of these utility functions are special cases of the Hyperbolic Absolute Risk Aversion (HARA) or Linear Risk Tolerance (LRT) family which has the form

$$u(x) = a\left(\frac{x}{\alpha}+\eta\right)^{1-\alpha} \text{for} \quad a\frac{1-\alpha}{\alpha} > 0 \text{ and } \frac{x}{\alpha}+\eta > 0\,. \tag{50}$$

The restrictions are required to ensure that $u' > 0$ and $u'' < 0$. So the domain of x has an upper bound if α is negative and a lower bound if α is positive and finite.[18] Risk aversion of the HARA utility is

$$A(x) = \left(\frac{x}{\alpha}+\eta\right)^{-1} \quad \text{or} \quad T(x) = \eta + \frac{x}{\alpha} \tag{51}$$

These hyperbolic and linear forms give rise to the names.

---

[17] This can be verified by using L'hospital's rule to take the limit,

$$\lim_{\gamma\to1}(x^{1-\gamma}-1)/(1-\gamma) = \lim_{\gamma\to1}[\partial(\exp((1-\gamma)\ell\text{n}\,x)-1)/\partial\gamma/(\partial(1-\gamma)/\partial\gamma)] = -\ell\text{n}\,x/(-1) = \ell\text{n}\,x.$$

[18] If α is a negative integer, the utility function is defined beyond the upper bound given, but it is decreasing in that region.

As $\alpha \to \infty$, $A(x) \to 1/\eta$ so this limiting case gives rise to exponential utility, $-e^{-x/\eta}$. For $\eta = 0$ relative risk aversion is constant and utility has the power form. Quadratic utility is $\alpha = -1$. LRT utility is often written as

$$u(x) = \tfrac{1}{\gamma}(x - \underline{x})^{\gamma} \quad \gamma \le 1 \qquad \text{or} \qquad u(x) = \tfrac{1}{1-\alpha}(x - \underline{x})^{1-\alpha}. \tag{52}$$

This form is less cumbersome, but cannot be so easily manipulated into exponential utility. Here $\underline{x}$ is a lower bound on the domain. It can be thought of as a subsistence level, the minimum level required. Conversely if $\underline{x}$ is negative, it may be thought of as a subsidy.

The Boundedness Debate

It is often assumed that utility must be bounded. This assumption is required in certain models mostly to ensure that expected utility is finite and therefore comparable across lotteries. Bounded utility arises naturally when $\mathcal{X}$ is a finite set and utilities of 1 and 0 can be assigned to the best and worst possible outcomes. But it does not necessarily arise when $\mathcal{X}$ is an unbounded set. In fact, when $\mathcal{X}$ is unbounded both above and below, it is impossible to have a utility function that is both bounded, increasing, and concave.[19] If only positive outcomes are valid, or $\mathcal{X}$ is otherwise naturally bounded below, then bounded utility is possible. Nevertheless, most of the commonly used utility functions, like power, log, and exponential are not bounded.

The problem with unbounded utility functions is that there are always some risky prospects that cannot be ranked by an unbounded utility function even though the original preference relation is assumed to be complete. For example consider the gamble that pays $u^{-1}(2^n)$ if heads is first seen on the n^{th} toss of a fair coin. If utility is unbounded above, it is possible to construct such a lottery, and its expected utility is

$$\sum_{n=1}^{\infty} 2^{-n} u\big(u^{-1}(2^n)\big) = \infty\,. \qquad (53)\backslash$$

A lottery that pays $ku^{-1}(2^n)$ is clearly k times as good in some sense but has only the "same" infinite expected utility. A lottery that pays $u^{-1}(2^n)$ if n is odd but zero if n is even is clearly worse, but it again has the "same" expected utility. Therefore, expected utility cannot be used to rank these choices though dominance obviously does so. The same problem arises if utility is unbounded below. A similar lottery that pays $u^{-1}(-2^n)$ has $-\infty$ expected utility. In this case, $\mathcal{X}$ need not even be bounded below. For example, logarithmic utility cannot evaluate such lotteries when assuming $\mathcal{X}$ is the positive reals or a subset $(0, L)$

In either case the proof breaks down because the Archimedean Axiom is violated. Combining such lotteries with others doesn't remove the infinite expected utility. This means that if all possible gambles can be ranked (the preference relation is complete), the utility function must be bounded. Fortunately, this is not a practical concern. The expected utility of a gamble for risk averse utility function can be infinite only if the expected outcome is infinite because $\mathbb{E}[\tilde{x}] \ge \mathbb{E}[u(\tilde{x})]$. As long as the possible outcomes are bounded, this problem will not arise, and in the real world (though not models) outcomes must be bounded by the totality of what is available. There can still be a problem with utility that is unboundedly negative. A log utility investor cannot compare the expected utility of two normally distributed variables; however, this problem can mostly be avoided as the investor would disdain all such choices.

A drawback of bounded utility is that it makes some apparently unreasonable compari-

[19] If u is concave, then $u(x) \le u(x_0) - (x_0 - x)u'(x_0)$, $\forall x < x_0$. But that becomes unboundedly negative as $x \to -\infty$.

sons. Suppose utility is increasing, concave, and bounded above, then there is always a level of initial wealth such that a 50-50 bet between losing a or winning b would not be accepted no matter how large b is. For example, consider $u(w) = -w^{-3}$. When the current wealth is 100, there is no 50-50 bet that loses 25 that will be accepted no matter how large the gain being offered. Rabin (2000) examines this problem in detail.

The Independence Axiom

Of the four axioms underlying EUT, the Independence Axiom (A4) is the one that has sparked the most debate. A violation of this axiom seems to be the usual culprit when utility maximization fails to describe real-world applications.[20] In fact, choices often still violate the axiom even after the discrepancy is pointed out to those making the choices and who profess that the axiom seems reasonable. Therefore, it is important to understand this axiom in detail. Mathematically, what the Independence axiom requires is that expected utility is linear in the probabilities; that is, that choices are based on an expected value of something.

The standard experiment demonstrating violation of the independence axiom is the Allais Paradox. This experiment has been repeated many times. The one reported here was run by Kahnemann and Tversky (1979) in the late 1970s. Seventy-two people were asked to choose between lotteries A and B[21]

$$A:\begin{cases} 33\% \text{ chance of } & 2500 \\ 66\% \text{ chance of } & 2400 \\ 1\% \text{ chance of } & 0 \end{cases} \quad \text{or} \quad B:\ 100\% \text{ chance of } 2400 \tag{54}$$

and also between lotteries C and D.

$$C:\begin{cases} 33\% \text{ chance of } & 2500 \\ 67\% \text{ chance of } & 0 \end{cases} \quad \text{or} \quad D:\begin{cases} 34\% \text{ chance of } & 2400 \\ 66\% \text{ chance of } & 0 \end{cases} \tag{55}$$

The majority of them (61%) choose the sure thing B over the risky A and choose C over D. Either of those choices is fine on its own; however, both choices together violate the independence axiom.

The violation is usually demonstrated as follows. Mentally think of lotteries A and B as

$$A:\begin{cases} 33\% \text{ chance of } & 2500 \\ 1\% \text{ chance of } & 0 \\ 66\% \text{ chance of } & 2400 \end{cases} \quad \text{or} \quad B:\begin{cases} 34\% \text{ chance of } & 2400 \\ \\ 66\% \text{ chance of } & 2400 \end{cases} \tag{56}$$

In each lottery, there is a 66% chance of winning 2400. So the preference for B over A must be due to a preference for the 34% chance at 2400 over the 33% chance at 2500 plus a 1% chance at nothing.

Now mentally think of lotteries C and D as

[20] Another culprit is a failure to deal with probabilities correctly. For example, there is evidence that people put undue emphasis on rare or salient outcomes or that they incorrectly update probabilities by not properly applying Bayes' Law.

[21] In his original paper, Allias (1953) described the choices in millions of francs as A: 500(10%), 100(89%), 0(1%); B: 100(100%), C: 100(11%), 0(89%); D: 500(10%) 0(90%).

$$C:\begin{cases}33\% \text{ chance of } 2500\\ 1\% \text{ chance of } 0\\ 66\% \text{ chance of } 0\end{cases} \quad \text{or} \quad D:\begin{cases}34\% \text{ chance of } 2400\\ \\ 66\% \text{ chance of } 0\end{cases} \tag{57}$$

This time, each lottery has a 66% chance of 0, but otherwise *C* matches *A* and *D* matches *B*. So a preference for *C* over *D* directly contradicts the previous choice if the Independence Axiom holds.[22]

The analysis above uses the splitting to show in a simple fashion that the violation is indeed of the Independence Axiom rather than one of the other axioms. However, it is not required to demonstrate that the choices are inconsistent with EUT. If lotteries *B* and *C* are preferred to *A* and *D*, then EUT says that the utility function must satisfy

$$\begin{aligned} B \succ A \quad &\Leftrightarrow \quad u(2400) > 0.33u(2500) + 0.66u(2400) + 0.01u(0) \\ &\Leftrightarrow \quad 0 > 0.33u(2500) - 0.34u(2400) + 0.01u(0) \\ C \succ D \quad &\Leftrightarrow \quad 0.33u(2500) + 0.67u(0) > 0.34u(2400) + 0.66u(0) \\ &\Leftrightarrow \quad 0 < 0.33u(2500) - 0.34u(2400) + 0.01u(0)\,. \end{aligned} \tag{58}$$

This is an obvious contradiction. The two relations in (58) cannot be true for any utility function whether or not it is risk averse. In fact they cannot be true even for utility functions that are not increasing.

Similar experiments have confirmed violations of the Independence Axiom in many contexts. Furthermore, even when participants are presented with the four basic axioms and an analysis of the problem, most of them do not alter their choices to conform to the Independence Axiom. Leonard Savage, one of the early developers of the theory of choice under uncertainty, made choices equivalent to *B* and *C* when first presented with the Allais Paradox, but concluded later that he was in error (Savage, 1954, pp 101-3). Other subjects did not so readily change their minds. For example, Moskowitz (1974, pp 232-7) reports 93% of those making Independence-violating choices persisted with their choices after being presented written arguments about the error. If discussion was also allowed, 73% still remained unconvinced their original choices were inconsistent.

The Allais Paradox can be illustrated with a Marschak-Machina triangle (more commonly referred to as a Machina triangle). The triangle is based on a set of lotteries with three possible outcomes, *L*, *M*, and *H* with $L < M < H$. Any lottery can be described by a point within the triangle bounded by the axes and the line segment connecting points (0,1) and (1,0). The horizontal and vertical values of a point give the probabilities for the *L* and *H* outcome, respectively; the probability of *M* is the residual, $1 - p_L - p_H$. The most preferred spot is (0, 1) at the upper left in which the best prize is assured. The least preferred spot is (1, 0) at the lower right. Preference among lotteries increase as the point moves up or left. Moving up increases the probability of the best prize and reduces the probability of the middle prize. Moving left reduces the probability of the worst prize and increases the probability of the middle prize. Moving up and left parallel to the diagonal bounding line increases the probability of the best prize and reduces the probability of the worst prize. All of these movements must increase expected utility regardless of risk

---

[22] It might be argued that people are indifferent between *A* and *B* and between *C* and *D*. In this case choosing *B* and *C* would not be a paradox. However, a slight change in the experiment can confirm this is not the case. Create a new lottery *D′* with a probability of receiving 2400 at 35%. Any expected utility maximizer indifferent between *A* and *B* must be indifferent between *C* and *D*. But *D′* must be preferred to *D* and therefore to *C*; however this choice is still not the one usually expressed.

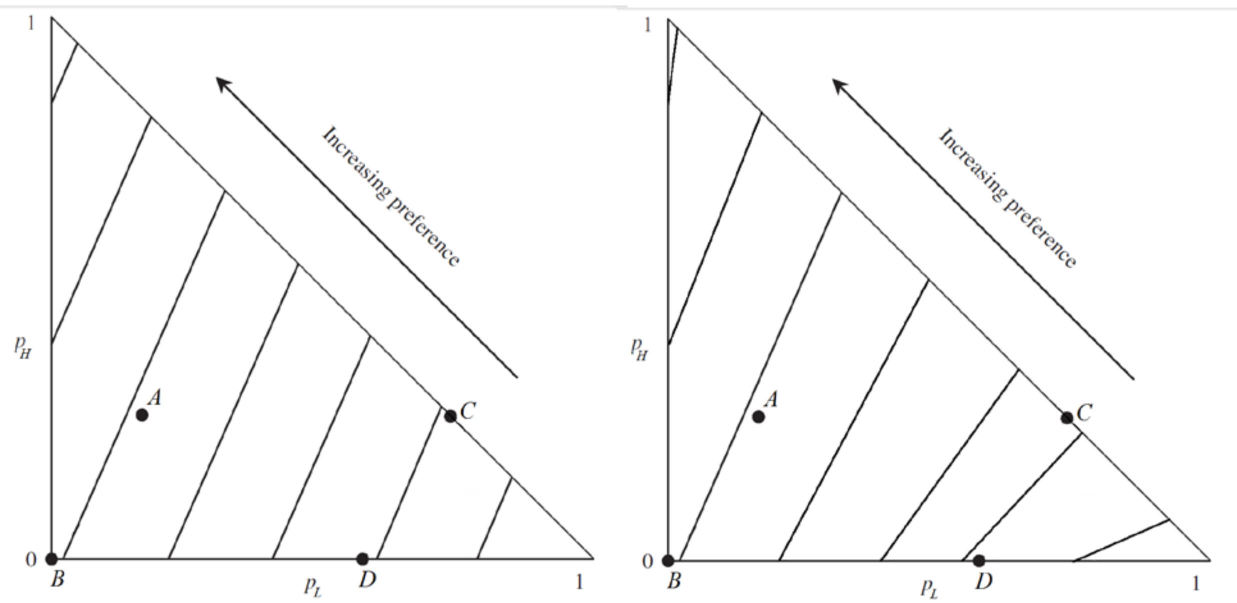


Figure 1.3 Machina Triangle Illustrating the Allais Paradox

This figure illustrates the Allais Paradox with Machina's Triangle. Under Expected Utility Theory, the indifference curves must be parallel straight lines so if *B* is preferred to *A*, then *D* must be preferred to *C*. If the indifference curves are not straight lines, then the common preferences are possible.

aversion (or risk preference).

Indifference curves are those points with the same expected utility, $u°$

$$\mathbb{E}[\tilde{u}] = u° = p_L u_L + (1 - p_L - p_H)u_M + p_H u_H = u_M + (u_L - u_M)p_L + (u_H - u_M)p_H$$
$$\Rightarrow p_H = \frac{u° - u_M}{u_H - u_M} + \frac{u_M - u_L}{u_H - u_M} p_L \, . \tag{59}$$

Because expected utility is linear in the probabilities, all indifference curves have the same positive slope $(u_M - u_L)/(u_H - u_M)$, and are, therefore, parallel straight lines sloping up to the right as illustrated in the left panel. A more risk averse individual has steeper sloped indifference curves,[23] but for any expected utility maximizer, the indifference curves must be straight and parallel.

The lotteries in the Allais paradox example above are at the points indicated. The panel on the left with parallel indifference curves describes the Independence Axiom. The panel on the right shows indifference curves for a decision maker for whom the Independence Axiom does not hold. For both, $B \succ A$. However, the only choice consistent with the Independence Axiom is then $D \succ C$. A less risk averse person might have more shallowly sloped indifference curves and prefer *C* to *D*, but then he would also prefer *A* to *B* because the indifference curves must be parallel. There is no other possibility for an expected utility maximizer (except indifference *A* ~ *B* and *C* ~ *D*). To explain the choices actually seen, indifference curves need to fan out, being steeper to the left and shallower to the right as shown in the right hand panel. Of course, without the Independence Axiom, they need not be straight lines either.

With more than three prizes, the same is true, though it cannot be so easily illustrated.

[23] For any utility function, the normalization $u_L = 0$ and $u_H = 1$ can be used. A more risk averse the agent has a more concave utility function so u_M will be larger increasing the slope in (59).

With k outcomes, all lotteries can be represented by the k dimensional unit simplex. Reasoning just like that in (59) shows that indifference surfaces are parallel hyperplanes in the simplex.

In his original paper, Allais (1953) dismissed the paradox describing the "inconsistent" choices as a flaw in expected utility theory description of choice. But it is more than that; it is a failure of rationality at least in some contexts. This is easiest to demonstrate with a small change to gamble A. The four gambles are now

$$\begin{gathered} A':\begin{cases}\Pr\{2500\}=\frac{33}{34}\\ \Pr\{0\}=\frac{1}{34}\end{cases} \quad \text{or} \quad B:\Pr\{2400\}=100\% \\ C:\begin{cases}\Pr\{2500\}=33\%\\ \Pr\{0\}=67\%\end{cases} \quad \text{or} \quad D:\begin{cases}\Pr\{2400\}=34\%\\ \Pr\{0\}=66\%\end{cases} \end{gathered} \tag{60}$$

Again most people choose C over D and B over A. However, C and D can also be expressed as

$$C:\begin{cases}\Pr\{A'\}=34\%\\ \Pr\{0\}=66\%\end{cases} \quad \text{or} \quad D:\begin{cases}\Pr\{B\}=34\%\\ \Pr\{0\}=66\%\end{cases}. \tag{61}$$

The 66% chances of getting 0 are common to both C and D so by the Independence Axiom they can be ignored in the choice. So obviously the only consistent choices are A and C or B and D.

One difference this time is that the choices go beyond just violating the Independence Axiom. They are irrational because they are not dynamically consistent. Consider someone who, like most people, chooses C and B. If offered a two-stage lottery which first determines if he gets nothing or a second lottery of A' or B, he will choose C, the lottery with A' as the second stage. However, if he is allowed to change his mind after the 34% chance has resolved that the second stage will occur, he will choose B over A'. Furthermore, he is well aware that he would do so even before the first stage is resolved. This is the dynamic inconsistency.

The Allais paradox is a particular example of the *common consequences effect*. The common consequence of a 66% chance of 2400 in A and B should be irrelevant in the choice between those two as should the 66% chance of nothing in both C and D. But those who choose B and D (or A and C) must be ignoring the common consequences as that is all that differs in the choices.

The common consequences effect is more generally described as

$$\begin{gathered} A:\begin{cases}\pi \text{ chance of} & x\\ 1-\pi \text{ chance of} & \mathbf{p}_1\end{cases} \quad \text{or} \quad B:\begin{cases}\pi \text{ chance of} & \mathbf{p}\\ 1-\pi \text{ chance of} & \mathbf{p}_1\end{cases} \\ C:\begin{cases}\pi \text{ chance of} & x\\ 1-\pi \text{ chance of} & \mathbf{p}_2\end{cases} \quad \text{or} \quad D:\begin{cases}\pi \text{ chance of} & \mathbf{p}\\ 1-\pi \text{ chance of} & \mathbf{p}_2\end{cases} \end{gathered} \tag{62}$$

where lottery $\mathbf{p}$ has outcomes both better and worse than x. Under the Independence Axiom, the choices must be either A and C (if $\mathbb{1}_x \succ \mathbf{p}$) or B and D (if $\mathbb{1}_x \prec \mathbf{p}$). However, researchers have found that when $\mathbf{p}_1$ is better than $\mathbf{p}_2$,[24] there is a tendency to choose A and D.

Another failure of the Independence Axiom is *Bergen's Paradox* also known as the *Common Ratio Effect*. When offered the choices

[24] Here "better than" means that $\mathbf{p}_1$ is preferred to $\mathbf{p}_2$ by every strictly increasing utility function. This is known as first-order stochastic dominance which is described in detail in the next chapter.

$$A:\begin{cases} 80\% \text{ chance of } & 4000 \\ 20\% \text{ chance of } & 0 \end{cases} \quad \text{or} \quad B:\begin{cases} 100\% \text{ chance of } & 3000 \\ 0\% \text{ chance of } & 0 \end{cases}$$
$$C:\begin{cases} 40\% \text{ chance of } & 4000 \\ 60\% \text{ chance of } & 0 \end{cases} \quad \text{or} \quad D:\begin{cases} 50\% \text{ chance of } & 3000 \\ 50\% \text{ chance of } & 0 \end{cases} \tag{63}$$

most study participants chose *B* and *C* with a few choosing *A* and *D*. Again both pairs of those choices violate the Independence Axiom. The generic description of the common ratio effect is

$$A:\begin{cases} \text{prob. } p \text{ of } & H \\ \text{prob. } 1-p \text{ of } & L \end{cases} \quad \text{or} \quad B:\begin{cases} \text{prob. } q \text{ of } & M \\ \text{prob. } 1-q \text{ of } & L \end{cases}$$
$$C:\begin{cases} \text{prob. } \alpha p \text{ of } & H \\ \text{prob. } 1-\alpha p \text{ of } & L \end{cases} \quad \text{or} \quad D:\begin{cases} \text{prob. } \alpha q \text{ of } & M \\ \text{prob. } 1-\alpha q \text{ of } & L \end{cases} \tag{64}$$

for $L < M < H$, $0 < p < q < 1$, and $0 < \alpha < 1$. Suppose *A* and *D* are the preferred choices. Then for any expected utility maximizer

$$\begin{aligned} pu_H + (1-p)u_L > qu_M + (1-q)u_L \quad &\Rightarrow \quad p(u_H - u_L) > q(u_M - u_L) \\ \alpha p u_H + (1-\alpha p)u_L < \alpha q u_M + (1-\alpha q)u_L \quad &\Rightarrow \quad \alpha p(u_H - u_L) < \alpha q(u_M - u_L)\,, \end{aligned} \tag{65}$$

but this is clearly impossible. If both *B* and *C* are chosen,[25] then all the inequalities in (65) have the opposite sign, which is also impossible. Here the violation of independence seems to arise from something about the probabilities rather than the utility function.

The common ratio effect can also be illustrated in a Machina triangle. Both gamble pairs (*A*, *B*) and (*C*, *D*) are connected by lines with slopes of 1

$$\frac{\Delta p_H^{AB}}{\Delta p_L^{AB}} = \frac{p-q}{1-p-(1-q)} = 1 \quad \frac{\Delta p_H^{CD}}{\Delta p_L^{CD}} = \frac{\alpha p - \alpha q}{1-\alpha p-(1-\alpha q)} = 1 \tag{66}$$

Because indifference curves are also straight lines, both *A* and *C* must be preferred or both *B* and *D* must be preferred.

Slovic (1969) reported a similar problem with probabilities. When offered payoffs based on four independent events each with probability *p* or based on a single event with probability p^4, participants found the four-way joint occurrence gambles more attractive when gains were involved, but the single occurrence gambles were more attractive when losses would result. This study indicates that not only are probabilities treated in the objective linear fashion that the Independence Axiom requires, but their treatment also depends whether the outcomes are good or bad. This is incorporated into Prospect Theory which is the topic of Chapter 12.

It has been suggested that people may not be use the objective probabilities directly. Instead they tend to put too much subjective emphasis on unlikely events with extreme outcomes. For example, they focus on the grand prize in a lottery and downplay the slim chance they have of winning it. This is known as probability weighting. It is basis of Rank Dependent Utility which is discussed in more detail in Chapter 11. Under Rank Dependent Utility, indifference surfaces need not be flat. Many other alternatives to EUT theory have been proposed to accommodate the Allais and Common Ratio Paradoxes. Some of these are discussed in the Further Notes section at the end of this chapter.

---

[25] There is evidence that the modal choice switches from preferring safety to gambling when the original high probability drops below some threshold around 50%.

Recursive Utility

Perhaps the most commonly used utility structure that does not satisfy the Independence Axiom is recursive utility. Recursive utility uses two functions to separate the evaluation of risk from aggregation over time. For two periods, recursive utility is represented as

$$\psi_0 = \Gamma\left[c_0, \Psi(\tilde{c}_1)\right] \qquad \text{with} \quad \psi_1 = \Psi(\tilde{c}_1) \equiv u^{-1}\left(\mathbb{E}[u(\tilde{c}_1)]\right). \tag{67}$$

The function Ψ gives the certainty equivalent of time-1 consumption, ψ_1. This is the deterministic amount of consumption that has the same utility as the expected utility of the random $\tilde{c}_1$. It is determined with the cardinal utility function *u*.

The function Γ is an aggregator that describes substitution across time. This is expressed as a time-0 certainty equivalent, ψ_0, of the complete two-period plan. The aggregation does not involve taking any expectations so it is an ordinal function, and any monotonic transform serves equally well.

For just two periods, there is nothing really recursive about this utility definition; only in a multi-period setting does the recursion become evident. If *T* is the final period, then

$$\begin{aligned} \psi_{T-1} &= \Gamma\left[c_{T-1}, \Psi(\tilde{c}_T)\right] \\ \psi_{T-2} &= \Gamma\left[c_{T-2}, \Psi(\tilde{\psi}_{T-1})\right] = \Gamma\left[c_{T-2}, \Psi\left(\Gamma[\tilde{c}_{T-1}, \Psi(\tilde{c}_T)]\right)\right] \\ &\vdots \\ \psi_0 &= \Gamma\left[c_0, \Psi(\tilde{\psi}_1)\right] = \Gamma\left[c_0, \Psi\left(\Gamma[\tilde{c}_1, \Psi(\tilde{\psi}_2)]\right)\right] = \cdots \end{aligned} \tag{68}$$

The most commonly used form of recursive utility is Epstein-Zin (also known as Epstein-Zin-Weil) with

$$\begin{aligned} &\Psi(\tilde{z}) \equiv \left(\mathbb{E}[\tilde{z}^{1-\gamma}]\right)^{1/(1-\gamma)} \qquad \Gamma(x,y) \equiv (x^{(\eta-1)/\eta} + \delta y^{(\eta-1)/\eta})^{\eta/(\eta-1)} \qquad \gamma, \eta > 0,\ \ 0 < \delta < 1 \\ &\Rightarrow \quad \psi_t = \left[c_t^{(\eta-1)/\eta} + \delta\left(\mathbb{E}[\tilde{\psi}_{t+1}^{1-\gamma}]\right)^{(\eta-1)/(\eta\alpha)}\right]^{\eta/(\eta-1)} \end{aligned} \tag{69}$$

Risks are assessed with a constant relative risk aversion of γ, intertemporal trade-offs are governed by a constant elasticity of substitution of η), and δ serves the role of subjective discounting.[26] For two-period models, the aggregator can more conveniently be written as

$$\bar{\psi}_0 = \frac{\eta}{\eta - 1}\left[c_0^{(\eta-1)/\eta} + \delta\left(\mathbb{E}[\tilde{c}_1^{1-\gamma}]\right)^{(\eta-1)/[\eta(1-\gamma)]}\right]. \tag{70}$$

This is permissible because the aggregator does not involve expectations so it is an ordinal function and any monotonic transformation can be applied. The disadvantage of these representations is that $\bar{\psi}_0$ is no longer a certainty equivalent in units of consumption. This is not crucial in a one-period model, but with more than a single period. it is important to keep ψ_t as a certainty equivalent so it can be compared to consumption in a consistent fashion.

---

[26] For $\gamma > 1$, CRRA utility is generally written as $c^{1-\gamma}/(1-\gamma)$ to make marginal utility positive. However, the denominator is just an arbitrary scaling factor; it is only important that it be negative to keep utility increasing. Here that is not needed as taking the $1-\gamma$ root serves the same purpose. The special case of $\gamma = 1$ corresponds to log utility as usual. When $\eta = 1$ the aggregator is the Cobb-Douglas function

$$\lim_{\eta\to 1} \ell\text{n}\,\psi_{t-1} = \lim_{\eta\to 1}\left(\tfrac{\eta}{\eta-1}\ell\text{n}[c_{t-1}^{(\eta-1)/\eta} + \delta\psi_t^{(\eta-1)/\eta}]\right) = \ell\text{n}\,c_{t-1} + \delta\ell\text{n}\,\psi_t \quad \Rightarrow \quad \lim_{\eta\to 1}\psi_{t-1} = c_{t-1}\psi_t^{\delta}\,.$$

As this is an ordinal function, we can apply the monotonic transformation raising it to the $\beta = 1/(1+\delta)$. This gives the usual Cobb-Douglas form of $\psi_{t-1} = c_{t-1}^{\beta}\psi_t^{1-\beta}$.

Recursive utility is not, in general, consistent with expected utility as the independence axiom need not hold; however, each separate use of the certainty equivalent function, ψ, applied over a single period does satisfy the axiom. It is only multi-period gambles for which independence is violated.

The following example illustrates the failure of the independence axiom for recursive utility. The consumer has Epstein-Zin preferences with $\alpha = ½$ and $\delta = 1$. Time-0 and time-2 consumption are both 4. For lottery *A*, time-1 consumption is also 4. For lottery *B*, time-1 consumption is 1 or 9 with equal probability. In isolation this time-1 gamble has an expected utility of $0.5(1^{1/2} + 9^{1/2}) = 2$ and a certainty equivalent of $\Psi(\tilde{c}) = 4$. So under the independence axiom, the lotteries should have the same valuation. But they do not for most Epstein-Zin preferences.

Time-2 consumption is certain so its certainty equivalent is just 4. The time-1 and time-0 certainty equivalents are

$$\tilde{\psi}_1 = (\tilde{c}_1^\rho + 4^\rho)^{1/\rho} \quad \psi_0 = [4^\rho + \left(\mathbb{E}[\tilde{\psi}_1^{1/2}]\right)^{2\rho}]^{1/\rho}\,. \tag{71}$$

For $\eta = \infty$, the aggregator is linear, and the evaluations under the Epstein-Zin preferences are

$$\begin{aligned}&A\text{:}\quad \psi_1 = 4+4 = 8 \quad\text{and}\quad \psi_0 = 4+8^1 = 12\\ &B\text{:}\quad \tilde{\psi}_1 = \tilde{c}_1 + 4 \quad \mathbb{E}[\tilde{\psi}_1^{1/2}] = \tfrac{1}{2}\sqrt{1+4}+\tfrac{1}{2}\sqrt{9+4} = 2.9208 \quad \psi_0 = 4 + 2.9208^2 = 12.5311\end{aligned} \tag{72}$$

For $\eta = 2$, the evaluations are

$$\begin{aligned}&A\text{:}\quad \psi_1 = (\sqrt{4}+\sqrt{4})^2 = 4 \quad\text{and}\quad \psi_0 = (\sqrt{4}+\sqrt{4})^2 = 4\\ &B\text{:}\quad \psi_1 = (\sqrt{\tilde{c}_1}+\sqrt{4})^2 \quad \mathbb{E}[\psi_1^{1/2}] = \tfrac{1}{2}(1+2)+\tfrac{1}{2}(3+2) = 4 \quad \psi_0 = [\sqrt{4}^2+\sqrt{4}^2]^{1/2} = 4\end{aligned} \tag{73}$$

For $\eta = 0.5$, the evaluations are

$$\begin{aligned}&A\text{:}\quad \psi_1 = (4^{-1}+4^{-1})^{1/(-1)} = 2 \quad\text{and}\quad \psi_0 = (4^{-1}+2^{-1})^{1/(-1)} = 1.3333\\ &B\text{:}\quad \psi_1 = (\tilde{c}_1^{-1}+4^{-1})^{1/(-1)} \quad \mathbb{E}[\psi_1^{1/2}] = \tfrac{1}{2}\sqrt{(1+\tfrac{1}{4})^{-1}}+\tfrac{1}{2}\sqrt{(\tfrac{1}{9}+\tfrac{1}{4})^{-1}} = 1.2793 \quad \psi_0 = [4^{-1}+(1.2793^2)^{-1}]^{1/(-1)} = 1.1614\end{aligned} \tag{74}$$

When the elasticity of substitution is low, the variation in consumption between the periods is disliked so the smoother lottery, *A*, is preferred. At a lower elasticity, the consumer is more than willing to accept variation. It is no coincidence that he is indifferent between the lotteries when $\eta = 2$. This is always true for Epstein-Zin utility when $\eta = 1/\gamma$ because it is then equivalent to time-additive utility. This is easily verified. If $\eta = 1/\gamma$, then $\eta/(\eta - 1) = 1 - \gamma$, and (70) becomes

$$\psi_0' = (c_0^{1-\gamma} + \delta\mathbb{E}[\tilde{c}_1^{1-\gamma}])\big/(1-\gamma)\,. \tag{75}$$

This is an expected utility evaluation so the Independence Axiom must hold. The same is true with more than two periods.

Epstein-Zin preferences include time-additive CRRA utility extending the class by decoupling risk aversion from the intertemporal elasticity of substitution. However, it does so at the expense of violating the Independence Axiom.

Subjective Probabilities

Throughout this development, probabilities have been taken as given; that is, objective. While this may be true for most casino games, it is certainly not true in other risky situations. For example, in betting on sports or investing in the stock market, the probabilities of the various outcomes are clearly not objective; they are determined subjectively by participants and typically differ. As Mark Twain wrote in *Pudd'nhead Wilson's Calendar* (1894) "It is difference of opinion that makes horse-races." This is also true in Finance.

In many, if not most, models in Finance probabilities or distributions are simply assumed and treated as objective. The exception is models dealing with information where we are concerned with the market reveals information and how beliefs are update. But as a basis for that we need to know that investors do have subjective probabilities. A more general approach than von Neumann Morgenstern shows that prospects are evaluated by computing expected utilities using subjective probabilities and a utility function. This approach is usually called Savage's Theory of Choice after its original developer Leonard Savage; it has been amplified and extended by many others.

The basic idea is that there is a set of all possible relevant events, $\mathcal{S}$, and the decision maker has a binary relation of relative likeliness defined over an algebra, $\mathcal{A}$, on $\mathcal{S}$. The algebra typically used is the set of all subsets of $\mathcal{S}$; that is, any possible combinations of events.[27] The binary likelihood relation $A \dot{\succsim} B$ means that event A is at least as likely as event B. Equally likely is defined as $A \dot{\sim} B$ if $A \dot{\succsim} B$ and $B \dot{\succsim} A$. Strictly more likely is defined as $A \dot{\succ} B$ if $A \dot{\succsim} B$ and not $A \dot{\sim} B$.

Intuitively relative likeliness should be a complete and transitive ordering. Mathematically, this makes $\dot{\succsim}$ a preference relation though it is awkward to call it that. There is one obvious additional requirement, $A \subseteq B \Rightarrow B \dot{\succsim} A$; that is, if B includes all the events that A does, then it must be at least as likely. In fact, if A is a strict subset of B, it should be strictly more likely unless if differs only by events deemed impossible.

Because the subjective probability measure is cardinal rather than ordinal, we also require something like the Independence Axiom. In this case, if $A \dot{\succsim} B$ and $C \cap A = C \cap B = \varnothing$, then $\{A \cup C\} \dot{\succsim} \{B \cup C\}$. This says, if C is mutually exclusive with both A and B, and if A is more likely than B, then $\{A$ or $C\}$ must be more likely than $\{B$ or $C\}$.

These conditions are necessary for establishing the existence of a qualitative probability measure, π, satisfying

$$\forall A, B \in \mathcal{A}, \quad \pi(A) \geq \pi(B) \Rightarrow A \dot{\succsim} B \tag{76}$$

There are other necessary conditions as well. For example, letting A^c denotes the complement of subset A, another necessary condition is if $A \dot{\succsim} B$ then $B^c \dot{\succsim} A^c$. This must be true because the statements $\pi(A) \geq \pi(B)$ is the same as $1 - \pi(B) \geq 1 - \pi(A)$. However, these other conditions are consequences of the assumptions already made about the relative likelihood relation.

Unfortunately, these necessary assumptions are not enough for sufficiency; the assumed likelihood ordering does not guarantee the existence of a probability measure when there are five or more states. Counterexamples like the following can arise. Denote the five outcomes in $\mathcal{S}$ by a, b, c, d, e, and subsets of $\mathcal{S}$ as ab, etc. Suppose relative likelihoods are characterized by

$$\begin{aligned} &abcde \dot{\succ} acd \dot{\succ} be \dot{\succ} abd \dot{\succ} ae \dot{\succ} cd \dot{\succ} bd \dot{\succ} abc \dot{\succ} \ldots \\ &\quad \ldots abc \dot{\succ} e \dot{\succ} bc \dot{\succ} ad \dot{\succ} d \dot{\succ} ac \dot{\succ} ab \dot{\succ} c \dot{\succ} b \dot{\succ} a \end{aligned} \tag{77}$$

with the remaining relative likelihoods defined by the complements; e.g., $A \dot{\succ} B$ if $B^c \dot{\succ} A^c$. One complement pair, acd and be, is included in (77). This is needed to establish the likelihood ordering of the complements relative to the explicit statements. The ordering is complete and

[27] The set of all subsets is called the power set; but any sigma algebra could be used. If some subsets are excluded from $\mathcal{A}$, it simply means that the decision maker is unwilling to make likelihood assessments about those subsets. Of course, expected utility choices concerning payoffs defined by those subsets will also be impossible.

transitive by construction as it describes all comparisons explicitly or implicitly as complements. Only "independence" need be verified.

It can be verified that these likelihood assessments are complete and transitive and satisfy the independence-like axiom. However, using just these orderings, we can conclude that

$$\begin{array}{lcccc} bc \stackrel{\bullet}{\succ} ad & \Rightarrow & \pi_b + \pi_c & > & \pi_a + \pi_d \\ d \stackrel{\bullet}{\succ} ac & \Rightarrow & \pi_d & > & \pi_a + \pi_c \\ ae \stackrel{\bullet}{\succ} cd & \Rightarrow & \pi_a + \pi_e & > & \pi_c + \pi_d \\ acd \stackrel{\bullet}{\succ} be & \Rightarrow & \pi_a + \pi_c + \pi_d & > & \pi_b + \pi_e \\ \hline & & 2\pi_a + \pi_b + 2\pi_c + 2\pi_d + \pi_e & > & 2\pi_a + \pi_b + 2\pi_c + 2\pi_d + \pi_e \,. \end{array} \tag{78}$$

Summing shows that this is impossible for probabilities.

Even when a probability measure is possible, just looking at likelihoods will not uniquely determine it. If there are only two states and $a \stackrel{\bullet}{\succ} b,$ then any value for π_a greater than one-half describes the beliefs. As more states are added, the subjective probabilities become further restricted, but they won't be uniquely set.

What is missing to guarantee the existence of a probability measure is an axiom like the Archimedean Axiom. No axiom like this is possible because the gambles are set by nature. This means we cannot consider probability mixtures of various gambles without going outside the set $\mathcal{S}$. The solution to this problem is obvious; the set considered must be expanded to allow such mixtures. Savage's Theory of Choice was the first to do this. A somewhat different but more intuitive method was developed by Anscombe and Aumann (1963). They added lotteries with objective probabilities to the mix. A simplified version of that approach is used below. The development will show both that the decision maker has subjective probabilities and uses them just like objective probabilities in computing expected utility.

There is a set $\mathcal{S}$ of S disjoint and exhaustive outcome events. The decision maker has the notion of the relative likelihood of the states described by $\stackrel{\bullet}{\succcurlyeq}$ but does not have objective probabilities for them. There is a set $\mathcal{G}$ of gambles that award a prize depending solely on the state that is realized. The prize comes from a set $\mathcal{X} \equiv [\underline{x}, \overline{x}]$ with $\underline{x} < 0 < \overline{x}.$ There is a null prize, $x = 0$, leaving the decision maker with the status quo. [28] Each gamble can be represented by an S component vector $\mathbf{x}$ of the prizes. The s^{th} component of $\mathbf{x}$ indicates the prize received when outcome state s is realized. In addition, there is a set $\mathcal{P}$ of discrete lotteries with objective probabilities that award the same prizes. Lotteries are described by the pair $(\mathbf{p}, \mathbf{x})$ giving the probabilities of each of several different prize levels. The decision maker has a preference relation $\succcurlyeq$ that is complete and transitive defined over $\mathcal{G} \cup \mathcal{P}$.

The goal is to find utility functions, $\hat{u}$ and u, and subjective probabilities, $\boldsymbol{\pi}^*$, such that

$$\mathbf{x} \succcurlyeq \mathbf{y} \Leftrightarrow \hat{u}(\mathbf{p}) \geq \hat{u}(\mathbf{q}) \Leftrightarrow \sum \pi_s^* u(x_s) \geq \sum \pi_s^* u(y_s)\,. \tag{79}$$

Note that the utility function $\hat{u}$ is defined over the gambles (and is an ordinal function) while u is defined over the prizes (and is a cardinal function). It is the latter function that is used in utility maximization under subjective probabilities as shown in the final relation in (79). The function $\hat{u}$ is only used in the development.

There are three obvious conditions we need to start:

[28] Upper and lower bounds and a status quo prize are not necessary, but simplify the presentation. Similarly, lotteries with continuous distributions can be allowed with some added complexity.

SA1) **Preference relation:** The ordering $\succcurlyeq$ over $\mathcal{G} \cup \mathcal{P}$ is complete and transitive

SA2) **Archimedean and Independence Axioms for Lotteries:** Axioms A3c and A4 hold for $\succcurlyeq$ on $\mathcal{P}$.

SA3) **Monotonicity:** Increasing a gamble's prize in any state or increasing the prize for any outcome in a lottery with fixed probabilities always increases the preference. That is, if $\mathbf{x}_1 > \mathbf{x}_2$, then $\mathbf{x}_1 \succ \mathbf{x}_2$ and $(\mathbf{p}, \mathbf{x}_1) \succ (\mathbf{p}, \mathbf{x}_2)$

Axioms SA1 and SA2 assure that choices among lotteries with objective probabilities are made using objective expected utility maximization. Axiom SA1 also requires that there be a common preference relation among gambles and lotteries. If this were not true, then lotteries could not be used to help our assessment of gambles. These two axioms are obviously necessary to get the desired result. Axiom SA3 indirectly guarantees that always getting the best prize possible in a gamble is better than the gamble which is better than always getting its worst prize. This axiom is needed to assure that the likelihood of every event is between never and always; i.e., $\mathcal{S} \mathrel{\dot{\succcurlyeq}} A \mathrel{\dot{\succcurlyeq}} \varnothing,\ \forall A \in \mathcal{A}.$ This, too, is obviously required and unobjectionable. The second part of the axiom is to assure that gambles and lotteries can be compared.

This is almost enough to derive subjective expected utility maximization. Consider any gamble **x** and the subset of lotteries with the same prizes (**p**, **x**). The lottery that assigns all the probability to the best (worst) outcome is equivalent to a gamble that always gives the best (worst) outcome. Axiom SA3 assures us that the value of the actual lotter is between the value of these two lotteries. Because lotteries are evaluated by objective expected utility we now only need change the probabilities on the lottery until $\mathbf{x} \sim \langle \mathbf{p}, \mathbf{x} \rangle$. There must be such a lottery because objective expected utility is continuous in the probabilities for its lotteries. Unfortunately, we cannot conclude that $\boldsymbol{\pi}^* = \mathbf{p}$ because there are typically multiple lotteries with the same payoffs but different probabilities that give the same expected utility. Only for gambles with just two possible payoffs will the matching probabilities be unique. So we need a method to separate and reconstitute gambles.

Define two gambles to be disjoint if there is no state for which they both award a prize other than the null prize. Because whatever state is realized applies to both gambles and the null prize is $x = 0$, two disjoint gambles are combined into a third simply by summing their payoffs, $\mathbf{x} = \mathbf{x}_1 + \mathbf{x}_2$. Applying this recursively any gamble can be decomposed into a set each member of which has only non-null outcome in a single state, $\mathbf{x} = \sum \mathbb{1}_s x_s$. Each of these gambles is equivalent to a unique lottery $\langle (p_s, 1-p_s)', (x_s, 0)' \rangle$. As this value of p_s is unique, we can now identify $\pi_s^* = p_s$. We can then recombine the lotteries to $\mathbf{p} \equiv \langle (p_1, 1-p_1, p_2, 1-p_2, \ldots)', (x_1, 0, x_2, 0, \ldots)' \rangle$ or equivalently $\mathbf{p} \equiv \langle (p_1, \ldots, p_S, 1-\sum p_s)', (x_1, \ldots, x_S, 0)' \rangle$. Assuming this is valid is the final required axiom.

SA4) **Additivity:** For all disjoint gambles $\mathbf{x}$ and $\mathbf{y}$ and their equivalent lotteries $\mathbf{x} \sim \langle \mathbf{p}, \mathbf{x} \rangle$ and $\mathbf{y} \sim \langle \mathbf{q}, \mathbf{y} \rangle$, the combined gambles and lotteries are also equivalent. $\mathbf{x} + \mathbf{y} \sim \langle (\mathbf{p}', \mathbf{q}')', (\mathbf{x}', \mathbf{y}')' \rangle$

The existence of lotteries that are individually equivalent to $\mathbf{x}$ and $\mathbf{y}$ is guaranteed by the first three axioms as previously discussed. It is the relation between the combinations that is important. This axiom does two things. First it is required to assure that the subjective probability of the union of two mutually exclusive events is the sum of the separate probabilities, which is an obvious restriction. Second, like the Independence Axiom, it assumes implicitly that only the state-by-state outcomes matter in the evaluation of a gamble. The mechanism or structure of determining the award is irrelevant.[29] Sufficiency of these conditions can now be demonstrated directly. The subjective probabilities are determined as shown and the utility function can be determined directly just by considering objective lotteries.

The resulting utility is state independent; that is, the utility of x is the same in every state. The contribution to expected utilities will differ of course, if states have different subjective probabilities of occurring. This method cannot be used to derive state-dependent utilities because the lotteries are resolved with objective probabilities so the proposed indifference match between $\sum p_i u(x_i)$ and $\sum \pi_s^* u_s(x_s)$ when $x_i = x_s$ is insufficient to guarantee that $\pi_s^* = p_s$ when $u_s(x_s)$ depends on the state as well as the prize in the state. However, state dependent subjective expected utility can also be derived axiomatically.[30]

The Ellsberg Paradox

Just as the expected utility evaluation does not live up its ideal in practice, subjective probabilities also display inconsistencies in application. The most common problem is illustrated by the Ellsberg (1961) Paradox. In its starkest, though not original, form the decision maker is offered a prize if he can guess the color of a ball to be picked from an urn. There are 300 balls that can be distinguished only by color. One hundred of the balls are red, and the remaining 200 balls are white and blue in unknown proportions. In the first decision, participants are to choose between A, winning \$100 if red is picked, and B, winning \$100 if white is picked. In the second decision, they are asked to choose between C, winning \$100 if either red or blue is picked, and D, winning \$100 if white or blue is picked. The amount won for each color of ball picked is shown of each choice in the table.

	Amount Won		
	red	white	blue
A:	100	0	0
B:	0	100	0
C:	100	0	100
D:	0	100	100

The vast majority of people offered a challenge like this strictly prefer to choose the color or color combination with known odds of winning. They choose A over B with a known $^1/_3$ chance of winning on red rather than white with winning chances between 0 and $^2/_3$. Then they pick D, white plus blue, with a known $^2/_3$ chance of winning rather than C, red with blue, with an unknown chance of winning between $^1/_3$ and 1. However, there are no subjective probabilities that could result in those choices. If they believe there are an equal number of white and blue balls, they should be indifferent both cases. If they think there are more white than blue balls, they should pick B and D. If they think there are more blue than white balls, they should pick A and C. And these would both be strict preferences.

To be precise, if w is the fraction of balls that are white, then the probability of receiving \$100 with choice B is w while the probability of winning \$100 with choice C is $^1/_3 + w$. So

$$A \succ B \quad \Rightarrow \quad \tfrac{1}{3} > w \qquad D \succ C \quad \Rightarrow \quad \tfrac{2}{3} > \tfrac{1}{3} + w\,. \tag{80}$$

Both choices cannot be correct. In fact, either A and C must first-order stochastically dominate B and D or vice versa.[31] Because it is a first-order stochastic dominance, risk aversion or its

[29] Also like the Independence Axiom, this "obvious" property seems to be violated in practice as demonstrated by the Ellsberg Paradox discussed below.

[30] See, for example, Karni (1993).

[31] Strictly speaking the Ellsberg Paradox only shows that Savage's Theory of Choice does not describe choices. It does not show that people do not have consistent subjective probabilities. We only know that they do not use consistent subjective probabilities in determining expected utility.

absence plays no role in this Paradox.

Various hypotheses have been offered to explain this paradox and to create a choice theory that incorporates it. Just as the Allais Paradox centered on violations of the Independence Axiom, the Ellsberg paradox comes from violations of the Additivity Axiom. An intuition for the violation can be developed by treating differently risk when the odds are known and uncertainty when the odds are not known. This distinction is known as *Knightian Uncertainty*.

Knightian Uncertainty, Ambiguity Aversion, and MMEU

Knightian Uncertainty is named for economist Frank Knight (1885–1972). His work emphasized the distinction between risk and uncertainty. Under his definitions, risk describes variation when the probability distribution is known objectively while uncertainty covers variation for which the probability distribution is unknown and therefore, at best, subjective. As illustrated by the Ellsberg Paradox, it appears that economic agents often make decisions differently in these two cases. In particular, they seem to prefer gambles for which the payoff distribution is objective over Bayesian equivalent gambles for which the payoff distribution is subjective. This preference is exactly the type expressed in the Ellsberg Paradox and is called *Ambiguity Aversion*.

Knightian uncertainty is often identified with maximization of the minimum expected utility (MMEU) though the latter is only one possible resolution of the uncertainty. Under MMEU people, when faced with unknown odds, evaluate various choices using a worst-case probability distribution for each.

Gardenfors and Sahlin (1982) described this theory. The basic idea is that rather than a probability distribution over outcome states there is a set $\mathcal{P}$ of probability distributions. The evaluation of a risky prospect is

$$V(\tilde{x}) = \min_{P\in\mathcal{P}} \mathbb{E}^P[u(\tilde{x})]\,; \tag{81}$$

that is, the gamble is assigned the smallest expected utility computed with the various probability distributions in $\mathcal{P}$. For example, suppose that a risk-neutral the decision maker in the Ellsberg is sure that the urn holds between 50 and 175 white balls. Then the expected utility of picking white and blue are

$$\mathbb{E}u_{\text{white}} = \min_{\frac{1}{6}\le\pi\le\frac{7}{12}} \pi\cdot 100 = 16\tfrac{2}{3} \quad \mathbb{E}u_{\text{blue}} = \min_{\frac{1}{12}\le\pi\le\frac{1}{4}} \pi\cdot 100 = 8\tfrac{1}{3} \tag{82}$$

Both of these choices are rejected in favor of choosing red with expected utility of $33^1/_3$. For the second set, the expected utilizes are

$$\mathbb{E}u_{\text{red or white}} = \min_{\frac{1}{2}\le\pi\le\frac{11}{12}} \pi\cdot 100 = 50 \quad \mathbb{E}u_{\text{red or blue}} = \min_{\frac{5}{12}\le\pi\le\frac{7}{12}} \pi\cdot 100 = 41\tfrac{2}{3}\,. \tag{83}$$

Both of these choices have smaller expected utility then the uncertainty-free choice of blue or white with an expected utility of $66\tfrac{2}{3}$. Gilboa and Schmeidler (1989) axiomatized MMEU and examined it closely.

One difficulty is that MMEU is extreme in looking at the minimum outcomes. Consider the following two gambles. Two urns each have 100 red or black balls. Urn A has between 40 and 50 red balls and urn B has between 35 and 70 red balls. You will be paid \$100 if you pick a red ball. What are the expected values of each choice? For a Bayesian, $40 \le \mathbb{E}u_{\text{A}} \le 50$ and $30 \le \mathbb{E}u_{\text{B}} \le 70$ so either could be better, though B is better under uninformative (or diffuse) priors because the expected number of red balls is 45 for urn A and 50 for urn B. The MMUE evaluations of the urns are 40 and 30 with A definitely better. This remains the MMUE choice even when urn B has between 39 and 100 red balls

Now suppose you are allowed to sample the urns with replacement first. You reach into each urn 10 times and pick out only red balls. How does your evaluation change? If you are a Bayesian, the expected payoff for both urns has increased, but B is looking better and better. Nevertheless, the MMEU evaluations are still 40 and 30. As sample continues, the Bayesian probability beliefs for the two urns converge to the correct answer (except with dogmatic priors), while the MMEU evaluations remain unchanged. If urn B holds more red balls than A, a Bayesian will always pick it after sufficient sampling, an MMEU decision maker never will.

Supplementary Notes

An Alternate Development of Utility Theory

An alternative development and entirely equivalent development of utility theory starts with the strict preference, $\succ$. In this case the two axioms are

A1′) **Asymmetry**: $\forall \mathbf{x}, \mathbf{y} \in \mathcal{X}, \ \mathbf{x} \succ \mathbf{y} \Rightarrow \mathbf{y} \not\succ \mathbf{x}.$ For all pairs of the choices, if **x** is strictly better than **y**, then **y** cannot be strictly better than **x**.

A2′) **Negative Transitivity**: $\forall \mathbf{x}, \mathbf{y}, \mathbf{z} \in \mathcal{X}, \quad \mathbf{x} \not\succ \mathbf{y} \text{ and } \mathbf{y} \not\succ \mathbf{z} \Rightarrow \mathbf{x} \not\succ \mathbf{z}.$ If **x** beats **y**, and **y** beats **z**, then **x** beats **z**.

Axiom (A2′) is obviously equivalent to (A2). It is perhaps not so obvious that (A1′) and (A1) are equivalent. That is left as an exercise to the reader. Once the weak preference is derived from the strict preference via $\mathbf{x} \not\succ \mathbf{y} \Rightarrow \mathbf{y} \succcurlyeq \mathbf{x}$, the remainder of the development is the same.

Alternatives to Expected Utility Theory

Once the Allais and other paradoxes were discovered, many alternatives to EUT were developed. Most of these relax the Independence Axiom. Some of them are described briefly.[32]

Rank-Dependent Utility — Quiggin (1982): RDU (originally called Anticipated Utility) alters the probabilities with a weighting function. Under RDU choices are made to maximize decision-weighted utility, $\sum_n \Omega(\sum_{i=1}^{n} \pi_i) - \Omega(\sum_{i=1}^{n-1} \pi_i) u(x_n)$. The decision weights are functions of the cumulative probability distribution of the outcomes so they are ordered $x_1 < x_2 < \cdots < x_n$. Empirically the weighting has an inverted S shape so that rare and extreme (either good or bad) events are overemphasized in evaluations.

RDU can explain the common ratio effect because decision weights do not necessarily adjust proportionally when probabilities do. The Allais Paradox can be explained similarly. RDU is discussed in detail in Chapter 11.

Chew-Dekel (Betweenness-Based) Preferences: A number of preference theories replace the Independence Axiom with the weaker Betweenness Axiom or some variation

$$\begin{aligned} &\mathbf{p} \sim \mathbf{q} \Rightarrow \mathbf{p} \sim \pi\mathbf{p} + (1-\pi)\mathbf{q} \sim \mathbf{q} \quad \forall \mathbf{p}, \mathbf{q} \in \mathcal{P} \\ \text{and} \quad &\mathbf{p} \succ \mathbf{q} \Rightarrow \mathbf{p} \succ \pi\mathbf{p} + (1-\pi)\mathbf{q} \succ \mathbf{q} \quad \forall \mathbf{p}, \mathbf{q} \in \mathcal{P}, \forall \pi \in (0,1) \,. \end{aligned} \tag{84}$$

Under the Betweenness Axiom there is no preference for or against randomization over indiffer-

[32] A few other alternatives to EUT are Moment Utility (Hagen, 1979), Optimism-Pessimism (Hey, 1984), Ordinal Independence (Segal, 1984), Dual Expected Utility (Yaair, 1987).

ent gambles. It bears this name because the indifference curves in a Machina Triangle are all straight lines. To see this note that any gamble that is a simple mixture of two others, $\mathbf{r} \equiv \pi\mathbf{p} + (1 - \pi)\mathbf{q}$ clearly lies on the line connecting them. Equation (84) guarantees that when **p** and **q** have the same utility, then all such **r** do as well verifying that the indifference curves are straight lines.[33] However, they need not be parallel as they are in EUT.

Chew-Dekel preferences are characterized by the Completeness, Transitivity, Archimedean, and Betweenness Axioms. When these axioms hold there exist two functions V and u, such that $\mathbf{p} \succcurlyeq \mathbf{q}$ if and only if $V(\mathbf{p}) \ge V(\mathbf{q})$ where V is defined implicitly by $V(\mathbf{p}) = \mathbb{E}_{\mathbf{p}}[\Gamma(\tilde{x}, V(\mathbf{p}))]$. When Γ depends only on x and not on **p**, then the Independence Axiom holds and EUT is valid with Γ as the regular cardinal utility function. Two non-EUT examples of Betweenness-based preferences are presented next.

Weighted Expected Utility — Chew and MacCrimmon (1979): Chew and MacCrimmon formulated Alpha-Utility Theory in their 1979 paper. [34] This has since grown into WEU. It is based on a specific form of betweenness that is a weakened form of independence or substitution

$$\forall \mathbf{p},\mathbf{q} \quad \mathbf{p} \sim \mathbf{q} \quad \Rightarrow \quad \forall \pi \in (0,1), \exists \rho \in (0,1) \text{ such that } \pi\mathbf{p} + (1-\pi)\mathbf{r} \sim \rho\mathbf{q} + (1-\rho)\mathbf{r} \quad \forall \mathbf{r}\,. \tag{85}$$

For the Independence Axiom π and ρ must be equal. Here they can differ, though ρ must be the same for all **r**.

Under weak independence, the preference relation over lotteries can be expressed as

$$V(\mathbf{p}) = \frac{\sum p_i w(x_i) u(x_i)}{\sum p_i w(x_i)} \equiv \sum p_i^w u(x_i) \quad \text{where } p_i^w \equiv \frac{p_i w(x_i)}{\sum p_i w(x_i)} \quad \forall w_i > 0\,. \tag{86}$$

WEU utility obviously reduces to EUT when the weights are constant.

WEU can be given the following interpretation though there is no presumption that this is how it is actually visualized. With no loss of generality, the weights can be scaled so they all are less than 1. Now consider a standard expected utility maximizer with utility function $u(x)$ who faces a lottery **p**. However, after nature has made the choice for a given prize x_i, the agent believes that nature consults a "hidden" lottery and only gives that prize with probability $w_i = w(x_i)$. With probability $1 - w_i$, nature "reneges" and redraws the original lottery independently of the first outcome. This continues until the "hidden" lottery indicates not to renege. The expected value of this lottery can be expressed recursively as

$$V(\mathbf{p}) = \sum p_i [w_i u(x_i) + (1-w_i) V(\mathbf{p})] = \sum p_i w_i u(x_i) + V(\mathbf{p}) \sum p_i (1-w_i)\,. \tag{87}$$

Because the probabilities sum to 1, the final term is $V(\mathbf{p})[1 - \sum p_i w_i]$, and (86) follows immediately.[35]

To further analyze WUT, consider lotteries with only three outcomes, $x_1 \prec x_2 \prec x_3$, whose utilities and weights are u_i and w_i. The indifference curve in the Machina triangle for utility V° is

---

[33] With k outcomes, lotteries can be described by a k–1 dimensional space analogous to the Machina triangle, and indifference curves will be k–2 hyperplanes.

[34] Specialized forms of WEU were described earlier by Samuelson (1950), Bolker (1966), and Jeffrey (1978).

[35] It is perhaps worth noting that in many experiments, participants are paid based only on the results of one or more trials selected at random. If they do not believe that the selection process is truly random, but rather that trials with some outcomes are more likely to determine the payment than others, then an EUT maximizer would behave exactly like a WEU maximizer.

$$p_3 = \frac{w_2(V^\circ - u_2) + w_1(u_1 - V^\circ)}{w_2(u_2 - V^\circ) + w_3(V^\circ - u_3)}\, p_1 + \frac{w_2(u_2 - V^\circ)}{w_2(u_2 - V^\circ) + w_3(V - u_3)}\,. \tag{88}$$

For EUT the weights are equal so the slope is $(u_2 - u_1)/(u_3 - u_2)$ which is independent of V° and all indifference curves are parallel straight lines. The slope in (88) is constant for a given V° so indifference curves are still straight lines in WEU, but the lines for different V° are not parallel as the slope depends on V°. The indifference curves can either fan out or fan in depending on the relative magnitudes of u_i and w_i.[36]

For continuous outcomes, a risk-premium for a zero-mean gamble, $\tilde{\varepsilon}$, can be defined in the usual way $u(x-\pi) \equiv \mathbb{E}[w(x+\tilde{\varepsilon})u(x+\tilde{\varepsilon})]/\mathbb{E}[w(x+\tilde{\varepsilon})]$. Using a Taylor expansion gives

$$\pi(x) = \left[-\frac{u''(x)}{u'(x)} - 2\frac{w'(x)}{w(x)} \right] \mathrm{var}[\tilde{\varepsilon}] \tag{89}$$

So the weighting effectively increases or decreases the Arrow-Pratt measure if it is decreasing or increasing in x. An exponential weighting function combined with exponential utility leads to constant absolute risk aversion before and after adjusting. Similarly, a power weighting function combined with power or log utility results in constant relative risk aversion.

Implicit Weighted Utility — Chew (1985) and Implicit Expected Utility — Dekel (1986): A further weakening of the Independence Axiom yields "Very Weak Independence"

$$\forall \mathbf{p},\mathbf{q},\mathbf{r} \quad \mathbf{p} \sim \mathbf{q} \quad \Rightarrow \quad \forall \pi \in (0,1),\ \exists \rho(\mathbf{r}) \in (0,1)\ \text{such that}\ \pi\mathbf{p} + (1-\pi)\mathbf{r} \sim \rho(\mathbf{r})\mathbf{q} + [1-\rho(\mathbf{r})]\mathbf{r}\,. \tag{90}$$

Here $\rho(\mathbf{r})$ can be different for different **r**'s. For Weak Independence, ρ must be constant while for the Independence Axiom ρ must be equal to π. So WEU and EUT are obviously special cases.

Very Weak Independence is still a form of the Betweenness Axiom. Utility choices can be represented as either

$$V(\mathbf{p}) = \frac{\sum p_i w(x_i, V(\mathbf{p})) u(x_i)}{\sum p_i w(x_i, V(\mathbf{p}))} \quad \text{or} \quad V(\mathbf{p}) = \sum p_i u(x_i, V(\mathbf{p}))\,. \tag{91}$$

The former is known as Implicit Weighted Utility. The latter is known as Implicit Expected Utility. What distinguishes these from WEU is that the weights depend on both the realized payoff and on the overall level of utility. As with WEU, the indifference surfaces or lines are flat but not necessarily parallel.

Disappointment Aversion — Gul (1991) and Routledge and Zin (2010): DA is a special case of Implicit Expected Utility. The original form was introduced by Gul (1991) with a generalization by Routledge and Zin (2010). Disappointment Aversion is characterized by

$$V(\mathbf{p}) \equiv \frac{\sum p_i w(x_i, V(\mathbf{p})) u(\tilde{x}_i)}{\sum p_i w(x_i, V(\mathbf{p}))} \quad \text{where} \quad w(x,V) = \begin{cases} 1 & u(x) \ge \beta V \\ 1+\alpha & u(x) < \beta V \end{cases} \quad \text{and}\ 0 < \beta\,. \tag{92}$$

When $\alpha = 0$, choices are as described by EUT.[37] Typically $\alpha > 0$ so those states where realized utility is low are over-weighted. Some other functional forms for w could be used as well. DA

[36] As WEU assumes transitivity for the preference ordering, the intersections of the indifference curves must lie outside the unit simplex, that is, at undefined lotteries with one or more probabilities negative. This can also be shown algebraically.

[37] This is also true if β is very large or very small so that $u(x) \ge \beta V$ or $u(x) < \beta V$ for all outcomes.

displays risk aversion when u is concave and $\alpha \geq 0$. For $\alpha < 0$ or non-concave u, there are some zero-mean risks that are preferred to the status quo.

DA offers one explanation of the Allais Paradox. Recall that the lotteries there were

$$A:\begin{cases} 33\% \text{ chance of } & 2500 \\ 66\% \text{ chance of } & 2400 \\ 1\% \text{ chance of } & 0 \end{cases} \quad \text{or} \quad B: \ 100\% \text{ chance of } 2400$$
$$C:\begin{cases} 33\% \text{ chance of } & 2500 \\ 67\% \text{ chance of } & 0 \end{cases} \quad \text{or} \quad D:\begin{cases} 34\% \text{ chance of } & 2400 \\ 66\% \text{ chance of } & 0. \end{cases} \tag{93}$$

The modal choices were $B \succ A$ and $C \succ D$. Suppose $u(x) = x$ so without disappointment, gambles would be valued at their expected values. In this case, the evaluations would be 2409, 2400, 825, and 816. So $A \succ B$ and $C \succ D$, and choices are consistent with EUT. However under DA with $\beta = 1$, there is disappointment about the bad outcomes of C and D and for the two lower outcomes for A. For $\alpha \geq 0.375$, disappointment is enough to switch the choice from A to B. But disappointment never switches the choice from C to D.

Local Expected Utility — Machina (1982): The models above that weaken the Independence Axiom were developed to account for observed preferences that violated it. Machina took a different, theoretical approach asking instead, "How robust are the concepts, tools, and results of expected utility was theory to failures of the Independence Axiom?" Starting with just the assumptions of a complete preference ordering and a stronger form of the Archimedean Axiom, along with a smoothness assumption,[38] he developed LEU (also called Nonlinear Expected Utility).

Under LEU, EUT is correct but only locally so. For a simple description of LEU, consider a finite set of prizes, then any gamble can be described by a vector **p** whose elements are the probabilities for each prize. The smoothness assumption is that the change in the value of a lottery **p** is locally linear with

$$V(\mathbf{p}+\Delta\mathbf{p}) - V(\mathbf{p}) = \sum \Upsilon(\mathbf{x};\mathbf{p})\Delta\mathbf{p} + o(\|\Delta\mathbf{p}\|). \tag{94}$$

So an infinitesimal change in probabilities causes a change in expected utility equal to the expected change in some utility function Υ. The difference between EUT and LEU is that in EUT, the function Υ depends only on **x** so the relation is exact regardless of the size of $\Delta\mathbf{p}$. Under LEU, Υ can vary with the initial probabilities as well so expected utility and the Independence Axiom are only local properties.

Because changes in expected utility are only locally linear, indifference curves in the Machina triangle do not even have to be straight lines. The simplest example of nonlinear LEU preferences has utility that is quadratic in the probabilities

$$\begin{aligned} V(\mathbf{p}) &= \sum p_i u(x_i) + \tfrac{1}{2}\alpha\left[\sum p_i v(x_i)\right]^2 \\ \Rightarrow \quad V(\mathbf{p}+\Delta\mathbf{p}) - V(\mathbf{p}) &= \sum u(x_i)\cdot\Delta p_i + \alpha\left[\sum p_i v(x_i)\right]\sum v(x_i)\cdot\Delta p_i + \alpha o(\|\Delta\mathbf{p}\|). \end{aligned} \tag{95}$$

When $\alpha = 0$, the change in utility is exactly linear in the changes in the probabilities.[39] In gener-

---

[38] In particular, he assumed the value function was Fréchet differentiable.

[39] Note also that the functions u and v cannot be the same. If they are and both are positive, which entails no loss of generality, then the ordinal $V(\mathbf{p}) = \Theta(\mathbf{p}) + \frac{1}{2}\Theta^2(\mathbf{p})$ is a monotonic transformation of $\Theta \equiv \sum p_i u(x_i)$, which is standard expected utility.

al, under LEU, as this example illustrates, the change in utility is only approximately linear in the changes in probabilities, and even in the limit of small probability changes, the change in utility depends on the initial probabilities as well as the particular outcome.

For lotteries with three outcomes, $L \prec M \prec H$, probability p_M can be substituted out from (95). The indifference curves in a Machina triangle are characterized by

$$\frac{dp_H}{dp_L} = -\frac{\partial V/\partial p_L}{\partial V/\partial p_H} = -\frac{u_L - u_M + \alpha(v_L - v_M)\sum p_i v_i}{u_H - u_M + \alpha(v_H - v_M)\sum p_i v_i}. \tag{96}$$

The indifference curves are no longer linear as the slopes depend on the probabilities at each point.

Regret Theory —Loomes and Sugden (1982): In EUT, the realized utility depends directly only on the outcome realized, and, perhaps, the choice made. A number of theories have been developed in which the possible choices and outcomes that did not occur also affect the realized utility. One of these theories is Regret Theory. Disappointment and regret sound similar, but the two theories differ. In DA, disappointment is measured relative to what the chosen random variable might have returned. In Regret Theory, regret depends on what other choices would have returned in the same situation.

The basic structure of regret theory is that there are gambles whose specific payoffs depend on a common set of states; x_{is} is the payoff on choice or gamble i in state s. To evaluate gambles, the agent uses a choiceless utility function, u, and a regret-rejoice function R. The former is the standard utility function developed for EUT. It assigns a utility to the level of wealth (or consumption, etc.) which does not depend on how that level was determined. The latter compares the utility realized to the utility that would have arisen had a different choice been made. This can be regret if the alternate choice would have done better or rejoicing if the alternate choice would have done worse. To choose between gamble i and gamble j, the agent makes the choice with the larger expected modified utility, $\mathbb{E}\left[R(u(\tilde{x}_i),u(\tilde{x}_j))\right]$.

Regret Theory can be illustrated with the three-state example in the table. Most people faced with the choices there choose the safe one, $\tilde{x}$, even though the risky outcome has a higher expected payoff. No regret is needed to explain that choice; risk aversion is sufficient. Regret and satisfaction is manifest if the realized utility for the same unit payoff on the safe choice is smaller in state A than in state B because the risky choice would have done better in state A. Similarly, satisfaction makes the realized utility of the safe choice greater in state C than in state B. The payoffs on the safe choice are the same so a standard utility function cannot explain this. Of course in this example, regret and satisfaction are not revealed by the choice.

state	*A*	*B*	*C*
probabilities	10%	80%	10%
$\tilde{x}$	1	1	1
$\tilde{y}$	5	1	0

In the specific development a regret utility function of the form $R(u_i,u_j) = u_i + Q(u_j - u_i)$ is used. Q is an increasing odd function with $Q(z) = -Q(-z)$. When $Q(z) \equiv 0$, Regret Theory is exactly the same as EUT. Typically it is assumed to have an inverted S shape, concave Q for $z < 0$ and convex for $z > 0$. This means that larger relative losses make a given choice less desirable. The inverted S shape for Q is based on empirical evidence; the theory works equally well with other functional forms.

state	*A*	*B*	*C*
probabilities	1/3	1/3	1/3
$\tilde{x}$	0	30	60
$\tilde{y}$	60	0	30
$\tilde{z}$	30	60	0

Regret Theory can lead to Common Outcome, Common Ratio, and Isolation Effect violations, but the major difference between Regret Theory and EUT is that choices may no longer be transitive. Suppose $Q(z) = \text{sgn}(z)z^2$, and the agent must choose between the three risky alternatives in the table. The entries are the realized utility of each outcome $u(x)$ before adjusted for regret. Each choice has an expected utility of 30 in isolation. When comparing $\tilde{x}$ to $\tilde{y}$, $\tilde{x}$ comes out worse by 60 one-third of the time and is better by 30 two-thirds of the time. So the regret function alters the realized utility of $\tilde{x}$ relative to $\tilde{y}$ by $\frac{1}{3}(-60^2+30^2+30^2)=-600$, making $\tilde{y}$ the preferred choice. In the same way, $\tilde{z} \succ \tilde{y}$ and $\tilde{x} \succ \tilde{z}$ so the preferences are not transitive.

The empirical evidence on Regret Theory is mixed. Harless (1992) suggests that regret is primarily a framing effect and is present most strongly when choices are presented in a form like the table above which emphasizes the state-by-state comparison.

Choquet Expected Utility is one axiomatization that justifies MMEU used under Knightian uncertainty. It starts by replacing the Additivity Axiom with a non-additive assessment called a capacity. The capacity is a function ψ on the algebra of events in Ω such that $\psi(\varnothing) = 0$, $\psi(\Omega) = 1$ and

$$\psi(A \cup B) + \psi(A \cap B) \ge \psi(A) + \psi(B)\,. \tag{97}$$

If (97) holds as an equality for all pairs of events, then the additivity Axiom is satisfied, and Choquet Expected Utility is the same as EUT. If, instead, (97) a strict inequality for some A and B, then choices depending on those two outcomes display ambiguity aversion.

The Ellsberg paradox discussed in the body of the chapter can be explained with $\psi(\{\text{red}\}) = \psi(\{\text{red or blue}\}) = \psi(\{\text{red or white}\}) = {}^1/_3$, $\psi(\{\text{blue}\}) = \psi(\{\text{white}\}) = 0$, $\psi(\{\text{blue or white}\}) = {}^2/_3$. The capacity equals the probabilities for the known events, but additivity fails because

$$\psi(\{\text{white or blue}\}) + \psi(\{\text{white and blue}\}) = \tfrac{2}{3} > \psi(\{\text{white}\}) + \psi(\{\text{blue}\}) = 0\,. \tag{98}$$

The set of additive functions $\mathcal{P}$ satisfying $P(A) \ge \psi(\text{A})$ for all $P \in \mathcal{P}$ is called the core of ψ. So

$$\psi(A) = \min_{P \in \mathcal{P}} P(A) \qquad \Rightarrow \qquad \mathbb{E}^{\psi}[u(\tilde{x})] = \min_{P \in \mathcal{P}} \mathbb{E}^{P}[u(\tilde{x})] \tag{99}$$

and the two theories are the same.

Robust Preferences are an extension of MMEU that avoids its extreme "pessimism". Robust preferences add a penalty function to the MMEU problem. The evaluation is

$$V(\tilde{x}) = \min_{P \in \mathcal{P}} \mathbb{E}^{P}[u(\tilde{x})] + \Gamma(P)\,. \tag{100}$$

Γ is a function that penalizes the choice of "extreme" probability distributions. Of course this requires a way of determining what distributions are extreme. One of the most used and studied penalty functions is the entropic penalty. The entropic penalty requires picking a benchmark distribution, P^*. A random variable, $\tilde{\omega}^P$, is then defined with a realization in each outcome state s equal to the ratio of the specific probability to benchmark probability[40]

$$V(\tilde{x}) = \min_{P \in \mathcal{P}} \mathbb{E}^{P^*}[\tilde{\omega}^P u(\tilde{x})] + \theta \mathbb{E}^{P^*}[\tilde{\omega}^P \ell\text{n}\, \tilde{\omega}^P] \quad \text{where } \omega_s^P \equiv p_s / p_s^*\,. \tag{101}$$

[40] For the benchmark distribution $\mathbb{E}^{P^*}[\tilde{\omega}] = 1$. It must assign a positive probability to every state. The other distributions in $\mathcal{P}$ need not do so.

For any random variable, the expectations in (101) and (81) are the same; i.e., $\mathbb{E}^{P*}[\tilde{\omega}^P \tilde{z}] = \mathbb{E}^P[\tilde{z}]$. When the penalty parameter is θ = ∞, the choice is made according to standard EUT with probability distribution *P**. At the other extreme, θ = 0, choices follow MMEU.

To illustrate consider the urn problem described in the text. To simplify the problem, assume that the probability of picking a red ball is a continuous variable between 40% and 50% for urn A and between 30% and 70% for urn B. The utility function is risk-neutral with $u(x) = x$. For any candidate probability density, $\omega^P(x) = p(x)/p^*(x)$. The penalty-adjusted MMEU problem for the urns is[41]

$$\begin{aligned} V(\tilde{x}) &= \min_{p(x)} \int_a^b \left[\frac{p(x)}{p^*(x)} x + \theta \frac{p(x)}{p^*(x)} \ell\text{n}\left(\frac{p(x)}{p^*(x)}\right)\right] p^*(x) dx + \lambda\left(1 - \int_a^b p(x) dx\right) \\ &= \min_{p(x)} \int_a^b p(x)\left[x + \theta[\ell\text{n}(p(x)) - \ell\text{n}(p^*(x))]\right] dx + \lambda\left(1 - \int_a^b p(x)dx\right). \end{aligned} \tag{102}$$

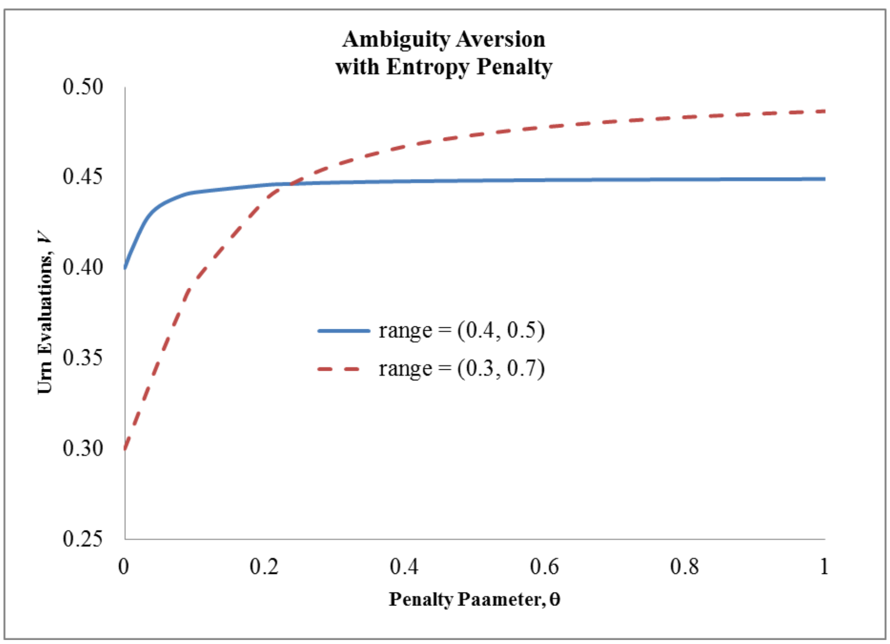
The first-order condition for $p(x)$ is

$$0 = x + [\ell\text{n}(p(x)) - \ell\text{n}(p^*(x))] + \theta - \lambda \quad \Rightarrow \quad p(x) = \frac{p^*(x)\exp(-x/\theta)}{\mathbb{E}^{P^*}[\exp(-x/\theta)]}. \tag{103}$$

Note that this relation is quite general. It holds for any choice of the benchmark distribution.[42] For an uninformative prior that is constant over the range, the value is

$$\begin{aligned} V(x) &= \frac{\int_a^b (b-a)^{-1} e^{-x/\theta} x dx}{\int_a^b (b-a)^{-1} e^{-x/\theta} dx} \\ &= \frac{e^{-a/\theta}(\theta + a) - e^{-b/\theta}(\theta + b)}{e^{-a/\theta} - e^{-b/\theta}}. \end{aligned} \tag{104}$$

The figure illustrates how the values vary with the penalty parameter θ. When θ is near zero, Urn B with the larger minimum expected value is chosen. As θ increases, the evaluation of urn A rises relative to urn B because its expected payoff is larger. The crossover occurs around θ = 0.238.



---

[41] The final term is the constraint that the density integrates to 1.

[42] For nonlinear utility, (103) is $p(x) = p^*(x)\exp[-u(x)/\theta] \big/ \mathbb{E}^{P^*}[\exp[-u(x)/\theta]]$.

References

(A partial list — my apologies)

Allais, M. 1953. “Le comportement de L’Homme Rationnel Devant Le Risque: Critique des Postulats et Axiomes de L’Ecole Americaine.” *Econometrica* 21: 503-546.

Kahnemann, David, and A. Tversky. 1979. “Prospect Theory: An Analysis of Decision under Risk.” *Econometrica* 47: 263-297.

Kreps, David. 1988. *Notes on the Theory of Choice*. Underground Classics in Economics. Westview Press.

Machina, Mark. 1982. “‘Expected Utility’ Analysis without the Independence Axiom.” *Econometrica* 50: 277-323.

Moskowitz, Harold. 1974. "Effects of Problem Representation and Feedback on Rational Behavior in Allais and Morlat-type Problems.” *Decision Sciences* 5: 225-242.

Pratt, John. 1964. “Risk-Aversion in the Small and in the Large.” *Econometrica* 32: 122-136.

Quiggin, J. 1982. “A Theory of Anticipated Utility.” *Journal of Economic Behavior and Organization* 3: 323-343.

Ross, Stephen. 1981. “Some Stronger Measures of Risk Aversion in the Small and in the Large.” *Econometrica* 49: 621-638.

Rothschild, Michael, and Joseph Stiglitz. “Increasing Risk I: A Definition.” *Journal of Economic Theory* 2: 225-243.

Savage, Leonard. 1954. *The Foundations of Statistics*. John Wiley and Sons, New York.

Slovic, P. 1969. “Maintaining the Attractiveness of a Gamble without Changing its Expected Utility.” *Journal of Experimental Psychology* 79: 139-145.

von Neumann, John, and Oskar Morgenstern. 1944. *Theory of Games and Economics Behavior*. Princeton University Press.