**BA 513: Ph.D. Seminar on Choice Theory**
**Professor Robert Nau**
**Fall Semester 2008**

**Notes and readings for class #2: axioms of subjective probability, expected utility, and subjective expected utility (revised January 11, 2008)**

In the last class we saw how ordinal utility functions, introduced by the marginalist economists in the 1870's, could be used to model choice problems involving purchase and exchange under conditions of certainty. In this class, we will consider the main elements of the theory of choice under *risk and uncertainty* that were developed in the mid-1900's: the theories of subjective probability, expected utility, and subjective expected utility. The primary readings are historical perspectives on these developments and are worth reading somewhat carefully. The secondary readings are excerpts from the original masterworks: they are worth skimming through (or at least reading the introductory parts) to get an appreciation for how the authors framed the problems for themselves.

**Primary readings**

1a. Excerpt on von Neumann's "sturm und drang" period from *Prisoner*'s *Dilemma* by William Poundstone
1b. "Classic and current notions of measurable utility" by Daniel Ellsberg (1954)
1c. "Retrospective on the utility theory of von Neumann and Morgenstern" by Peter Fishburn, 1989
1d. "The Invention of the Independence Condition for Preferences" by Peter Wakker and Peter Fishburn, 1995

**Supplementary reading**

2a. "Foresight: its logical laws, its subjective sources" by Bruno de Finetti, 1937 (mainly chapter I)
2b. Excerpt from *Theory of Games and Economic Behavior* by von Neumann and Morgenstern, 1947
2c. Excerpts from *The Foundations of Statistics* by Leonard J. Savage, 1954

## 2.1 Historical background

We have seen that some elements of rational choice theory had already been developed by the end of the 19th Century (and even much earlier). Probability theory and its application to problems of gambling was highly evolved, the concept of maximizing expected utility had been introduced (but had long lain dormant), and the idea that utility-maximizing agents should converge to an equilibrium in a competitive market had been explored. Several other major innovations remained to be made by the pioneers of the 1940's and 1950's, who laid the axiomatic foundations of modern rational choice theory. These innovations did not spring out of thin air, but were the logical extension of a number of theoretical developments that took place in the early 1900's.

a) **Axiomatic fever.** Around the turn of the century, the axiomatic method became *de rigueur* in mathematics, led by the efforts of David Hilbert, Bertrand Russell, and Alfred North Whitehead. Their ambition was to reduce all of mathematics—or as much of it as possible—to a relatively small set of primitive assumptions (axioms) from which all other facts could be deduced as theorems by applying principles of logic. For example, the existence of numbers, and the use of numbers in scientific models, was no longer taken for granted. Rather, one should begin with a notion of an abstract set whose elements were *ordered* in some way or another, and by imposing suitable axioms on the ordering relation, the possibility of assigning numbers to the elements of the set could then be deduced as a theorem. (The existence of real and complex numbers can be axiomatized in this way, starting from abstract, qualitative properties of sets of points.) The axiomatic method was soon embraced by various other disciplines that relied on mathematical methods. By the 1920's, in the writings of authors such as Frank P. Ramsey (a Cambridge mathematician and philosopher) and Ragnar Frisch (a Swedish economist who coined the term "econometrics" and later received the first Nobel prize in economics), the formulation of choice-theoretic models in terms of axioms of ordering was well established—something that was not seen in papers written by economists and statisticians only 20 years earlier. (By the way, the dream of axiomatizing all of mathematics foundered when Kurt Gödel proved his famous incompleteness theorem in 1931. For a perspective on this result, see *Gödel, Escher, Bach* by Douglas Hofstader.)

b) **Primitive preferences.** In the early 1900's, the idea of "utility" as a primitive psychological concept fell into disrepute, as we have already seen, because of doubts about the necessity or even the possibility of measuring it. In its place, economists began to emphasize the notion of *preference* as a psychologically primitive concept. An individual might not be able to assign a meaningful numerical utility to an object, but presumably she could always state which of two alternatives she *prefers*, or else state that she is precisely indifferent between them. (Indifference-curve analysis is based on this idea that the individual should be able to identify equivalence classes of commodity bundles among which she is indifferent, which she can then arrange in increasing order of preference.) Since preference is a *qualitative, binary relation*, it seemed to be ideally suited as the primitive operation on which an axiomatic theory of choice could be built. Thus, the same ordering relation that mathematicians were already using in areas such as number theory could be imported directly into choice theory, with the interpretation of preference now attached to it. *Indeed, modern choice theory is a product of the convergence of these ideas: the use of axiomatic methods in general—and axioms of qualitative ordering relations in particular—and the adoption of the concept of preference as a psychologically primitive concept.* (In the 1930's, Paul Samuelson developed an axiomatic theory of "revealed preference" in which the primitive concept is that of a "choice function" rather than a binary preference relation. Samuelson's approach is superficially different from the binary-preference approach used by Ramsey, von Neumann and Morgenstern, and Savage. Depending on the axioms that choice functions and preferences are assumed to satisfy, they may or may not have exactly the same implications for behavior, although for our purposes they are essentially equivalent modeling frameworks. The connections between them have been extensively explored in the writings of Amartya Sen.)

c) **Concepts of probability.** Although the idea of quantifying uncertainty in terms of numerical probabilities was well known in the context of gambling and insurance, it played essentially no role in marginalist (or for that matter, classical) economics. In the early 20th Century, economists began to recognize that uncertainty was central to the modeling of phenomena such as financial investments and entrepreneurial decisions, and John Maynard Keynes and Frank Knight wrote famous treatises on probability in the 1920's. At the same time, there was much disagreement (among statisticians and philosophers as well as economists) about the interpretation that should be attached to the concept of probability. The four best-known interpretations (at least three of which are still around today) are the *classical* interpretation, the *frequentist* (or *objective*) interpretation, the *logical* (or *necessary*) interpretation, and the *subjective* interpretation. Under the classical interpretation (articulated by early probability theorists such as J. & D. Bernoulli and Laplace), you try to decompose an uncertain situation into "equally likely cases" and then compute the probability of an event as the ratio of the number of "favorable" cases to the total number of cases. (This method is only applicable if there are considerations of symmetry that justify the assumption of equally likely cases.) According to the frequentist view, which underlies statistical methods of hypothesis testing, the probability of an event is its frequency of occurrence in a large (potentially infinite) number of repeated trials. (Of course, not all events are repeatable, which somewhat limits the applicability of this approach too.) According to the logical view, which was embraced by Keynes, probabilities are attached to sentences or propositions (rather than events per se) and are deduced by logical reasoning from the truth values assigned to their premises. (But it is not clear how the truth values of the premises are to be determined, as Ramsey pointed out in his criticism of Keynes. Keynes admitted that some probabilities were indeterminate under this approach, and he especially felt that it would be impossible to describe the behavior of the stock market in terms of logical probabilities.) According to the subjective view (which was advocated by Knight, Ramsey, de Finetti, and Savage, and which underlies the Bayesian methods of statistics that are becoming dominant today, probability is a subjective degree of belief that can be assigned to any event (repeatable or not), which can be measured by psychometric methods (such as eliciting preferences among gambles), and which ought to be updated by application of Bayes' rule upon receipt of new data. The subjective view has been criticized by those who feel that measurements of prior probabilities are often unreliable and by those who deny that reasonable people who have access to the "same" information should ever disagree about the value of a probability. Much of the literature in economics that deals with choice under uncertainty assumes that agents have mutually consistent "correct" beliefs about the relevant events, another issue to which we shall return.

**d) The independence condition for preferences ⇒ cardinal utility.** In attempting to establish a new game-theoretic foundation for economics, von Neumann and Morgenstern felt it was necessary to first axiomatize a concept of *cardinal* measurable utility that could serve as a "single monetary commodity" whose expected value the players would seek to maximize through their strategy choices. Recall from the previous class that choice under conditions of certainty is typically modeled using utility functions that are merely *ordinal*. Under such conditions, it is meaningless to say something like "the utility of *y* is exactly halfway between the utility of *x* and the utility of *z*." Von Neumann and Morgenstern showed that you *can* make a statement like this if you imagine that the decision maker has preferences not only among tangible consequences such as *x*, *y*, and *z*, but also among probability distributions over those consequences. Then you can say that *by definition*, a 50-50 gamble between *x* and *z* has a utility that is exactly halfway between the utility of *x* and the utility of *z*, and if *y* is equally preferred to such a gamble, then it too has a utility that is exactly halfway between the utility of *x* and the utility of *z*. There are several important caveats, though. First, you don't get something for nothing. You need much richer psychological data to construct a cardinal utility function, namely data about preferences among probability distributions over consequences, not just data about preferences among the consequences themselves. It is no longer enough for the decision maker to judge that she prefers an apple to a banana to a cantaloupe—she must also be able to determine a unique number $\alpha$ such that consuming a banana is equally preferred to an $\alpha$ chance of consuming an apple and a $1-\alpha$ chance of consuming a cantaloupe, which is not a trivial gustatory task. (The task is much easier in the special case where the consequences are merely different amounts money. Then the probabilistic mixture is a gamble, which is a familiar object of choice to most persons.) Second, those richer preferences must satisfy an *independence condition*, namely that *x* is preferred to *y* if and only if an $\alpha$ chance of *x* and a $1-\alpha$ chance of *z* is preferred to an $\alpha$ chance of *y* and a $1-\alpha$ chance of *z*, regardless of the value of *z*—i.e., a preference for *x* over *y* is independent of common consequences received in other events. (The independence condition is implicit in von Neumann and Morgenstern's axiomatization, but it was not fully appreciated until the early 1950's.) Third, once you have introduced probabilities into the definition and measurement of utilities, you have made a bargain with the devil, and you can't get rid of them again. You cannot attach any cardinal meaning to the utility values *except* in a context that involves risky choices: if the utility of *y* is found to lie exactly halfway in between the utilities of *x* and *z*, then you cannot conclude that the increase in happiness (or whatever) that results from the exchange of *x* for *y* is exactly the same as the increase that results from the exchange of *y* for *z*. All you can say is that *y* is equally preferred to a 50-50 gamble between *x* and *z*. Thus, the von Neumann-Morgenstern utility function is not the "same" utility function that had been sought by earlier generations of utilitarians and marginalists.

*However…* at around the same time that von Neumann and Morgenstern were writing and revising their book, several economists (including Sono, Samuelson and Leontief) discovered that a simple behavioral restriction on preferences-under-certainty implies that a consumer must have an additively-separable utility function, i.e., a utility function of the form $U(x_1, x_2, \ldots, x_n) = u_1(x_1) + u_2(x_2) + \ldots + u_n(x_n)$, which is unique up to an increasing linear transformation and is therefore cardinally measurable. (The marginalist economists of the 1870's had originally assumed that the utility function was additively separable, on grounds of intuition and/or mathematical convenience, although it was later pointed out by Edgeworth and others that this was economically unrealistic, since it ruled out effects of complementarity and substitutability between commodities.) The restriction on preferences that is needed for additive-cardinal-utility-under-certainty is simply another version of the independence condition implicitly assumed von Neumann and Morgenstern (and later explicitly assumed by Savage), although in the context of consumer theory it is known as "coordinate independence" or "factor independence" or "conjoint independence." As previously discussed in section 1.7 of the notes from class 1, the coordinate independence axiom requires that, when there are three or more commodities, the preference ordering among any subset of commodities is independent of common holdings of the other commodities—thus, the marginal rate of substitution between apples and bananas depends only on the current holdings of apples and bananas, not on holdings of cantaloupes.[1] A fully general proof of this result, and its connection with the theory of choice under uncertainty, was finally given in 1960 in a paper by Debreu on "Topological Methods in Cardinal Utility Theory." (The 1995 paper by Fishburn and Wakker, in this week's readings, provides more history.) The same independence condition, under the name of "preferential independence," provides the foundation for the theory of additive multiattribute utility developed in the 1970's by Keeney and Raiffa, Fishburn, and others. We will return to the multi-commodity version of the independence condition when we consider "state-preference theory," which is Arrow-Debreu's theory of choice under uncertainty that is obtained by applying consumer theory to state-contingent commodities. But the punch line is this: *preferences that are independent of common consequences received in other events or in other commodities are representable by an additive utility function that is cardinally measurable*.

**Comparative axioms of SP, EU, and SEU.**

The dominant theories of choice under risk and uncertainty are de Finetti's theory of subjective probability (SP), von Neumann and Morgenstern's theory of expected utility (EU), and Savage's theory of subjective expected utility (SEU). Savage's theory is a merger of the other two. Over the last 50 years there have been many revisions and variations of these theories, and depending on which book you read, you might see somewhat different sets of axioms, theorems, and proofs. I have drawn up a table with my own compilation, which is intended to highlight the similarities.

[1] Recall that consumer theory also assumes quite rich preferences. There the decision maker must be able to determine a unique number $\alpha$ such that consuming a banana is equally preferred to consuming an $\alpha$ fraction of an apple and a $1-\alpha$ fraction of a cantaloupe, although most persons find that a deterministic mixture of commodities (which is a tangible meal in its own right) is easier to think about than a probabilistic mixture.

# The axioms at a glance

| | de Finetti | von Neumann-Morgenstern (as revised by Jensen) | Savage |
|---|---|---|---|
| **Theory of:** | Subjective probability | Expected utility | Subjective expected utility |
| **Features of the environment** | States | Abstract set closed under probabilistic mixtures | States, consequences |
| **Objects of comparison** | Events $A$, $B$, $C$ (sets of states) | Probability distributions $f$, $g$, $h$ | Acts $f$, $g$, $h$ (mappings from states to consequences); *constant* acts $x$, $y$, $z$ |
| **Primitive relation** $\succcurlyeq$ | Comparative likelihood | Preference | Preference |
| **Reflexivity axiom** | $A \succcurlyeq A$ | $f \succcurlyeq f$ | $f \succcurlyeq f$ |
| **Completeness axiom** | For all $A$, $B$: either $A \succcurlyeq B$ or $B \succcurlyeq A$ or both | For all $f$, $g$: either $f \succcurlyeq g$ or $g \succcurlyeq f$ or both | P1a: For all $f$, $g$: either $f \succcurlyeq g$ or $g \succcurlyeq f$ or both |
| **Transitivity axiom** | For all $A$, $B$,$C$: if $A \succcurlyeq B$ and $B \succcurlyeq C$ then $A \succcurlyeq C$ | For all $f$, $g$, $h$: if $f \succcurlyeq g$ and $g \succcurlyeq h$ then $f \succcurlyeq h$ | P1b: For all $f$, $g$, $h$: if $f \succcurlyeq g$ and $g \succcurlyeq h$ then $f \succcurlyeq h$ |
| **Independence axiom (a.k.a. substitution, cancellation, separability, sure-thing principle)** | If $A \cap C = B \cap C = \varnothing$ then $A \succcurlyeq B$ iff $A \cup C \geq B \cup C$ | If $f \succcurlyeq g$ then $\alpha f + (1{-}\alpha)h \succcurlyeq \alpha g + (1{-}\alpha)h$ for all $\alpha$ strictly between 0 and 1. The converse is also true as a theorem, whence: $\alpha f + (1{-}\alpha)h \succcurlyeq \alpha g + (1{-}\alpha)h$ iff $\alpha f + (1{-}\alpha)h^* \succcurlyeq \alpha g + (1{-}\alpha)h^*$ (Compare to Savage's P2) | P2: For all $f$, $g$, $h$, $h^*$, if event $A$ is non-null, then[2] $Af + (1{-}A)h \succcurlyeq Ag + (1{-}A)h$ iff $Af + (1{-}A)h^* \succcurlyeq Ag + (1{-}A)h^*$ |
| **State-independent utility axiom ("value can be purged of belief")** | | | P3: For all constant acts $x$, $y$, $z$, if event $A$ is non-null, then $x \succcurlyeq y$ iff $Ax + (1{-}A)z \succcurlyeq Ay + (1{-}A)z$ |
| **Qualitative probability axiom ("belief can be discovered from preference")** | | | P4: For all events $A$ & $B$ and constant acts $x$, $y$, $x^*$, $y^*$ such that $x \succ y$ and $x^* \succ y^*$: $Ax + (1{-}A)y \succ Bx + (1{-}B)y$ iff $Ax^* + (1{-}A)y^* \succ Bx^* + (1{-}B)y^*$ |
| **Nontriviality axiom** | $\Omega \succ A \succ \varnothing$ for any $A$ that is "uncertain" | There exist $f$ and $g$ such that $f \succ g$ | P5: There exist $f$ and $g$ such that $f \succ g$ |
| **Continuity axiom** | There exists a fair coin | If $f \succ g \succ h$, then $\alpha f + (1{-}\alpha)h \succ g \succ \beta f + (1{-}\beta)h$ for some $\alpha$ and $\beta$ strictly between 0 and 1 (i.e., there is no heaven or hell nor lexicographic preferences) | P6: If $f \succ g$, then for any $x$ there exists a finite partition of the set of states into events $\{A_i\}$ such that $(1{-}A_i)f + A_i x \succ g$ and $f \succ (1{-}A_i)g + A_i x$ for any event $A_i$ in the partition (i.e., there is no heaven or hell and there exists a fair coin) |
| **Dominance axiom** | | | P7: If $f(s) \succ g(s)$ on every state $s$ in event $A$, then $Af + (1{-}A)h \succcurlyeq Ag + (1{-}A)h$ |
| **Conditional probability axiom** | If $A \subseteq C$ and $B \subseteq C$ then $A\|C \succcurlyeq B\|C$ iff $A \succcurlyeq B$ | | |

[2]My notation: $Af + (1\text{-}A)h$ means the act that agrees with $f$ on $A$ and agrees with $h$ on $A^c$.

In all three theories, there is a primitive ordering relation, denoted as usual by $\succcurlyeq$.[2] In SP theory, the objects of comparison are *events A*, *B*, etc., which are subsets of some grand set of possible *states of nature*. Thus, for example, *A* might be the event "France wins the next world cup" and *B* might be the event "England wins the next world cup." The ordering relation between events is one of *comparative likelihood*: $A \succcurlyeq B$ means that event *A* is considered "at least as likely" as event *B*.

In EU theory, the objects of comparison are *probability distributions f*, *g*, etc., over some set of possible *consequences*, which could consist of amounts of money or other tangible rewards or prizes, or they could simply be elements of some abstract set. Thus, for example, *f* might be a 50% chance of winning \$100 and a 50% chance of losing \$100, while *g* might be a 10% chance of receiving a free 5-course dinner at Maxim's and a 90% chance of getting thrown into the Seine River off the Pont Neuf, assuming that the set of prizes includes all four of these consequences. The ordering relation is one of *preference*: $f \succcurlyeq g$ means that distribution *f* is "at least as preferred" as distribution *g*.

SEU theory merges the main elements of these two frameworks: the objects of comparison are *acts*, which are real or imaginary mappings from states of nature to consequences. Thus, an act is an assignment of some definite consequence to each state of nature that might conceivably occur. For example, act *f* might be "win \$100 if France wins the next world cup, lose \$100 otherwise" while act *g* might be "win a free dinner at Maxim's if England wins the next world cup and get dunked in the Seine otherwise." (Actually, it is a bit more complicated than this: according to Savage, a consequence is not something you receive in addition to whatever else you already have. Rather, it is a complete description of a "state of the person", your final state of health, wealth, happiness, and whatever else you may care about. We will return to this thorny issue a few classes hence.) The ordering relation is *preference*, as in EU theory.

All three theories use some very similar-looking axioms. First, they all assume that the relation is *complete*: for any *A* and *B*, either $A \succcurlyeq B$ or $B \succcurlyeq A$ or both. In other words, you are always able to determine at least a weak direction of ordering between two alternatives—you are never entirely undecided. (This is a *very* questionable assumption, but we will let it pass for now.) Second, they all assume the ordering relation is *transitive*, i.e., $A \succcurlyeq B$ and $B \succcurlyeq C$ implies $A \succcurlyeq C$. (If you start by assuming that preferences are complete, it would be strange not to assume them also to be transitive, although some authors such as Fishburn have explored relaxations of this condition anyway.) Third, they all assume the relation has a property called *independence* or *cancellation*, which means that in determining the order of two elements (such as $A \succcurlyeq B$), there is some sense in which *you can ignore common events or consequences on both sides*.

The **independence condition for subjective probability** looks like this:

$$A \succcurlyeq B \Leftrightarrow A \cup C \succcurlyeq B \cup C \text{ for any } C \text{ that is disjoint from both } A \text{ and } B.$$

[2] I use the non-strict relation $\succcurlyeq$ throughout, whereas some authors use the strict $\succ$. This is just a matter of taste. You can get to the same results either way, except for some differences on sets of measure zero.

Thus, if you join the same non-overlapping event to two events whose likelihoods are being compared, you don't tip the balance of likelihood. For example, if you think it is at least as likely that France will win the world cup ($A$) as that England will win it ($B$), then you must think it is at least as likely that France-or-Italy ($A \cup C$) will win as that England-or-Italy ($B \cup C$) will win. This condition obviously also implies:

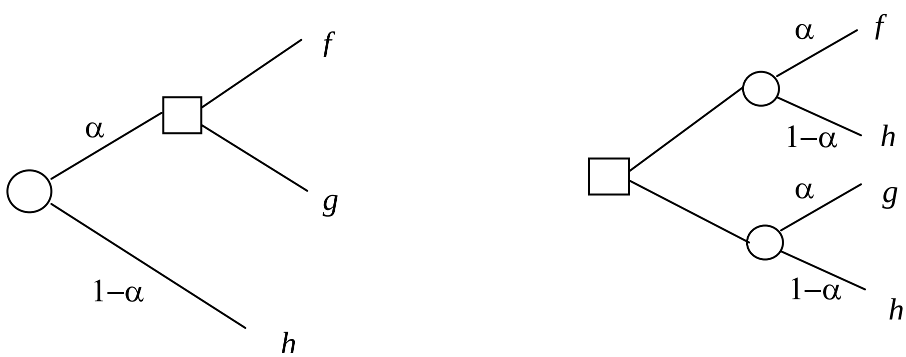
$A \cup C \succcurlyeq B \cup C \Leftrightarrow A \cup C^* \succcurlyeq B \cup C^*$ for any $C$, $C^*$ that are disjoint from both $A$ and $B$.

Thus, a common disjoint event on both sides of a likelihood comparison can be replaced by another common disjoint event without tipping the balance.

The **independence condition for expected utility** is:

$$f \succcurlyeq g \Leftrightarrow \alpha f + (1-\alpha)h \succcurlyeq \alpha g + (1-\alpha)h.$$

Here the expression $\alpha f + (1-\alpha)h$ denotes an *objective mixture* of the distributions $f$ and $h$—i.e., an $\alpha$ chance of receiving $f$ and a $1-\alpha$ chance of receiving $h$, where the chances are objectively determined as if by a roulette wheel. Thus, if $f$ and $g$ are objectively mixed with the *same* other distribution $h$, in exactly the *same* proportions, you don't tip the balance of preference between them. There is an obvious decision-tree interpretation of this condition, namely that you should choose the same decision branch (up or down) in the following two trees:




The argument goes like this: suppose you would choose $f$ over $g$ in the first tree, if the decision node were reached. Then presumably you would also choose $f$ over $g$ if you had to precommit yourself before knowing the outcome of the initial chance node (which might lead you involuntarily to $h$ instead). But then you are effectively choosing between the two lotteries depicted in the second decision tree. (Are you convinced? Would you always make the same choice in two such trees? It is possible to draw trees in which not everyone does so, as we will discuss a few classes hence. For example, it may matter whether the two chance nodes in the second tree represent the same event or two different events with the same probabilities.) The independence condition is necessary for the usual backward induction ("rollback") method of solving a decision tree. The condition also implies:

$$\alpha f + (1-\alpha)h \succcurlyeq \alpha g + (1-\alpha)h \Leftrightarrow \alpha f + (1-\alpha)h^* \succcurlyeq \alpha g + (1-\alpha)h^*,$$

which says that if $f$ is preferred to $g$ when they are both mixed with a common third distribution $h$, then the direction of preference isn't changed if they are mixed with a different common distribution $h^*$ instead. Thus, the common element $h$ can be replaced with any other common element.

The preceding result looks very similar to the **independence condition for subjective expected utility** (Savage's axiom P2, otherwise known as the "sure thing principle"), which says

$$Af + (1-A)h \succcurlyeq Ag + (1-A)h \quad \Leftrightarrow \quad Af + (1-A)h^* \succcurlyeq Ag + (1-A)h^*$$

for any non-null event $A$. (A non-null event is an event that has non-negligible probability in the sense that at least some preferences among acts are affected by the consequences received in that event.) Here, the expression $Af + (1-A)h$ is a *subjective mixture* of $f$ and $h$: it means you get the consequence specified by act $f$ in every state where event $A$ is true, and you get the consequence specified by act $h$ in all other states. So this condition says that if you are comparing two acts which happen to agree with each other in some states (namely the states in which $A$ is not true), then you can substitute different agreeing consequences in those states without tipping the balance. In other words, all you care about are the states where two acts don't agree with each other; in the states where they agree, you don't care *how* they agree. This property is necessary to be able to define the concept of a *conditional preference* between two acts: the relation $Af + (1-A)h \succcurlyeq Ag + (1-A)h$ means that $f$ is preferred to $g$ *conditional on event A*.

There is also **a second kind of independence condition** that also appears in SEU theory, namely an assumption about the *state-independence* of preferences for particular consequences. This assumption (Savage's P3) states that if $x$, $y$, and $z$ are *constant acts* (acts that yield the same consequence in every state of nature) and $A$ is any non-null event, then:

$$x \succcurlyeq y \quad \Leftrightarrow \quad Ax + (1-A)z \succcurlyeq Ay + (1-A)z$$

Thus, if the consequence $x$ is preferred to the consequence $y$ "for sure," then it is also preferred to $y$ conditional on any event, other things being equal. This assumption is needed to help separate utilities from probabilities—i.e., to ensure that relative utilities for consequences don't depend on the states in which they are received. (We will return to this question a few classes hence.)

The important thing to note is that, in all these conditions, *you can effectively ignore or cancel any common additive factors (e.g., C, h, z, etc.) on opposite sides of the ordering relation.* This property is necessary in order for the relation to be represented by a probability distribution that is *additive* across mutually exclusive events and/or a utility function that is a *linear in probability*. The additivity or linearity of the preference representation, in turn, means that the direction of preference or comparative likelihood between two alternatives depends only on the *difference* between them. This observation leads to an alternative—and much simpler—way of deriving the existence of a probability distribution and/or utility function from the axioms by using the separating hyperplane theorem, as we will see shortly.

**The dual relation between Subjective Probability and Expected Utility theory**

In the axioms-at-a-glance table above, an important difference between the axioms of expected utility and those of subjective probability and subjective expected utility is that numbers (namely, real-valued objective probabilities) appear explicitly in the EU axioms, but not in the SP and SEU axioms. The latter two sets of axioms are purely qualitative, and the fact that numbers show up in the representations of beliefs and preferences is a consequence of the axioms, and in particular, this requires the assumption that the spaces of events and acts are rich enough to be infinitely subdivided. This is what the continuity axioms are for: they guarantee that probabilities and utilities will turn out to be measurable on a continuum. Although purists may prefer to derive numbers from purely qualitative axioms, there are a number of practical advantages of letting numbers appear directly in the axioms. First it makes the content of the axioms more transparent and it leads to easier proofs of the theorems. But perhaps more importantly, it facilitates the measurement of probabilities and utilities, because it allows the decision maker to pull numbers directly out of the air (or wherever) while contemplating only a finite set of alternatives. For example, in EU theory, if $b$ and $w$ are considered to be the "best" and "worst" consequences, the decision maker can assign a numerical utility to an intermediate consequence $x$ by subjectively assessing the value of $\alpha$ such that $x$ is equally preferred to the probabilistic mixture $\alpha b + (1-\alpha)w$. Under the standard SP and SEU axioms, strictly speaking, a potentially infinite number of alternatives must be compared in order to determine a real-valued numerical scale of probability or utility. However, there are alternative axioms of SP and SEU that *do* explicitly involve numbers and which better highlight the structural similarities of the theories and the fundamental role of the separating hyperplane theorem.

In his famous 1937 paper, de Finetti himself pointed out that there are actually two equivalent ways to axiomatize the theory of subjective probability. One way is in terms of a binary relation of comparative likelihood (as illustrated in the table above). The other way is in terms of the *acceptance of monetary gambles*. The second way leads to simpler mathematics, and as I will show in a moment, the same mathematics can also be used to derive the theory of expected utility. The two theories are dual to each other—i.e., they are really the *same* theory but their working parts are merely labeled and interpreted in different ways.

Suppose that there is some set $S$ of mutually exclusive, collectively exhaustive states of nature, and let distributions of monetary wealth over those states be represented by vectors. Thus, if $\boldsymbol{f}$ is a vector representing a wealth distribution, then $f(s)$ is the amount of money you get if state $s \in S$ occurs. Now imagine that you have well-defined preferences among such wealth distributions, and as usual write $\boldsymbol{f} \succcurlyeq g$ to mean that $\boldsymbol{f}$ is weakly preferred to $g$. Behaviorally, this means that if you were initially endowed with wealth distribution $g$, you would be willing to *exchange* it for distribution $\boldsymbol{f}$, exactly as if you were offering to trade commodities. What axioms should such preferences satisfy? Let's begin with the familiar axioms:

**A0: (reflexivity)** $\boldsymbol{f} \succcurlyeq \boldsymbol{f}$ **for all** $\boldsymbol{f}$**.**
**A1: (completeness) for any** $\boldsymbol{f}$ **and** $\boldsymbol{g}$**, either** $\boldsymbol{f} \succcurlyeq \boldsymbol{g}$ **or** $\boldsymbol{g} \succcurlyeq \boldsymbol{f}$ **or both.**
**A2: (transitivity) if** $\boldsymbol{f} \succcurlyeq \boldsymbol{g}$ **and** $\boldsymbol{g} \succcurlyeq \boldsymbol{h}$**, then** $\boldsymbol{f} \succcurlyeq \boldsymbol{h}$**.**

Let's also assume that more money is strictly preferred to less, i.e., that preferences for wealth are strictly monotonic:

**A3: (strict monotonicity) if $f(s) \succ g(s)$ for all $s$, then $f \succ g$ (i.e., NOT $g \succcurlyeq f$ ).**

And just for good measure, let's impose

**A4: (independence) $f \succcurlyeq g \Rightarrow \alpha f + (1-\alpha)h \succcurlyeq \alpha g + (1-\alpha)h$ for all $h$ and $0 < \alpha < 1$.**

Is the latter a reasonable axiom in this setting, and what does it imply? First, note that in the presence of the completeness axiom, the independence axiom also works in reverse, i.e.,

$$\alpha f + (1-\alpha)h \succcurlyeq \alpha g + (1-\alpha)h \;\Rightarrow\; f \succcurlyeq g \text{ for all } 0 < \alpha < 1.$$

This is so because the completeness axiom requires the decision maker to have a definite direction of preference between $f$ and $g$, and if it were not the same as the direction of preference between $\alpha f + (1-\alpha)h$ and $\alpha g + (1-\alpha)h$, the independence axiom would produce a contradiction. (If we did *not* assume completeness, then it would be possible for the decision maker to have a definite preference between $\alpha f + (1-\alpha)h$ and $\alpha g + (1-\alpha)h$ but *no* definite preference between $f$ and $g$.) Next, by applying the independence axiom in both directions, we obtain the further implications:

$$f \succcurlyeq g \Rightarrow \alpha f \succcurlyeq \alpha g \quad \text{for all } \alpha > 0 \text{ (by choosing } h=0\text{), and also}$$

$$f \succcurlyeq g \Rightarrow f + h \succcurlyeq g + h \quad \text{for all } h.$$

This means that the decision maker effectively has *linear utility for money*, because scaling both $f$ and $g$ up or down by the same positive factor and/or increasing the decision maker's initial wealth by a constant amount $h$ does not lead to a reversal of preference. Since we are only worried about axiomatizing probability at the moment, we won't let that bother us. In any case, it is a reasonable assumption when differences among the wealth distributions $f$, $g$, and $h$ are not too large.

The preceding results lead to several more important conclusions. First, *the direction of preference between any two wealth distributions depends only on the difference between them.* In other words, if $f \succcurlyeq g$ and meanwhile $f^*$ and $g^*$ are two other distributions such that that $f^* - g^* = f - g$, then also $f^* \succcurlyeq g^*$. To see this, note that $f^* - g^* = f - g$ implies $f^* - f = g^* - g$. You can then get from $f \succcurlyeq g$ to $f^* \succcurlyeq g^*$ by adding $f^* - f$ on the left and $g^* - g$ on the right. Second, *inequalities between different pairs of wealth distributions can be added*: if $f \succcurlyeq g$ and also $f^* \succcurlyeq g^*$, then it follows that $f + f^* \succcurlyeq g + g^*$. Proof: (i) $f \succcurlyeq g$ implies $f + f^* \succcurlyeq g + f^*$ by adding $f^*$ to both sides; (ii) $f^* \succcurlyeq g^*$ implies $f^* + g \succcurlyeq g + g^*$ by adding $g$ to both sides; and (iii) $f + f^* \succcurlyeq g + g^*$ then follows by transitivity.

It is natural to refer to differences between wealth distributions as **gambles**, because exchanging one wealth distribution for the other means that you win money in some states and lose money in others relative to what you would have had otherwise. For clarity, we will henceforth call them "\$-gambles." By the preceding analysis, there are two kinds of \$-gambles: "acceptable" and "unacceptable." If $f \succcurlyeq g$, then $f - g$ is an acceptable \$-gamble, otherwise it is unacceptable. Let $G$ denote the set of acceptable \$-gambles. Then the axioms previously applied to preferences, and the implications subsequently derived from them, imply that $G$ satisfies the following axioms of its own:

**B0: (reflexivity) $\mathbf{0} \in G$**
**B1: (completeness) for any $x$, either $x \in G$ or $-x \in G$ or both**
**B2: (linearity) if $x \in G$, then $\alpha x \in G$ for any $\alpha > 0$**
**B3: (additivity) if $x \in G$ and $y \in G$, then $x + y \in G$**
**B4: (coherence) $G$ contains no strictly negative vectors**

*These are precisely the assumptions used by de Finetti in his gamble-based method of axiomatizing subjective probability, and (by the arguments above) they are logically equivalent to A0–A4.* Now we can apply the **separating hyperplane theorem** (which was used in the previous lecture to prove the welfare theorem): B0, B2 and B3 imply that $G$ is a convex cone whose vertex is the origin, B1 implies that it is at least a half-space, if not the whole space, while B4 says that $G$ does not contain any points in the *open negative orthant*, which is also a convex set. (Refer to Figure 1.8 in the notes from lecture 1.) Hence there is a hyperplane that separates $G$ from the open negative orthant, and $G$ must be (only) a half-space, so the hyperplane is unique, and moreover it passes through the origin. Let $\boldsymbol{\pi}$ denote the normal vector of this separating hyperplane, which is unique up to positive scaling. Then $\boldsymbol{x} \bullet \boldsymbol{\pi} \geqslant 0$ for all $x \in G$ and $\boldsymbol{x} \bullet \boldsymbol{\pi} < 0$ for all $x$ that are strictly negative. (Recall that $\boldsymbol{x} \bullet \boldsymbol{\pi}$ denotes the inner product, otherwise known as the dot product or matrix product or sumproduct, of the vectors $\boldsymbol{x}$ and $\boldsymbol{\pi}$. That is, $\boldsymbol{x} \bullet \boldsymbol{\pi} = \sum_{i=1}^{n} x_i \pi_i$ .) The latter condition means that $\boldsymbol{\pi}$ cannot have any negative elements, and without loss of generality we can normalize it so that it is a probability distribution. Then $\boldsymbol{x} \bullet \boldsymbol{\pi}$ is the expectation of the \$-gamble $\boldsymbol{x}$ according to the probability distribution $\boldsymbol{\pi}$. This analysis can be summarized as follows:

**SP THEOREM:** If preferences among wealth distributions over states satisfy A0-A1-A2-A3-A4 (or equivalently, if acceptable \$-gambles satisfy B0-B1-B2-B3-B4), then there exists a unique subjective probability distribution $\boldsymbol{\pi}$ such that $f \succcurlyeq g$ if and only if the expectation of the wealth distribution $f$ is greater than or equal to the expectation of distribution $g$ according to $\boldsymbol{\pi}$ (or, equivalently, $\boldsymbol{x}$ is an acceptable \$-gamble if and only if its expectation is non-negative according to $\boldsymbol{\pi}$) .

This is **de Finetti's theorem on subjective probability**, although de Finetti used a more indirect linear-algebraic proof that referred to determinants of matrices rather than separating hyperplanes.

De Finetti's theorem shows that someone who gambles rationally (and who also has linear utility for money—at least for small gambles) behaves "as if" she assigns definite subjective probabilities to events, and she gambles so as to maximize the expected value of her wealth.

Hence, subjective probabilities exist and they can be observed through behavior. De Finetti actually goes further to argue that *only* subjective probabilities exist. In other words, *all* probabilities are ultimately subjective: there is no such thing as "the" probability of an event. There is just my probability and your probability. We merely find it easier to agree on some probabilities—such as those of coin flips—than others. De Finetti's subjective theory of probability (which was further extended by Savage to cover situations involving non-linear utility and/or non-monetary consequences) provides the foundation for a theory of statistical inference that supports rational decision making under uncertainty. This theory is commonly called *Bayesian* statistical theory because it emphasizes the use of Bayes' rule to update "prior" subjective probabilities to obtain "posterior" subjective probabilities after observing sample information. This is to be distinguished from so-called *classical* or *frequentist* statistical theory which attempts to formulate rules of inference (hypothesis tests, etc.) that do not refer to anyone's subjective prior probabilities. De Finetti essentially proved that classical statistical methods are an "incoherent" basis for decision making because in some situations they will inevitably lead to the acceptance of strictly negative gambles. Hence, in practical terms, there is no rational alternative to using subjective probabilities in statistical reasoning.

Now let's reverse the role of money and probabilities. Let $C$ be some set of consequences or prizes, and let $f$, $g$, etc., denote objective probability distributions over those prizes. Thus, $f(c)$ now represents the *probability* of receiving *prize c* rather than an amount of money in a state of the world. Let's impose the same axioms as before, with one small difference:

**A0: (reflexivity)** $f \succcurlyeq f$ **for all** $f$
**A1: (completeness) for any** $f$ **and** $g$**, either** $f \succcurlyeq g$ **or** $g \succcurlyeq f$ **or both**
**A2: (transitivity) if** $f \succcurlyeq g$ **and** $g \succcurlyeq h$**, then** $f \succcurlyeq h$
**A3′: (non-triviality) there exist** $f^*$ **and** $g^*$ **such that** $f^* \succ g^*$ **(i.e., NOT** $g^* \succcurlyeq f^*$**)**
**A4: (independence)** $f \succcurlyeq g \Rightarrow \alpha f + (1-\alpha)h \succcurlyeq \alpha g + (1-\alpha)h$ **where** $0 < \alpha < 1$**.**

The small difference is that the strict monotonicity condition (A3) is weakened to a non-triviality condition (A3′): there should be *some* pair of alternatives such that one is strictly preferred to the other. We can't be any more specific about the alternatives for which this is true because we don't know a priori which prizes are definitely better than others. These are just von Neumann and Morgenstern's axioms of expected utility, as reformulated by Jensen (1967).

By the same logic as before, we conclude that the direction of preference between any two probability distributions depends only on the *difference* between those two distributions. Note that a difference between two probability distributions also a gamble in the sense that it changes your *risk profile*—it just does so by shifting the probabilities attached to fixed prizes rather than by shifting the amounts of money attached to events with fixed subjective probabilities. Let's call such gambles "*p*-gambles" to distinguish them from \$-gambles. You may notice that a *p*-gamble satisfies some implicit constraints that a \$-gamble does not. First of all, the elements of a *p*-gamble must sum to zero. This is unimportant—it just means that *p*-gambles lie in a lower-dimensional linear subspace. Second, the elements of a *p*-gamble cannot be larger than 1 in magnitude. This is also unimportant, because it would suffice to consider only "small" gambles anyway. Hence, we may conclude that there exists some set $G$ of "acceptable" *p*-gambles such that $f \succcurlyeq g$ if and only if $f - g \in G$, and where $G$ satisfies the following properties:

**B0: (reflexivity) $0 \in G$.**
**B1: (completeness) for any $x$, either $x \in G$ or $-x \in G$ or both.**
**B2: (linearity) if $x \in G$, then $\alpha x \in G$ for any $\alpha > 0$.**
**B3: (additivity) if $x \in G$ and $y \in G$, then $x + y \in G$.**
**B4′: (non-triviality) there is at least one non-zero *p*-gamble *not* in *G*.**

Now we proceed almost as before. Conditions B0, B2 and B3 imply that $G$ is a convex cone whose vertex is the origin. Condition B1 guarantees that it is at least a half-space in the linear subspace of vectors whose elements sum to zero, while B4′ prevents it from being the whole space, hence the bounding hyperplane is unique when restricted to this subspace. Let $\boldsymbol{u}$ denote the normal vector of this hyperplane (which is unique up to a positive scale factor), and note that the elements of $\boldsymbol{u}$ sum to zero because it is in the same subspace. Then $\boldsymbol{u}$ has the interpretation of a *utility function for prizes*, because $f \succcurlyeq g$ if and only if $f \bullet \boldsymbol{u} \succcurlyeq g \bullet \boldsymbol{u}$, and the quantities on the left and right of this expression are just the expectations of $\boldsymbol{u}$ under the distributions $f$ and $g$, respectively. The function $\boldsymbol{u}$ is subject to arbitrary positive affine transformations, because this will just affinely transforms all the expected utilities without changing the ordering of the probability distributions. This analysis can be summarized as follows:

**EU THEOREM:** If preferences among probability distributions over consequences satisfy A0-A1-A2-A3′-A4 (or equivalently, acceptable *p*-gambles satisfy B0-B1-B2-B3-B4′), there exists a utility function $\boldsymbol{u}$, unique up to positive affine transformations, such that $f \succcurlyeq g$ if and only if the expected utility of $f$ is greater than or equal to the expected utility of $g$ (or, equivalently, $x$ is an acceptable *p*-gamble if and only if it yields a non-negative change in expected utility according to $u$).

This is **von Neumann-Morgenstern's theorem on expected utility**, and we see that it is isomorphic to de Finetti's theorem on subjective probability, merely with the roles of probabilities and payoffs reversed. Both theorems are obtained by using the separating hyperplane theorem to separate a set of acceptable changes in wealth from a set of changes that are a priori irrational (namely, sure losses), just as we did in the case of the welfare theorem in lecture 1. Here, the wealth vector is a vector of monetary payoffs assigned to states *or* probabilities assigned to prizes, whereas before it was a vector of commodities possessed with certainty, but the rationality principle is exactly the same.

As I mentioned earlier, von Neumann and Morgenstern's theorem astonished the economic world when it was first presented, because it succeeded in reviving the previously moribund concept of cardinally measurable utility and making it once again a cornerstone of economic analysis (As Ellsberg famously remarked in 1954: "It appeared that a mathematician had performed some elegant sleight of hand and produced, instead of a live rabbit, a dead horse.") But von Neumann and Morgenstern's utility function is technically not the "same" utility function that had been sought by utilitarian economists of earlier generations. If $u$ denotes an individual's vNM utility function for wealth, this does not literally mean that $u(x)$ measures the individual's value for wealth $x$ (whatever that might mean). Rather, it is merely "as if" the individual assigns the value $u(x)$ to wealth $x$ prior to taking expectations *when making choices* *under conditions of risk* (objective probabilities). Thus, strictly speaking, a vNM utility function is an "index of choice under risk" rather than a subjective measure of pure value. But if it walks like a duck and quacks like a duck, it is tempting to go ahead and call it a duck, and certainly von Neumann and Morgenstern intended that their utility measure should be treated exactly as if it were a new kind of money.

In the preceding derivations, I've glossed over a slight technical condition, namely the requirement of *continuity* of the preference relation, which translates into *closure* of the set of acceptable $-gambles or *p*-gambles. Thus, we really ought to add the following:

**A5 (continuity): for fixed *f*, the set of *g* such that $f \succcurlyeq g$ is closed, and vice versa**

…and its counterpart:

**B5 (closure): *G* is closed.**

The continuity/closure requirement is needed to rule out unbounded utilities (e.g., heaven or hell as an outcome) as well as lexicographic preferences (e.*g*., between two commodity bundles you always strictly prefer the one that has more apples, regardless of the number of bananas, but if they have the same number of apples, then you prefer the one that has more bananas).

To summarize the results so far, expected utility theory is *usually* derived by posing the question: what condition should binary preferences satisfy beyond the usual axioms of reflexivity, completeness, transitivity, and continuity? The conventional answer is that they should also satisfy the *independence* condition. We have seen, however, that there is another way to characterize expected utility preferences—or for that matter, subjective probability beliefs. Transitivity and independence together are equivalent to the assumption that there exists a set of *acceptable gambles* (preferred changes in your wealth distribution over states or your probability distribution over prizes). A distribution *f* is then preferred to another distribution *g* if and only if the difference $f - g$ is an acceptable gamble. You satisfy the axioms of subjective probability or expected utility if and only if (a) your set of acceptable gambles is a convex cone, (b) it excludes at least some gambles—especially those that lead to a sure loss—and (c) it is always the same set of gambles, regardless of your current wealth or risk profile. The normal vector of the hyperplane separating the acceptable gambles from the unacceptable ones is just your subjective probability distribution or utility function.

**Subjective Expected Utility (a la Anscombe-Aumann)**

Next, let's reformulate Savage's subjective expected utility model in such a way that its primitive measurements also involve numbers, along the same lines as the SP and EU models above, following the approach of Anscombe and Aumann. The latter two authors observed that Savage's theorem could be derived more simply by merging Savage's concept of an act (a mapping of states to consequences) with von Neumann and Morgenstern's concept an act (an objective probability distribution over consequences). Anscombe and Aumann proposed that the objects of choice should be *mappings from states to objective probability distributions over consequences*. Such objects are known as "horse lotteries" in the literature. This is supposed to conjure up an image of a horse race in which each horse carries a different lottery ticket offering objective probabilities of receiving various prizes. The horse race will be run first to determine the lottery ticket that will be played—i.e., the state of the world you will be in—and that lottery will then be used to randomly select a prize. Your task is to determine your preferences among all possible ways of assigning objective lottery tickets to horses. By applying the von Neumann-Morgenstern axioms of expected utility to horse lotteries, together with one more assumption, you get probabilities and utilities together.

Just to fix ideas, suppose that you are offered various deals in which you may win some money depending on who wins the next World Cup, where the relevant states are considered to be "England," "France" or "some other country". Suppose that possible consequences (prizes) are $0, $500, and $1000. One example of a horse lottery would be the following:

| | *England* | *France* | *Other* |
|---|---|---|---|
| $0 | 0 | 0 | 1 |
| $500 | 1 | 1 | 0 |
| $1000 | 0 | 0 | 0 |

The numbers in the cells are the objective probabilities ("chances") of receiving a particular prize if a particular state occurs. If you possessed this horse lottery you would receive $500 if England or France won and would receive nothing otherwise. Another example of a horse lottery would be the following:

| | *England* | *France* | *Other* |
|---|---|---|---|
| $0 | 0.8 | 0.5 | 0.6 |
| $500 | 0.2 | 0.5 | 0 |
| $1000 | 0 | 0 | 0.4 |

If you possessed this one, then if England won you would have a 0.8 chance of receiving nothing and a 0.2 chance of receiving $500 (determined by, say, a subsequent roulette wheel spin); if France won, your chances of the same rewards would be 0.5 and 0.5, respectively; and if any other country won you would have a 0.6 chance of $0 and 0.4 chance of $1000. Notice that in choosing between this horse lottery and the previous one you would implicitly have to take into account your values for the three prizes as well as your beliefs concerning the likelihoods of the three states. Under the Anscombe-Aumann theory, you are assumed to start out with definite preferences among *all* such horse lotteries that might be specified with respect to a given set of states and a given set of prizes. Arguably it is hard to think of preferences among such complicated objects as being psychologically primitive. If you had to assess your preferences over a large set of horse lotteries, it would be easier to construct them by starting from an assignment of probabilities to states and utilities to prizes and then computing expected utilities, but according to the theory you start out with preferences that just happen to satisfy appropriate axioms of rationality, and from the axioms we infer that it is "as if" you assign probabilities and utilities and then compute expected utilities. That's the way the axiom game is played.

A convenient (in fact, essential) property of horse lotteries is that they can be mixed together in continuous proportions just like the probability distributions that are the objects of choice in von

Neumann and Morgenstern's EU theory. Thus, for example, a 50-50 mixture of the two preceding horse lotteries would look like this:

| | *England* | *France* | *Other* |
|---|---|---|---|
| \$0 | 0.4 | 0.25 | 0.8 |
| \$500 | 0.6 | 0.75 | 0 |
| \$1000 | 0 | 0 | 0.2 |

Notice that the number in each cell of this table is just a 50-50 mixture of the numbers in the corresponding cells of the previous two tables (e.g., 0.4 = 0.5(0)+0.5(0.8) in the upper left). This is the nature of the mixing operation that is performed whenever the independence axiom (A4) is applied to horse lotteries.

Henceforth, let *f*, *g*, and *h* denote arbitrary horse lotteries. Think of them as vectors (or matrices) with doubly-subscripted elements, where $f(c, s)$ denotes the objective probability that the horse lottery *f* assigns to consequence *c* when state *s* occurs (i.e., when horse *s* wins the race). Furthermore, let *x*, *y*, and *z*, denote horse lotteries that are "constant" in the sense that they yield the same objective probability distribution over consequences in every state of the world (i.e., no matter which horse wins). Let the usual axioms A0-A5 apply to preferences among horse lotteries, together with one additional axiom, which is essentially the same as Savage's P3:

**A6 (state-independence): $x \succcurlyeq y \Rightarrow Ax + (1-A)z \succcurlyeq Ay + (1-A)z$ for all constant *x*, *y*, *z* and every event *A***

Here, $Ax + (1-A)z$ denotes the horse lottery that agrees with *x* if *A* occurs and agrees with *z* otherwise, where *A* is an event (a subset of the states of the world). From axioms A0-A5 we immediately obtain, via the separating hyperplane theorem, the existence of an *expected-utility* vector *v* such that $f \succcurlyeq g$ if and only if $f \bullet v \succcurlyeq g \bullet v$ The elements of the vector *v* are subscripted the same as those of *f* and *g*, with $v(c, s)$ representing the expected utility of consequence *c* when it is to be received in state *s*. Intuitively, $v(c, s)$ combines the *probability* of the state with the *utility* of the consequence. Axiom A6 then implies the further restriction that the expected utilities assigned to different consequences in state *s* must be *proportional* to the expected utilities assigned to the same consequences in any other state, because it requires *conditional preferences* among objective probability distributions over consequences to be the same in all states of the world. It follows that *v* can be decomposed as $v(c, s) = p(s)u(c)$, where $p(s)$ is a (unique!) non-negative probability assigned to state *s* and $u(c)$ is a state-independent utility assigned to consequence *c*. The unique probability distribution is determined by normalizing the constants of proportionality of the utility functions in different states.

To see how this works, suppose that for the 3-state, 3-prize problem described above, after all your preferences have been assessed from many comparisons of horse lotteries, the elements of the unique expected-utility vector *v* turn out to look like this in tabular form:

| | *England* | *France* | *Other* |
|---|---|---|---|
| \$0 | −3 | −12 | −15 |
| \$500 | 1 | 4 | 5 |
| \$1000 | 2 | 8 | 10 |

Notice that $v$ has special structure. First, *the column sums are all zero*—i.e., the elements of $v$ sum to zero within each state. This property follows from the fact that $v$ can be assumed without loss of generality to lie in the same linear subspace as the set of differences between pairs of horse lottery vectors, whose elements sum to zero within each state because the probabilities assigned to consequences by each horse lotteries sum to 1 within each state. Second, *the columns are directly proportional to each other*—i.e., $v( . , s)$ is the "same" von Neumann-Morgenstern utility function over consequences in every state $s$, merely with a different positive scale factor. The latter property is implied by the state-independence axiom (A6), which requires conditional preferences among objective lotteries to be the same in every state. By convention, the *utilities* attached to particular consequences are assumed to be identical—not merely proportional—in different states, in which case the ratio of $v(c, s)$ to $v(c, s')$ can be interpreted as the ratio of the subjective *probabilities* of states $s$ and $s'$. (The importance of this convention is discussed in more detail below.) In the example above, the scale factors of the utilities in states 1-2-3 are in the ratios 1:4:5. By renormalizing the scale factors so that they sum to 1, we obtain the implied subjective probability distribution $p = (0.1, 0.4, 0.5)$. (Evidently you think France is 4 times as likely as England to win, and so on.) Finally, by dividing these probabilities out of the elements of $v$, we obtain the implied state-independent utility function $u = (-30, 10, 20)$. Of course, $u$ is ultimately subject to arbitrary positive affine transformations. For example, if we wanted to scale $u$ so that its lower and upper bounds were 0 and 1, we could write $u = (0, 0.8, 1)$ instead. Thus, the difference in utility between \$0 and \$500 is evidently 80% as great as the difference between \$0 and \$1000, as though you have diminishing marginal utility for money.

Notice that once the expected-utility vector $v$ has been determined, the expected utility of any horse lottery is obtained by merely calculating its inner product ("sumproduct") with $v$. Thus, for example, the first horse lottery above (which yields \$500 if England or France wins and \$0) otherwise has an expected utility of −10 (= 1 + 4 − 15), the second horse lottery (which yields a 0.4 chance of \$0 if England wins, etc.) has an expected utility of −11.2, and their 50-50 mixture has an expected utility of −10.6, which (naturally) is midway in between. (If the state-independent utility function is renormalized to a 0-1 scale, the expected utilities of the three horse lotteries become 0.4, 0.376, and 0.388, respectively.)

As a special case, the representation of preferences derived above applies to horse lotteries that place all the probability mass on a single consequence in every state, which are ordinary "Savage acts" (i.e., mappings from states to definite consequences). Thus, preferences among Savage acts are represented by a unique probability distribution $p$ over states and a state-independent utility function $u$ over consequences, with act $f$ preferred to act $g$ if and only if the expected utility of $f$ is greater than or equal to the expected utility of $g$. Voila! This analysis can be summarized as follows:

**SEU THEOREM:** If preferences among horse lotteries satisfy A0-A1-A2-A3′-A4-A5-A6, then there exists a unique probability distribution *p* and a state-independent utility function *u* that is unique up to positive affine transformations such that $f \succcurlyeq g$ if and only if the expected utility of *f* is greater than or equal to the expected utility of *g*.

This elegant result enables both probabilities and utilities to be directly measured by (for example) asking the decision maker to determine the objective mixture of best and worst consequences that is equally preferred to an arbitrary intermediate horse lottery, exactly as in the von Neumann-Morgenstern framework. However, as in the von Neumann-Morgenstern framework, the objects of comparison are highly artificial insofar as they involve objective randomization of things that are not objectively randomized in everyday experience. From a foundational perspective, a much more serious issue is that, as in Savage's model, the uniqueness of the derived probability distribution depends on the implicit assumption that consequences are not only similarly ordered by preference in every state of the world (which is the content of axiom A6 above), but they also have the *very same numerical utilities* in every state of the world—whatever that might mean. This assumption is problematic because it is not a property of preferences *per se*, and it cannot be confirmed or falsified by observing real choices. It is merely an arbitrary convention, which enables the expected-utility vector *v* to be conveniently decomposed into the product of a unique probability distribution *p* and a *state-independent* utility function *u*. Comparing your own utilities across different states of the world is conceptually similar to comparing the utilities of two different persons, which most theorists regard as difficult if not hopeless, a point that Aumann himself made in his famous 1971 letter to Savage. Savage defended his convention on the grounds that it is theoretically possible to define "consequences" in such a way as to make it true, but not everyone agrees. We will return to this point a few lectures hence.

### The Dutch Book Argument and incomplete preferences

The monetary-gamble version of de Finetti's subjective probability theorem is commonly known as the "Dutch Book Argument," abbreviated as DBA. Suppose you accept gambles for money in accordance with axioms B1, B2, and B3. This means that for any gamble *x* that might be proposed, you must be willing to accept either *x* or *–x*, if not both. Further, if you are willing to accept *x*, then you should also accept $\alpha x$, where $\alpha$ is a positive coefficient chosen (say) by a gambling opponent. And further still, if you are willing to accept *x* and *y* separately, then you should also be willing to accept them together—i.e., you should accept *x*+*y*. Now, do your acceptable gambles also satisfy B4? If so, then by the SP theorem, there is a unique probability distribution that rationalizes all your gambles in the sense that it assigns them all non-negative expected value, and you are evidently a probabilistic thinker who only accepts gambles that increase your expected wealth. If your gambles do *not* satisfy B4, then by definition you are willing to accept a strictly negative gamble—i.e., a *sure loss*, also known as a *Dutch book* for reasons that are obscure. (Even my Dutch colleagues don't know!) So if you *must* gamble, then to protect yourself against the possibility of a sure loss at the hands of a clever opponent, you should first assign probabilities to events and then accept only those gambles that have non-negative expected value.

The Dutch book argument has often been used as a hammer to drive home the point that rational individuals should have beliefs representable by subjective probability distributions. This bludgeoning tactic has been criticized by some philosophers, who object to the element of coercion implicit in axioms B1, B2, and B3. Why should you be *forced* to accept any gambles at all? The real problem is with the *completeness axiom*, B1, which says that for any gamble *x*, you must accept either *x* or –*x*. If B1 is dropped (which seems like a very good idea to me), then you may accept any gambles (or not) that you please, but the ones that you do accept are still subject to B2 and B3 within reasonable limits (small stakes). In this case, your set *G* of acceptable gambles generally won't be an entire half-space: it will be some smaller convex cone of vectors, as in Figure 1.7 of the notes from class 1. If you are rational, there will still be a hyperplane separating this cone of gambles from the open negative orthant, so as to avoid a Dutch book, but the separating hyperplane and its normal vector won't be unique. This means that your preferences will now be represented by a **convex set of probability distributions**, each of which assigns non-negative expectation to all of your acceptable gambles. The convex set of distributions and arbitrary positive multiples thereof is the so-called **dual cone** of the set of acceptable gambles. This result leads naturally to a theory of **lower and upper probabilities** (or more generally, lower and upper expectations for random variables), which many theorists today regard as a more plausible and more robust model of subjective beliefs.

To formalize the subjective probability model with incomplete preferences, suppose that you accept some finite collection of \$-gambles with respect to a finite set of states of nature. Let *M* denote the matrix whose row vectors are the payoff vectors of those gambles. If you subscribe to B2 and B3, then you will also accept any non-negative linear combination of these gambles, which can be represented by the matrix product $\boldsymbol{w} \bullet M$, where *w* is some vector of non-negative coefficients applied to different gambles. You violate axiom B4—i.e., you are exposed to a Dutch book—if there is some non-negative $\boldsymbol{w}$ such that $\boldsymbol{w} \bullet M < 0$ (i.e., $\boldsymbol{w} \bullet M$ is a strictly negative vector). By applying Lemma 1, the following "incomplete" theorems are obtained

**SP THEOREM with incomplete preferences:** If preferences among wealth distributions over states satisfy A0-A2-A3-A4-A5 (or equivalently, if acceptable \$-gambles satisfy B0-B2-B3-B4-B5), then there exists a convex set *P* of subjective probability distributions such that $f \succcurlyeq g$ if and only if the expectation of the wealth distribution $f$ is greater than or equal to the expectation of distribution $g$ for every distribution $\boldsymbol{\pi}$ in *P* (or, equivalently, $\boldsymbol{x}$ is an acceptable \$-gamble if and only if its expectation is non-negative for every distribution $\boldsymbol{\pi}$ in *P*) .

This result was shown by C.A.B. Smith in a 1961 paper and numerous subsequent authors have refined it further. Peter Walley's 1991 book *Statistical Reasoning with Imprecise Probabilities* recasts much of Bayesian theory in terms of lower and upper probabilities. Similar results are obtained when you drop the completeness axiom from EU or SEU theory: you end up with convex sets of utility functions or expected-utility functions.

**EU THEOREM with incomplete preferences:** If preferences among probability distributions over consequences satisfy A0-A2-A3-A4-A5 (or equivalently, acceptable *p*-gambles satisfy B0-B2-B3-B4-B5), there exists a convex set *U* of utility functions, unique up to positive affine transformations, such that $f \succcurlyeq g$ if and only if the expected utility of $f$ is greater than or equal to the expected utility of $g$ for every utility function $\boldsymbol{u}$ in $U$ (or, equivalently, $\boldsymbol{x}$ is an acceptable $p$-gamble if and only if it yields a non-negative change in expected utility for every utility function $\boldsymbol{u}$ in $U$ ).

**SEU THEOREM with incomplete preferences:** If preferences among horse lotteries satisfy A0-A2-A3-A4-A5-A6, then there exists a convex set $V$ of expected-utility functions, at least one[3] of which can be decomposed as the product of a probability distribution $p$ and a state-independent utility function $u$, such that $f \succcurlyeq g$ if and only if the expected utility of $f$ is greater than or equal to the expected utility of $g$ for every expected-utility function $v$ in $V$.

(For references on these results, see Aumann's 1962 paper "Utility Theory without the Completeness Axiom," Schervish, Seidenfeld and Kadane's 1995 paper "A Representation of Partially Ordered Preferences," Dubra et al.'s 2004 paper "Expected Utility Without the Completeness Axiom," and my 2006 paper "The Shape of Incomplete Preferences.")

**Aggregation of beliefs and "ex post" arbitrage**

An important and useful feature of the Dutch Book Argument, and the no-arbitrage principle more generally, is that it is a standard of rationality that applies to groups as well as to individuals—perhaps even better. As an example, consider the following hypothetical scenario. Suppose that a geological test is to be conducted prior to drilling an oil well, and three interested parties make the following statements:

> *Alice:* The chance of striking oil here is at least 40%.
>
> *Bob*: If the geological test turns out to be negative, the chance of striking oil can't be any more than 25%.
>
> *Carol*: Even if there's oil here, the chance of a negative test result is at least 50%.

The term "chance" is used here to refer to rates at which the agents are willing to bet money on the conditional or unconditional outcomes of events. Here, for example, Alice indicates that she is willing to bet *on* the event that oil is found at odds of 40:60 in favor. In other words, she will accept a small bet with net payoffs in the following proportions:

| | Positive, Oil | Positive, No Oil | Negative, Oil | Negative, No Oil |
|---|---|---|---|---|
| Payoff to Alice | \$3 | –\$2 | \$3 | –\$2 |

Meanwhile, Bob indicates that he is willing to bet against finding oil at odds of 75:25 against, *given* a negative test result. In other words, Bob will accept a small bet with net payoffs in the proportions:

[3] For technical reasons which are too obscure to go into here, when the completeness axiom is dropped from the AA version of the SEU model, it does not necessarily follow that *all* the elements of the representing set *V* can be decomposed into products of probabilities and state-independent utilities—another subtle axiom is needed for that.

| | Positive, Oil | Positive, No Oil | Negative, Oil | Negative, No Oil |
|---|---|---|---|---|
| Payoff to Bob | \$0 | \$0 | –\$3 | \$1 |

(Note that Bob's bet against finding oil is called off—i.e., the net payoff is zero—if the test result is positive.) Finally, Carol indicates that she is willing to bet on a negative test result at 50:50 odds, given that oil is found, meaning she will accept a small bet with net payoffs proportional to:

| | Positive, Oil | Positive, No Oil | Negative, Oil | Negative, No Oil |
|---|---|---|---|---|
| Payoff to Carol | –\$1 | \$0 | \$1 | \$0 |

Notice that we have only a very incomplete representation of any one person's beliefs. Alice (who is perhaps a geologist familiar with the area) has given information about her prior belief concerning the presence of oil. Bob (who is perhaps an oil driller with some experience in using the geological test under similar conditions) has given information about his posterior belief in the presence of oil, given a negative test result. Carol (who is perhaps a technician familiar with properties of the test) has given information about her belief concerning a "false negative" test result.

Note that no single agent has given a very complete account of his or her beliefs: each has given a bound on a single conditional or unconditional betting rate, which determines a convex set of probability distributions rather than a unique distribution. Alice's beliefs are represented by the set of all probability distributions for which the probability of striking oil is 40% or greater. Bob's beliefs are represented by the set of all probability distributions for which the conditional probability of striking oil, given a negative test result, is less than or equal to 25%. Carol's beliefs are represented by the set of all probability distributions for which the conditional probability of a negative test result, given that oil is present, is at least 50%.

Now consider a betting opponent who can make bets simultaneously with all three individuals. From the opponent's perspective, it makes no difference if the betting rates are posted by three different persons or by one person. Everyone's money is equally good—that's why money is useful. So, from the opponent's perspective, it is as if Alice, Bob, and Carol are speaking with one voice. Can the observer construct a Dutch book against them? By the Dutch Book Argument, he cannot do so if and only if there is at least one probability distribution that assigns non-negative expected value to all their bets. In this case, it so happens that there is a *unique* probability distribution that rationalizes the bets in this fashion, namely:

| | Positive, Oil | Positive, No Oil | Negative, Oil | Negative, No Oil |
|---|---|---|---|---|
| Probability | 20% | 0 | 20% | 60% |

From the observer's perspective, it is as if a single "representative agent" has revealed a unique subjective probability distribution on states of the world.

So, are Alice, Bob, and Carol behaving rationally? Maybe and maybe not! Notice that their "representative" probability distribution assigns zero probability to the event {Positive, No Oil}.

What happens if this event occurs? The answer is that there is an *ex post* arbitrage opportunity, because the observer can place a combination of bets that will never lose money for him but will *win* money for him if the event {Positive, No Oil} occurs, namely:

| | Positive, Oil | Positive, No Oil | Negative, Oil | Negative, No Oil |
|---|---|---|---|---|
| Payoff to Alice | $3 | –$2 | $3 | –$2 |
| Payoff to Bob (×2) | $0 | $0 | –$6 | $2 |
| Payoff to Carol (×3) | –$3 | $0 | $3 | $0 |
| **Total** | **$0** | **–$2** | **$0** | **$0** |

(The observer has taken the liberty of applying appropriate multipliers to the bets with Bob and Carol: the agents offered to accept any small bets with net payoffs *proportional* to those shown earlier, with non-negative multipliers to be chosen at the discretion of the observer.) A negative aggregate payoff for the agents is a profit for the observer. The observer is thus seen to earn a riskless profit if no oil is found following a positive test result.[4] I submit that in this case he would be justified in considering the players to have behaved irrationally. In fact, I will argue a few lectures hence that *no ex post arbitrage* is the "correct" standard of rationality to be applied in both single and multi-agent choice problems (including noncooperative games).

An objection that might be raised against the no-ex-post-arbitrage standard of collective rationality is that even if the observer earns a riskless profit, it does necessarily mean that any *individual* has behaved irrationally. Each will merely blame the others for betting at the wrong odds, thereby creating the observer's windfall. The players are entitled to their subjective beliefs, and their own bets may be rational by their own lights. But the rejoinder to this objection is that they are *all* guilty of irrationality because (as in the apples-bananas example from last week) they are all witnesses to all of the bets: any one of them can exploit the same opportunity as the observer. For example, suppose Alice takes the opposite side of the bets with Bob and Carol in the table above. Her payoff then would be as follows:

| | Positive, Oil | Positive, No Oil | Negative, Oil | Negative, No Oil |
|---|---|---|---|---|
| Payoff from Bob (x2) | $0 | $0 | $6 | –$2 |
| Payoff from Carol (x3) | $3 | $0 | –$3 | $0 |
| **Total** | **$3** | **$0** | **$3** | **–$2** |

The bottom line is the same bet she has already offered to accept—except that it yields $2 more if no oil is found after a positive test result! Meanwhile, Bob and Carol can do similarly. So it is as if the players all see a piece of paper lying on the ground saying "Get the bet of your choice *and* receive an extra $2 if no oil is found following a positive test result." If they all decline to pick it up, they evidently all believe it won't pay off—and if it doesn't pay off, that apparent

[4] In order to instantiate all of the gambles, it is necessary to know the outcomes of both events: performing the test and drilling for oil. However, if the test is performed but no drilling is performed following a negative result, then all of the gambles must be called off, and the aggregate payoff is the same as if drilling had occurred, namely zero. Hence, from the observer's perspective, it suffices that the test will be conducted and that drilling will occur if the result is positive.

belief is vindicated. But if it does pay off, they are all irrational ex post. Note that if any agent had attempted to exploit the arbitrage opportunity, the opportunity presumably would have evaporated: she would have raised the stakes until a point was reached at which she or some of the other agents would have wished to stop betting at the same rates. In so doing, she would have changed the rules of the game.

The preceding result, namely that the occurrence of an event which has been assigned zero probability (either implicitly or explicitly) creates an *ex post* arbitrage opportunity, is true in general, based on the following more specific version of Lemma 1 from class 1:

**Lemma 2:** For any matrix $M$, either there exists a non-negative vector $\boldsymbol{w}$ such that $\boldsymbol{w} \bullet M \leq 0$, with strict negativity in the $j^{\text{th}}$ position, or else there exists a (not necessarily unique) probability distribution $\pi$, which assigns strictly positive probability to state $j$, such that $M\pi \geq 0$.

The probability distribution of the representative agent has the appearance of a "common" probability distribution for the agents, but caution must be exercised in interpreting it that way. First, the rates at which agents are willing to bet money are influenced not only by their judgments of the relative likelihood of events but also by their valuation of money in those events. Other things being equal, an agent would rather bet money on an event in which, were it to occur, money would be worth more than in other events. Indeed, if we suppose that the agents are subjective expected utility maximizers, then their betting rates ought to equal the renormalized product of their true probabilities and relative marginal utilities for money—quantities that are called "risk neutral probabilities" in the finance literature.[5] The fact that betting rates may not be "true" probabilities is not necessarily problematic: betting rates are actually more useful than probability judgments because they suffice to determine monetary gambles that the agents will accept. We will discuss risk neutral probabilities in more detail in the coming weeks.

Another caveat in interpreting the representative agent's probability distribution as a commonly-held distribution is that (absent the completeness axiom) the individual agents are not required to have uniquely defined personal probabilities. In general, the betting rates offered by a single rational agent will be consistent with some convex set of probability distributions, and the *intersection of all such sets* is the set of probability distributions characterizing the representative agent. Hence, in general, the representative agent will have "sharper" beliefs than any of the individuals. Elster strident claims that it makes no sense to speak of "a family's beliefs" because "there are no such things." Here we see, to the contrary, that there is a precise sense in which a group may have better defined beliefs than any of its constituent individuals.

[5] Bob might become wealthy if he drills and strikes oil and nearly bankrupt if he drills and does not strike oil. Money earned from bets therefore might have higher marginal utility for him if he does not strike oil. For example, suppose that Bob's true posterior probability of striking oil after a negative test result is 50%, but his marginal utility for money is three times higher in the event of no oil. Then his risk neutral probability for not striking oil will be three times as great as his risk neutral probability for oil.

**Summary of main results and arguments thus far:**

- No-arbitrage is the fundamental principle of rationality and the separating-hyperplane theorem is the key mathematical tool underlying the theories of subjective probability, expected utility, and subjective expected utility, as well as competitive equilibrium under certainty.

- The effect of the independence condition is to guarantee that the utility function that represents preferences is *additively separable*, which in turn guarantees that your set of acceptable "gambles" (i.e., directions of preferred changes in wealth) is the same no matter what your current wealth position

- The completeness axiom is *inessential* to all these theories; if it is dropped, you merely get more sensible representations of preferences by *convex sets* of probabilities, utilities, or expected utilities, rather than by single points. The no-arbitrage principle (including the Dutch Book Argument) has even more normative force if completeness is not assumed.

- When beliefs are elicited in terms of acceptable monetary gambles, in the manner of de Finetti, they can be aggregated across individuals in a natural way.

- The slightly stronger principle of no *ex post* arbitrage is perhaps a more appropriate standard of rationality, according to which an individual or group is judged irrational if and when an event to which they assigned zero probability is observed to happen.

## GUIDE TO THE READINGS

### 1a. Excerpt on von Neumann's "sturm und drang" period from *Prisoner's Dilemma* by William Poundstone

This short passage provides some background on the axiomatic movement in mathematics in the early 20th Century and how it affected von Neumann, who was one of Hilbert's protégés. It also contains a famous passage from von Neumann concerning the evolution of mathematical models from the "classical" stage to the "baroque" stage. This passage has been cited by social scientists (e.g., Jon Elster) who worry that rational choice theory has long since reached the baroque stage.

### 1b. "Classic and current notions of measurable utility" by Daniel Ellsberg, 1954

This paper gives an excellent summary of the controversy that surrounded von Neumann and Morgenstern's resurrection of the concept of cardinally measurable utility. Ellsberg is a lucid and entertaining writer, and this paper was written while echoes of the original debates were still reverberating. By the way, this is the same Daniel Ellsberg who later cooked up "Ellsberg's paradox" and who still later played a key role in the downfall of President Richard Nixon. While working as an analyst at the Rand Corporation in the 1960's, Ellsberg leaked to the press a large bundle of classified military documents about the Vietnam War. The publication of these documents, which came to be known as the "Pentagon Papers," contributed greatly to the erosion of public support for the war. Nixon formed his clandestine band of "plumbers" to put an end to such leaks, and one of their first illegal errands was to burglarize the office of Ellsberg's psychiatrist in an effort to obtain embarrassing information. The plumbers were later caught red-handed while raiding the offices of the Democratic Party in the Watergate hotel, and the rest is history.

### 1c. "Retrospective on the utility theory of von Neumann and Morgenstern" by Peter Fishburn, 1989

This article places von Neumann and Morgenstern's contribution to utility theory in an historical perspective and discusses some of the controversy that their theory provoked at the time, with the benefit of more hindsight than was available to Ellsberg. It also gives a simpler account of the key technical ideas and presents Jensen's much simpler system of axioms for expected utility. (Study these rather than von Neumann and Morgenstern's version!)

### 1d. "The Invention of the Independence Condition for Preferences" by Peter Wakker and Peter Fishburn, 1995

Nowadays it is obvious that the independence condition for preferences is the key assumption in von Neumann and Morgenstern's theory of expected utility, but this was not immediately apparent at the time the first edition of their book appeared (1944), nor even when the more detailed second edition appeared (1947). The axiom did not appear in the book in anything like its modern form, which led to much confusion that was not cleared up until the early 1950's. In this article, Fishburn and Wakker trace the history of the "invention" of this idea, showing that it is central to several different branches of decision theory—including consumer theory as well as decision under risk and uncertainty—and that the underlying mathematical model had been developed long before von Neumann and Morgenstern used it in the context of decision under risk. They also review in great detail the history of the early years of expected utility, based on interviews they conducted with many of the original protagonists.

**2a. "Foresight: its logical laws, its subjective sources" by Bruno de Finetti, 1937 (mainly chapter I)**

This paper contains a number of seminal ideas. First it sketches the outline of theory of subjective probability expressed in terms of axioms of comparative probability (rather than preference). Next, the paper presents an alternative approach: "a direct, quantitative, numerical definition of the degree of probability attributed by a given individual to a given event, in such a fashion that the whole theory of probability can be deduced immediately from a very natural condition having an obvious meaning. It is a question simply of making mathematically precise the trivial and obvious idea that the degree of probability attributed by an individual to a given even is revealed by the conditions under which he would be disposed to bet on that event." That is, the probability that an individual attributes to an event $E$ is defined as the number $p$ such that, for an arbitrary positive or negative stake $S$, the individual would be willing to exchange the certain quantity of money $pS$ for a lottery in which she receives $S$ if $E$ occurs and zero otherwise. "This being granted, once an individual has evaluated the probabilities of certain events, two cases present themselves: either it is possible to bet with him in such a way as to be assured of winning, or else this possibility does not exist. In the first case one clearly should say that the evaluation of the probabilities given by this individual contains an incoherence, an intrinsic contradiction; in the other we will say the individual is coherent. *It is precisely this condition of coherence which constitutes the sole principle from which one can deduce the whole calculus of probability*" [emphasis added] The principle of "coherence" is otherwise known as the principle of "no arbitrage" or "no Dutch books": the individual's betting behavior should not present arbitrage opportunities or Dutch books to a clever opponent. (As we proceed through the course, we will see that the principle of coherence, or no-arbitrage, can be used as the foundation not only for the calculus of probability, but most of the rest of rational choice theory as well.) In the 1937 paper, de Finetti proves that coherent probabilities, defined in this manner, must obey the familiar additive and multiplicative laws of probability. His proofs are based on the evaluation of determinants of systems of linear equations. (These results are contained in Chapter I of the paper, which is all that is required reading.) Nowadays, the same results can be proved more easily by invoking the separating hyperplane theorem. De Finetti sidesteps the question of the distortions that would be introduced in his measurement scheme by diminishing (or otherwise state-dependent) marginal utility for money, and he admits that this is an apparent shortcoming, although the use of money bets has other practical benefits. (Later we will see that the "apparent" shortcoming may actually be a blessing in disguise.)

The later sections of the paper (which are *not* required reading!) introduce another seminal idea: the concept of *exchangeability*. Under a subjective interpretation of probability, events are rarely ever "independent" in the usual statistical sense. For example, if you repeatedly toss a coin whose fairness is unknown (or even better, an asymmetrical object such as a thumbtack), the tosses may be "independent" in a causal sense but probabilistically *dependent* in the sense that you learn about the process and modify your probabilities as data accumulates: your probability of heads on the second or third toss therefore may be different from your probability of heads on the first toss. However, the tosses are still subjectively *exchangeable* if you don't care about how they are labelled—i.e., if all that matters at any point is the total number of heads and tails observed so far. For example, if you regard the tosses as exchangeable, your probability of getting heads on the 4th toss after observing 2 heads and 1 tail in the first three tosses should be the same regardless of whether the actual sequence was HHT or HTH or THH. De Finetti proves a remarkable theorem stating that coherent probabilities assigned to long sequences of exchangeable events must be generated "as if" by a probability distribution over the unknown probability parameters. (The topic of exchangeability leads to deep water—if you are interested in learning more, see the chapter on it in David Kreps' book *Notes on the Theory of Choice*.)

**2b. Excerpt from *Theory of Games and Economic Behavior* by von Neumann and Morgenstern, 1947**

This book started the modern rational choice revolution by reviving the concept of cardinally measurable utility and proposing a theory of games as the foundation for economics. As was standard by that time, von Neumann and Morgenstern used the concept of *preference* as a primitive ordering relation. The objects of preference in their economic system are *probability distributions* over an abstract set of prizes, which might or might not be quantities of money. (Actually, the objects of preference are not explicitly assumed to be probability distributions—they are simply elements of some abstract set that is closed under the operation of taking mixtures—but the theory is clearly intended to apply to probability distributions over prizes.) Von Neumann and Morgenstern laid down a set of axioms for such preferences which implied the existence of a utility function, unique up to positive linear transformations, such that the more preferred of any two such objects was the one that yielded the higher expected utility. In the 1944 edition of the book, the axiomatization of expected utility was stated almost in passing. Von Neumann and Morgenstern wanted to have a quantity that would serve as "a single monetary commodity" for representing payoffs in games where randomized strategies could be used, but they were aware of the problem of diminishing marginal utility for money, so they made the assumptions necessary to derive the existence of a utility function whose expected value the players would wish to maximize. Their masterstroke was to introduce an "addition" operation for utilities in terms of probabilistic mixtures. The defining property of von Neumann Morgenstern utilities is that, if the utilities of prizes (or distributions) $x$ and $y$ are denoted by $u(x)$ and $u(y)$ respectively, then a lottery which yields an $\alpha$ chance of receiving $x$ and a $1-\alpha$ chance of receiving $y$, denoted $\alpha x + (1-\alpha)y$, must have utility $\alpha u(x) + (1-\alpha)u(y)$. That is,

$$u(\alpha x + (1-\alpha)y) = \alpha u(x) + (1-\alpha)u(y)$$

Thus, utility has the property of "linearity in the probabilities." It quickly became clear that this definition of utility was an important innovation in its own right, so von Neumann and Morgenstern added more details to the second edition published in 1947. Nevertheless, the additional details were still somewhat obscure, and over the next few years there was considerable debate about the content of their expected utility axioms. Eventually it was realized that the key property of *preferences* in von Neumann and Morgenstern's axiom system was the property now known as *independence*, namely that $x$ is preferred to $y$ if and only if $\alpha x + (1-\alpha)z$ is preferred to $\alpha y + (1-\alpha)z$, for any other prize $z$ and any number $\alpha$ strictly between 0 and 1.

*For a more accessible discussion of the vNM axioms, see the article by Fishburn. Actually, no one uses the axioms that were originally presented by vNM. Simpler but equivalent set of axioms have since been given by many other authors. Fishburn and many others endorse the version given by Jensen in 1967.*

Von Neumann and Morgenstern's concept of cardinally measurable utility is NOT the same concept that was envisioned by utilitarian philosophers and marginalist economists. The vNM utility function does not necessarily represent "strengths of preference" among prizes nor the quantity of "happiness" produced by additional increments of wealth under conditions of certainty. Rather, it describes an individual's preferences with respect to *risky* gambles among different prizes, which may reflect intrinsic aversion to risk as well as diminishing marginal utility for money or other commodities. (We will have more to say about this next week.) Von Neumann and Morgenstern's approach provides a justification for Bernoulli's use of an expected-value operation in conjunction with a utility function, which had no real justification under the utilitarians' and marginalists' interpretation of utility. Some contemporary economists were unhappy that von Neumann and Morgenstern even used the discredited term "utility" to describe the quantity whose existence followed from their axioms—they felt that vNM muddied the waters by using an archaic and "loaded" term for an important new concept. But by now we have a multitude of axiom systems for different kinds of ordinal and cardinal utility—utility under certainty, expected utility under risk, expected utility under uncertainty, multi-attribute utility, non-expected utility, etc.—so the concept introduced by Bernoulli and Bentham has become a tree with very wide branches indeed.

Strictly speaking, von Neumann and Morgenstern's utility theory is a theory of decision making under *risk*—i.e., under objectively known probabilities. As such, it has a rather limited range of applicability. Decisions under risk (according to the narrow, objectivistic interpretation of the term) are rarely observed in nature outside of casinos or parlor games in which mechanical randomization devices are used. But this is still a controversial matter, as Elster's pessimistic remarks about subjective probability indicate. Many economists, too, still take a basically objective view of probability, conceiving of the economy as an objective stochastic process. Von Neumann and Morgenstern mention in a footnote that a more general theory could undoubtedly be constructed using subjective probabilities, but it was left to Savage to carry out this step.

Of course, von Neumann and Morgenstern's main ambition was not to axiomatize probability or utility but to construct a theory of games that would serve as the foundation for economics, providing a connecting bridge between the "Robinson Crusoe" theory of the utility maximizing individual and the marginalists' theory of many small (and powerless) agents interacting in a competitive market. (Von Neumann had originally proved the minimax theorem for two-player games in 1928.) Their contributions in this regard were ultimately overshadowed by the later work of Nash, Shapley, Shubik, Harsanyi, Aumann, and others. Most of von Neumann and Morgenstern's book is now considered unreadable. However, their program of building a game-theoretic foundation for economics has gradually come to fruition: most current textbooks in microeconomics have a heavy game-theory flavor. Moreover, game theory has become the tool of choice in most rational choice applications outside of economics, which are typically concerned with interactions among small numbers of actors in non-market settings.

One more important thing to note about von Neumann and Morgenstern's book is that it is filled with analogies between economics and physics, which is hardly surprising, considering von Neumann's background. This is "social physics" in its most unabashed form. Von Neumann and Morgenstern compare the state of mathematical economics in the mid-20th Century to that of mathematical physics in the period just before Newton. What is perhaps most interesting is that they cross-pollinated economics with concepts from "classical" physics—with its mechanistic and rather static view of the universe—rather than with concepts of modern physics such as the *uncertainty principle* (according to which some quantities are intrinsically indeterminate because of interactions between the object and the measuring instrument on very small scales), the principle of *relativity* (according to which only relative, not absolute, measurements can be performed on very large scales) or the *entropy law* of thermodynamic systems (which governs the evolution of systems over time as well as the statistical properties of information).

### 2c. Excerpts from *The Foundations of Statistics* by Leonard J. Savage, 1954

Until very recently, the field of statistics was dominated by methods of inference and hypothesis testing pioneered by Neyman and Pearson, which are based on a frequentist interpretation of probability. Frequentist statistics asks and answers the backasswards question: given that hypothesis is true, what is the probability of the data? This approach to statistical inference is not axiomatically based and it is well known to be "incoherent" (in de Finetti's sense) as a basis for decision making—although experienced statisticians know where the traps lie and how to avoid them.

Savage set out to construct an axiomatic foundation for statistics that would integrate the theory of inference with a theory of coherent decision making, in order to be able to answer the more practical question: given the data, what is the probability that the hypothesis is true? (Or, even more practical: given the data, what *decision* should be made?) Savage felt that a fully general theory would need to combine a subjective interpretation of probability, as had been proposed by Ramsey and de Finetti, with a theory of expected utility along the lines developed by von Neumann and Morgenstern. Savage followed Ramsey and von Neumann and Morgenstern in using *preference* as the primitive psychological concept in his theory. But the objects of preference in his theory were not the abstract, probabilistic mixtures that appeared in von Neumann and Morgenstern's theory. In the true spirit of the axiomatic method, Savage sought to derive the existence of numerical probabilities from qualitative axioms imposed on preferences, rather than assume their existence on a priori grounds. Moreover, the objects of preference in his framework are concrete and easily identified with elements of real decision problems. Savage's theory starts from the assumption that there is a set of mutually exclusive and collectively exhaustive events called "states of the world" (a.k.a. states of nature) and a set of "consequences" that an individual might experience. A particular mapping from states of the world to consequences is called an "act," and corresponds to a real or hypothetical alternative that an individual might choose. For example, the states of the world might correspond to possible future values of stock or commodity prices, consequences might be sums of money, and acts might consist of the purchase or sale of various financial assets yielding different payoffs as a function of the state. An individual is then assumed to have preferences over the set of *all* real or imaginary acts that can be constructed from the given sets of states and consequences. As a practical matter, the set of states of the world and the set of consequences can only be defined with a finite amount of detail in any real application, which led Savage to introduce a very important distinction between "small worlds" that we can model and the "grand world" in which we really live. As we will see later, the validity and usefulness of Savage's framework hinges to a great extent on the question of whether it is possible to carry out the construction of a small world that adequately represents the grand world *and* satisfies the axioms he wishes to impose on it.

Savage's axioms are similar to de Finetti's axioms of comparative probability and to von Neumann and Morgenstern's axioms of expected utility up to a point. All of these axiom systems assume that the ordering relation (whether preference or comparative probability) is both *complete* and *transitive*. (Completeness means that any two objects are either strictly ordered or exactly equivalent—the individual is never unable to make up her mind. Transitivity means that if $x \succcurlyeq y$ and $y \succcurlyeq z$ then $x \succcurlyeq z$. That is, there can be no "cycles" of strict preference. These two conditions are combined as axiom P1 in Savage's system.) All three axiom systems also include an *independence* condition, which means essentially that common elements in any pair of alternatives can be ignored or cancelled out. (This is P2 in Savage's system, and it came under heavy assault in the non-expected utility revolution of the 1980's.) Savage's framework also requires several other specialized axioms, the most important of which are P3 and P4. These latter two axioms allow subjective probabilities to be disentangled from subjective utilities. (We will have much more to say about these axioms a few weeks hence.) The remaining Savage axioms, P5-P6-P7, are mainly technical conditions (nontriviality, continuity, and dominance) and are relatively uncontroversial.

Savage's model, and in particular its independence axiom, were immediately attacked by Maurice Allais, the French economist who had his own "neo-Bernoullian" theory of decision making in which preferences were determined not only by the mean value of utility but also by higher-order moments—e.*g*., the variance of utility. Allais concocted a famous example of a paradoxical decision problem in which most individuals choose to violate the independence condition. Savage discussed this example in his book and showed how it helped him to "correct an error" in his own thinking while remaining faithful to his axioms. (See the section on "Historical and critical comments on utility.") Others have remarked that they would just as soon satisfy their own preferences and "let the axioms satisfy themselves." Versions of the Allais paradox can be constructed both for Savage's and for von Neumann and Morgenstern's theories. In my opinion, the paradox is more compelling in von Neumann and Morgenstern's framework than in Savage's, but it is not the most important challenge to either theory. (This is a matter for further discussion in class.)

Whereas von Neumann and Morgenstern's theory of expected utility seems adequate for "designed games" in which the payoff matrix of the game and the players' choices of possibly-randomized strategies are common knowledge by construction, Savage's theory of *subjective* expected utility seems more suitable for the modeling of games that arise in nature, in which agents must form their own subjective beliefs about both the payoffs (i.e., consequences) and strategic intentions of their opponents. Yet, formally, Savage's model refers to a decision problem in which a single agent is engaged in a game against impersonal forces of nature. (We will delve deeper into the question of whether Savage's model is suitable foundation for game theory in another few weeks.)

A simpler derivation of the subjective expected utility model was given in 1963 by Anscombe and Aumann. The primitive objects of choice in their theory are so-called "horse lotteries," which are mappings from states of the world to objective probability distributions over consequences. Thus, a horse lottery is an amalgam of a "Savage act" and a "von Neumann-Morgenstern act." By a double application of the von Neumann-Morgenstern axioms to horse lotteries, they derive essentially the same representation as Savage. The Anscombe-Aumann formulation is mathematically simple and elegant, but it is not quite as "pure" as Savage's in the sense that it uses numbers (i.e., objective probabilities) directly in its axiom system rather than inferring their existence from qualitative relationships among acts that are specified entirely in set-theoretic terms. Additional elegant reformulations of subjective expected utility have been given by Peter Fishburn and Peter Wakker in a number of books and papers.

Savage's book and his other publications and collegial lobbying efforts succeeded in establishing an alternative paradigm of "Bayesian" statistics in the 1950's and 1960's. Bayes' theorem plays a distinguished role in the theory: it describes how "prior" beliefs are modified by data to become "posterior" beliefs. The essential difference between Bayesian and frequentist methods is that Bayesian methods are capable of incorporating the decision maker's subjective prior beliefs into the analysis in a systematic way. Both Bayesians and frequentists agree that the *likelihood function*, which expresses the probability of the data as a function of the unknown parameters, summarizes the information contained in the data. But frequentist methods stop at that point and use the likelihood function for purposes of inference; Bayesians use the likelihood function to multiply the prior distribution, yielding the posterior distribution. Savage was also instrumental in bringing de Finetti's work (much of which had been published only in Italian or French) to the attention of the anglophone world.

For several decades, Bayesian methods were widely regarded to be impractical because the calculation of posterior distributions required the evaluation of integrals that often lacked analytic solutions and had too many parameters to be evaluated by numerical methods. As a graduate student I took a course in Bayesian decision analysis from Dennis Lindley, one of Savage's early disciples, in which he lamented that it would never be possible to perform a numerical integration in as many as 10 dimensions. In the last decade or so, new computational methods and faster computers have emerged to surmount this problem. Using "Monte Carlo Markov chain" methods, it is now feasible to solve Bayesian models with *hundreds of thousands* of parameters, and Bayesian methods are on the cutting edge of developments in statistics and econometrics—especially here at Duke, where the Department of Statistical Science houses the foremost group of Bayesian statisticians in the world.