Notes on decision theory

Sam Wycherley*

December 20, 2023

These are a collection of notes covering topics in decision theory. They are drawn from a number of sources and I claim no originality to any of the content. The more classic material owes a lot to Fishburn (1970). I wrote these when I was interested in decision theory, but I subsequently lost interest (these days, I work on much less abstract things, like trying to improve the US Treasury market), so these are much less extensive than I had originally planned. Nevertheless, I wrote pretty extensive notes on a lot of classic decision theory, which is useful to know in detail for pretty much any economist, so I've decided to post them. I'd recommend Tomasz Strzalecki's decision theory notes for anyone who's interested.

Some housekeeping: I usually write "wlog" for "without loss of generality" and "iff" for "if and only if". If I write "s.t." then I mean "such that" and if I write "wrt", I mean "with respect to".

Contents

*Email: wycherley@stanford.edu.

I use the following scheme for numbering axioms and assumptions:

(**C***) Continuous preference axioms.

(**R***) Revealed preference axioms.

(**S***) Separability axioms.

(**V***) Von Neumann-Morgenstern expected utility theory axioms.

(**P***) Savage axioms.

(**E***) Extraneous probability approach axioms.

1 Decisions under certainty

Let X be a set of *alternatives*, $|X| > 1$. This is the set of possible decision objects. These might be bundles of goods or services, candidates running for election, risky bets or investment opportunities, job offers, meal choices, and so on – anything that you need to make a decision about in a given situation.

Given a set of alternatives X, there are three ways we might go about modelling the choices of a decisionmaker, assuming for now that the decisionmaker makes choices deterministically:

1. *Preference relations.* We begin with an exogenously given binary relation $\succ$ on X that directly defines the decisionmakers *preferences* (we will work with strict preferences $\succ$, but we could just as easily work with weak preferences $\succsim$). The decisionmaker (strictly) prefers an alternative x to y if $x \succ y$, and if the decisionmaker does not prefer one to the other, we say she is indifferent between x and y. We assume the decisionmaker makes choices that she most prefers.

   This is a fairly general approach, and we will frequently take preferences as basic throughout these notes.

2. *Utility functions.* We take a utility function $u : X \to \mathbb{R}$ on the set of alternatives X and assume the decisionmaker makes choices that maximizes her utility function. These days, the utility function is not taken to have any intrinsically meaningful interpretation (in the sense of, say, representing Benthamite utils of happiness). It is instead just a convenient tool for ordering the alternatives in X.

   This is the least general of the three approaches (there are preference relations that cannot be represented by a utility function), but it is by far the most convenient. It is effectively impossible to do much interesting economic theory by eschewing utility. Often in applied theory, economists (especially macroeconomists) are sloppy and you see them talk about "preferences $u(x) = \ldots$". Though this is not a cardinal sin (there is a unique mapping from utility functions to the preferences they represent), in these notes, we will not do that.

3. *Choice correspondences.* For different menus (subsets of X) that a decisionmaker might face, we take the choices that the decisionmaker would be willing to make as basic. This defines a choice correspondence $c(M) \subseteq M$ on each menu M. One advantage of this is that while we do not know the decisionmaker's preferences, we can observe the choices they make and attempt to infer their preferences from the resulting choice data. This exercise is at the heart of revealed preference theory.

We will go through these in the order above, starting with preference relations, relating preference relations to utility functions, and eventually relating choice correspondences to preference relations. These approaches suit different purposes. As mentioned, utility functions are extremely convenient for economic analysis, and it is difficult to say much without them. Choice correspondences provide the most general framework, though choice correspondences that cannot be "rationalized" by some preference relation are not particularly interesting. Revealed preference also carries with it a libertarian philosophy that, while usually reasonable, is in some ways quite limiting. On observing choice data, we cannot always take it as given that the decisionmaker is choosing the option they genuinely most prefer and not making a mistake. This is most obvious in cases where the decisionmaker suffers a cognitive impairment, such as dementia.

As a rule, we will take the preference relation approach as the backbone of these notes.

1.1 Preferences with transitive indifference

We start by representing the *preferences* of a decisionmaker with some binary relation over the set of alternatives X. This will be not just any binary relation, but a binary relation endowed with allegedly reasonable properties. Definitions of the basic properties of binary relations are given in the mathematical appendix to the notes.

A standard approach is to define preferences so that indifference is transitive. We will take this as a default assumption, but it is a questionable choice, as Example 2 illustrates.

Definition 1 (Weak order)**.** A *weak order* $\succ$ on a set of alternatives X is an asymmetric and negatively transitive binary relation on X.

We will derive some more properties of $\succ$. First, note negative transitivity can be rephrased in a more intuitive way:

Lemma 1. *A binary relation R on X is negatively transitive iff for all $x, y, z \in X$, we have that xRz implies xRy or yRz.*

Proof. Proof is by contraposition. Suppose $x\neg Ry$ and $y\neg Rz$. Then negative transitivity implies $x\neg Rz$. Hence xRz implies xRy or yRz. □

Moreover:

Proposition 1. *If $\succ$ is a weak order then it is irreflexive, transitive and acyclic.*

Proof. For any $x \in X$, since $\succ$ is asymmetric, $x \succ x$ would imply $x \not\succ x$, yielding a contradiction, so $\succ$ is irreflexive.

Fix $x, y, z \in X$ and suppose $x \succ y$ and $y \succ z$. Then Lemma 1 implies $x \succ z$ or $z \succ y$, and the latter is contradicted by acyclicality. Hence $x \succ z$, so $\succ$ is transitive.

Take $x_1, \ldots, x_n \in X$ and suppose $x_k \succ x_{k+1}$ for all $k = 1, \ldots, n-1$. By transitivity, $x_1 \succ x_n$. By irreflexivity, it follows that $x_1 \neq x_n$. Hence $\succ$ is acyclic. □

Whether these properties are reasonable is debatable. For individual *strict* preferences, asymmetry is obviously reasonable, since you cannot simultaneously strictly prefer x to y and y to x, and you cannot strictly prefer something to itself.

Transitivity is superficially reasonable, in that otherwise preferences might be incoherent: it is odd to imagine strictly preferring a banana to an apple and an apple to an orange, yet strictly preferring the orange to the banana. However, there are situations where transitivity might break down. One is in settings involving aggregation, most commonly encountered in social choice theory but easily adapted:

Example 1 (Condorcet cycle)**.** Suppose you are a very introspective decisionmaker and whenever it comes to ranking the alternatives $X = \{a, b, c\}$, you carefully draw up a list of factors $F = \{1, 2, 3\}$ and construct a strict ranking $\succ_k$ for each factor k. You weigh up the factors equally and strictly prefer an alternative x over y if $x \succ_k y$ for at least two factors k, i.e. a majority. Suppose you come to the following ranking:

$$\begin{array}{ccc} \succ_1 & \succ_2 & \succ_3 \\ a & c & b \\ b & a & c \\ c & b & a \end{array}$$

Then $a \succ b$, $b \succ c$ but $a \not\succ c$, so your strict preference relation $\succ$ is not transitive.

Negative transitivity is even less obvious. For example, suppose $X = (0, \infty) \times (0, \infty)$ where $(x_1, x_2) \in X$ is a bundle of x_1 cookies and x_2 chocolate bars. We might reasonably expect a decisionmaker to have $(5, 5) \succ (4, 4)$ (the decisionmaker prefers more to less). Negative transitivity would then imply that either $(5, 5) \succ (8, 2)$ or $(8, 2) \succ (4, 4)$. But these comparisons are potentially much harder to make – I know I prefer more of everything to less without necessarily taking a stance on how I might make tradeoffs between the two goods.

A weak order $\succ$ need not be connected: there may be alternatives $x, y \in X$ for which a decisionmaker does not strictly prefer x to y or y to x. There are two possibilities in this case. The primary possibility is that the decisionmaker is indifferent between x and y, which we denote by $x \sim y$. Indeed, for now we consider only this possibility.

Definition 2 (Indifference)**.** Given a strict preference relation $\succ$ on X, we define the indifference relation $\sim$ on X so that $x \sim y$ if $x \not\succ y$ and $y \not\succ x$.

The *weak preference relation* $\succsim$ on X is then defined by $x \succsim y$ if $y \not\succ x$, i.e. if $x \succ y$ or $x \sim y$. We say such a weak preference relation is *induced by* $\succ$.

Indifference has a few possible sources. A decisionmaker might be indifferent between two alternatives x and y if she does not consider there to be any meaningful difference between the two – this is "true indifference" in the sense we mean it in everyday use. Or perhaps she is uncertain over her preferences, and so finds it difficult to make a judgment call, or is otherwise unable/unwilling to compare the two. These latter reasons do not quite map as well to how we use "indifference" in everyday usage. Really these reasons point us to a second possibility other than indifference: incomparability. In the case of "true indifference", the decisionmaker does not care whether x or y is chosen, whereas if she is uncertain about her preferences or unable to make a comparison, she might care quite a bit and be unwilling or unable to decide. For example, suppose you apply for a bunch of PhD programmes without knowing that much about them, and you receive offers from Stanford and Princeton. Naturally, you can only accept one of the offers, and which offer you accept is very consequential – it will determine the next half decade of your life and shape your subsequent career in profound ways. If you do not already have a strict preference over Stanford or Princeton, then it is implausible that you really are indifferent between them – rather, you would be unwilling to make a decision until you have collected more information and formed your preferences.

Putting such critiques to the side for now, we'll proceed under the standard framework.

Proposition 2. *Suppose $\succ$ is a weak order on X (that is, $\succ$ is asymmetric and negatively transitive), and let $\succsim$ be the weak preference relation on X induced by $\succ$. Then*

(i) *for all $x, y \in X$, precisely one of $x \succ y$ or $y \succ x$ or $x \sim y$ holds;*

(ii) *$\sim$ is an equivalence relation, that is, it is symmetric, reflexive and transitive;*

(iii) *$x \succ y$ and $y \sim z$ implies $x \succ z$, and $x \sim y$ and $y \succ z$ implies $x \succ z$;*

(iv) *$\succsim$ is transitive and connected;*

(v) *if we define $\succ'$ on $X/_\sim$ (the set of equivalence classes on X under $\sim$) so that $a \succ' b$ iff $x \succ y$ for some $x \in [a]$ and $y \in [b]$, then $\succ'$ is a weakly connected weak order on $X/_\sim$.*

Proof. (i) By asymmetry, if $x \succ y$ then $y \not\succ x$, and by definition $x \not\sim y$. The case if $y \succ x$ is identical with x, y switched. If $x \not\succ y$ and $y \not\succ x$, then $x \sim y$ by definition.

(ii) If $x \not\succ y$ and $y \not\succ x$, then by definition $x \sim y$ and $y \sim x$, so $\sim$ is symmetric. Since $x \not\succ x$ by irreflexivity of $\succ$, we have $x \sim x$, so $\sim$ is reflexive. Finally, if $x \sim y$ and $y \sim z$ then (a) $x \not\succ y$ and $y \not\succ z$, so negative transitivity of $\succ$ implies $x \not\succ z$ and (b) $z \not\succ y$ and $y \not\succ x$, so negative transitivity implies $z \not\succ x$. Hence $x \sim z$, establishing that $\sim$ is transitive.

(iii) Suppose $x \succ y$, $y \sim z$. Towards contradiction, suppose $x \not\succ z$. Then either $x \sim z$ or $z \succ x$. If $x \sim z$, then transitivity of $\sim$ implies $x \sim y$, contradicting $x \succ y$. If $z \succ x$, then transitivity of $\succ$ implies $z \succ y$, contradicting $y \sim z$. The case where $x \sim y$, $y \succ z$ is similar.

(iv) Take any $x, y, z \in X$ and suppose $x \succsim y$ and $y \succsim z$. Either: (a) $x \succ y$, $y \succ z$, which implies $x \succ z$ by transitivity of $\succ$; or (b) $x \sim y$ and $y \succ z$, or $x \succ y$ and $y \sim z$, both of which imply $x \succ z$ by (iii); (c) $x \sim y$ and $y \sim z$, which imply $x \sim z$ by transitivity of $\sim$. Hence in all cases, $x \succsim z$, so $\succsim$ is transitive. Connectedness is immediate from (i).

(v) First suppose $a \succ' b$ and $b \succ' a$ for $a, b \in X/_\succ$. Then $x \succ y$ and $y' \succ x'$ for $x, x' \in [a]$ and $y, y' \in [b]$. But (iii) then implies $x' \succ y$ and so $x' \succ y'$, yielding a contradiction. Hence $\succ'$ is asymmetric.

Next suppose $a \succ' b$ for $a, b \in X/_\sim$, and $x \succ y$ for $x \in [a]$ and $y \in [b]$. Then for any $c \in X/_\sim$ and any $z \in [c]$, negative transitivity of $\succ$ implies $x \succ z$ or $z \succ y$

by Lemma 1. Hence $a \succ' c$ or $c \succ' b$, so $\succ'$ is negatively transitive by the same lemma.

Finally, suppose $a \neq b$ for $a, b \in X/_\sim$. Then $[a]$ and $[b]$ are disjoint subsets of X, so $x \in [a]$ and $y \in [b]$ implies $x \not\sim y$, so by (i), either $x \succ y$ or $y \succ x$. Hence $a \succ' b$ or $b \succ' a$, so $\succ'$ is weakly connected.

□

Transitivity of the indifference relation $\sim$ is also not necessarily reasonable in many situations. This is easiest to see in a variation on the *sorites paradox*:[1]

Example 2 (Luce, 1956)**.** Suppose you take sugar in your coffee, but you do not have a particularly sweet tooth. In particular, you would prefer a coffee with one sugar cube over a coffee with five sugar cubes mixed in. Imagine that there are 401 cups of coffee and the kth coffee prepared with $(1 + 400/k)x$ grams of sugar (where x is the mass of one sugar cube). Imagine the cups are not in order, so the only way you can assess how much sugar is in the cup of coffee is by taste. Now, you cannot discern a difference between the kth cup and the $(k+1)$th cup of coffee, so will be indifferent between them, at least in the sense that if you are prepared to choose the kth cup then you are prepared to choose the $(k+1)$th cup and vice versa.[2] However, it is obvious that you will not be indifferent between the first and last cup of coffee, since you can detect that the last is much sweeter than the first.

So far, we have taken the strict preference relation $\succ$ as basic and derived the indifference relation $\sim$ and weak preference relation $\succsim$ from $\succ$. This leads to a weak preference relation that has strong properties like connectedness. This limits the interpretation of $\succsim$: it is something *we* derived as analysts, rather than something belonging to the decisionmaker. As noted, we cannot really interpret $\sim$ as indifference in the sense that we might usually mean in everyday use under this approach.

We could instead take $\succsim$ as basic. This has the advantage that $\succsim$ need not be connected – so we can draw a distinction between indifference and incomparability. In this case, the indifference relation $\sim$ induced by $\succsim$ is defined so that $x \sim y$ if both $x \succsim y$ and $y \succsim x$, and the relation $\succ$ induced by $\succsim$ is defined so that $x \succ y$ if $x \succsim y$ and $y \not\succsim x$.

However, while in theory $\succsim$ allows us to distinguish between indifference and incomparability, we almost always assume $\succsim$ is connected, so every element of X can be compared. If $\succsim$ is a weak preference relation – i.e. if it is connected and transitive, then whether we take strict or weak preference relations as basic is immaterial:

[1] The *sorites* paradox, due to the ancient Greek philosopher Eubulides of Miletus, is as follows: a grain of sand is not a heap, and adding a grain of sand to a non-heap of sand does not turn it into a heap. Yet 10,000 grains of sand collected together is a heap, so if we kept adding sand one grain at a time, at some point we must have gone from not having a heap to having a heap. The paradox, as with Luce's sugar example, is because we cannot discern the difference between k and $k+1$ grains of sand but we can discern between 1 and 10,000 grains.

[2] Of course, even if you can't discern the difference, you might not be indifferent in the sense that if you knew the precise sugar content of the two cups, you would strictly prefer the lower sugar cup on e.g. health grounds. This will not be reflected in your choices, but it may be important when it comes to your welfare.

Proposition 3. *Suppose $\succsim$ is a connected, transitive and reflexive binary relation on X. Define $\succ$ so that $x \succ y$ iff $x \succsim y$ and $y \not\succsim x$, and define $\sim$ so that $x \sim y$ if $x \succsim y$ and $y \succsim x$. Then $\succ$ is a weak order, i.e. it is asymmetric and negatively transitive.*

Moreover, let $\sim'$ and $\succsim'$ be the indifference and weak preference relations induced by $\succ$. Then $\sim'$ and $\sim$ agree, and $\succsim$ and $\succsim'$ agree.

Proof. First, we show $\succ$ is a weak order. Take $x, y \in X$ and suppose $x \succ y$. Then $x \succsim y$ but $y \not\succsim x$, so $y \not\succ x$. Hence $\succ$ is asymmetric. Now take $x, y, z \in X$ and suppose $x \not\succ y$ and $y \not\succ z$. Since $\succsim$ is connected and reflexive, $y \succsim x$ and $z \succsim y$. Since $\succsim$ is transitive, $z \succsim x$. Hence $x \not\succ z$, so $\succ$ is negatively transitive.

Next, take $x, y \in X$. If $x \sim y$ then $x \not\succ y$ and $y \not\succ x$, so $x \sim' y$. Likewise, if $x \not\sim y$, then precisely one of $x \succsim y$ and $y \succsim x$ holds, so either $x \succ y$ or $y \succ x$, and so $x \not\sim' y$.

Again taking $x, y \in X$, if $x \succsim y$, then $y \not\succ x$, and so $x \succsim' y$. The case of $y \succsim x$ is analogous. □

1.1.1 Ordinal utility representations

If we were restricted to analyse preferences purely in terms of binary relations, we would not get very far. It is far more convenient when preferences can be summarized by some order-preserving function (a *utility function*) that maps the set of alternatives X into the real numbers.

Definition 3 (Utility function). Given a weak order $\succ$ on a set X, a function $u : X \to \mathbb{R}$ is called a *utility function representing* $\succ$ if

$$x \succ y \quad \text{iff} \quad u(x) > u(y)$$

for all $x, y \in X$.

Fortunately, if $\succ$ is a weak order and $X/_\sim$ is countable, we can always represent preferences by a utility function:

Theorem 1 (Birkhoff, 1948; Suppes & Zinnes, 1963). *Suppose $\succ$ is a binary relation on set of alternatives X and suppose $X/_\sim$ is countable. Then $\succ$ is a weak order on X iff there exists a utility function $u : X \to \mathbb{R}$ representing $\succ$.*

Proof. Suppose $\succ$ is a weak order on X. Let $\{a_0, a_1, \dots\}$ be an enumeration of $X/_\sim$ and let $\{q_0, q_1, \dots\}$ be an enumeration of the rationals $\mathbb{Q}$. Let $\succ'$ be the weak order on $X/_\sim$ induced by $\succ$.

Set $u(a_0) = 0$. Suppose we have defined u for all x_i, $i < k$. Then for a_k, precisely one of the following cases applies:

(i) $a_k \succ' a_i$ for all $i < k$, in which case take $u(a_k) = k$, or

(ii) $a_i \succ' a_k$ for all $i < k$, in which case take $u(a_k) = -k$, or

(iii) $a_i \succ' a_k \succ' a_j$ for some $i, j < k$ such that there is no $m < k$ with $a_i \succ' a_m \succ' a_j$. In this case, take $u(a_k) = q_r$ for the first q_r in the enumeration $\{q_0, q_1, \dots\}$ such that $u(a_i) > q_r > u(a_j)$ (since $\mathbb{Q}$ is dense in $\mathbb{R}$, such a rational number exists).

Now by construction, $u(a_k) \neq u(a_i)$ for all $i < k$ and $a_i \succ' a_j$ iff $u(a_i) > u(a_j)$ for all $i, j \leq k$. Since this holds for all k, it holds for all $a_k \in X/_\sim$. Now, for every $a \in X/_\sim$, define $u(x) = u(a)$ for all $x \in [a]$. Then u is defined on X and $x \succ y$ iff $u(x) > u(y)$ given Proposition 2(v).

Conversely, suppose $\succ$ is a binary relation on X and there is a function $u : X \to \mathbb{R}$ such that $x \succ y$ iff $u(x) > u(y)$. Take $x, y, z \in X$. Since $>$ is asymmetric and negatively transitive, we immediately have that $\succ$ is asymmetric and negatively transitive. □

From Theorem 1, given a weak order $\succ$, we also have that $x \succsim y$ iff $u(x) \geq u(y)$ and $x \sim y$ iff $u(x) = u(y)$.

In general, the result does not extend to uncountable sets of alternatives, as the following counterexample shows.

Definition 4 (Lexicographic order)**.** Given a product set $X = \bigtimes_{i=1}^{k} X_i$, a partial order $>$ on X is a (strict) *lexicographic order* if it takes the form $x > y$ if $x_1 > y_1$ or if there is some $j \leq k$ such that $x_i = y_i$ for all $i < j$ and $x_j > y_j$.

Note the lexicographic order is asymmetric and negatively transitive. Consider $x, y \in X$ and take $>$ to be the lexicographic order on X. Write $x \geq y$ if $x > y$ or $x = y$. For $x, y \in X$, if $x > y$, then $x_j > y_j$ for some j and $x_i = y_i$ for all i such that $i < j$ (if any). Given $x_i = y_i$ for all such i, $y > x$ would imply $y_j > x_j$, yielding a contradiction. Hence $>$ is asymmetric. For negative transitivity, take $x, y, z \in X$. If $x \not> y$ and $y \not> z$, then (i) $x = y = z$ or (ii) $y > x$ and $z \geq y$ or (iii) $z > y$ and $y \geq x$. If (i) holds then $x \not> z$ is immediate. If (ii) holds then there is some j such that $y_i = x_i$ for all $i < j$ if any and $y_j > x_j$, and since $z \geq y$, $z_i \geq y_i$ for all $i \leq j$. Hence $z_i \geq x_i$ for all $i \leq j$ with strict inequality for some such i, so $z > x$, implying $x \not> z$. The case of (iii) is similar. Hence $>$ is negatively transitive.

Example 3 (Lexicographic preferences)**.** Let $X = [0,1] \times [0,1]$ and let $>$ be the lexicographic order on X. Suppose $x \succ y$ iff $x > y$ in the lexicographic order, i.e. $x \succ y$ iff either $x_1 > y_1$ or both $x_1 = y_1$ and $x_2 > y_2$. Since the lexicographic order is asymmetric and negatively transitive, $\succ$ is a weak order.

Suppose there is some utility function $u : X \to \mathbb{R}$ that represents $\succ$. Then for all $x_1 \in [0,1]$, we have that $(x_1, 1) \succ (x_1, 0)$, so $u(x_1, 1) > u(x_1, 0)$. Let $\rho(x_1) = u(x_1, 1) - u(x_1, 0)$. Then $\rho(x_1) > 0$ for all $x_1 \in [0,1]$, and so we have

$$[0,1] = \bigcup_{n=1}^{\infty} \left\{ x_1 \in [0,1] \mid \rho(x_1) > \frac{1}{n} \right\}.$$

Since the left hand side is uncountable, the right hand side is a countable union, and the countable union of countable sets is countable, it follows that for some n, say $\bar{n}$, $E_{\bar{n}} := \{x_1 \in [0,1] \mid \rho(x_1) > 1/\bar{n}\}$ is uncountable.

Now let $K = u(1,1) - u(0,0)$ and choose $N \in \mathbb{N}$ so that $N > K\bar{n}+1$.[3] Choose a subset $E = \{y_1, \dots, y_N\} \subset E_{\bar{n}}$ with precisely N elements, indexed so that $y_1 < y_2 < \cdots < y_N$. Now since $(y_n, 0) \succ (y_{n-1}, 1)$, we have $u(y_n, 0) > u(y_{n-1}, 1)$ for all $n = 2, \dots, N$. Thus

$$u(y_n, 0) - u(y_{n-1}, 0) > u(y_{n-1}, 1) - u(y_{n-1}, 0) > \frac{1}{\bar{n}}.$$

Now we have

$$\begin{aligned} K &= u(1,1) - u(0,0) \\ &= [u(1,1) - u(y_N, 0)] + [u(y_1, 0) - u(0,0)] + \sum_{n=2}^{N} [u(y_n, 0) - u(y_{n-1}, 0)] \\ &> \sum_{n=2}^{N} \frac{1}{\bar{n}} > \frac{N-1}{\bar{n}} > K, \end{aligned}$$

yielding a contradiction. Hence there is no utility function representing $\succ$.

To make progress, then, we need to impose further conditions on the preference relation $\succ$:

Definition 5 (Order density)**.** Let R be a binary relation on a set X. We say that a subset $X_0 \subseteq X$ is *R-order dense in X* if for every $x, y \in X$: xRy and $x, y \notin X_0$ implies there is some $z \in X_0$ such that xRz and zRy.

This is not a very easily interpreted property, but it fails in situations like Example 3.

Theorem 2 (Birkhoff, 1948; Luce & Suppes, 1965)**.** *Let $\succ$ be a binary relation on X. Then there is a utility function representing $\succ$ iff $\succ$ is a weak order on X and there is a countable subset $Z \subset X/_\sim$ such that Z is $\succ'$-order dense in $X/_\sim$, where $\succ'$ is the weak order induced on $X/_\sim$ by $\succ$.*

Proof. Suppose for all $x, y \in X$, $x \succ y$ iff $u(x) > u(y)$. Then $\succ$ is a weak order on X and $\succ'$ is a weakly connected weak order on $X/_\sim$ such that $a \succ' b$ iff $u(a) > u(b)$ with $u(a) = u(x)$ whenever $x \in [a]$. Let $\mathcal{I} = \{[a, b] \mid a < b \text{ and } a, b \in \mathbb{Q}\}$. Now, for each interval $I \in \mathcal{I}$ for which there is some $a \in X/_\sim$ such that $u(a) \in I$, select one such a and put $a \in A$. This generates a countable set A. Now define

$$K = \{(b, c) \mid b, c \in X/_\sim - A,\ \ b \succ' c \text{ and } b \succ' a \succ' c \text{for no } a \in A\}.$$

Then if $(b, c) \in K$, we have $b \succ' a \succ' c$ for no $a \in X/_\sim$ since otherwise there exists some $d \in A$ s.t. $b \succ' d \succ' c$ and for all $x \in (u(c), u(b))$, there is an interval $I \in \mathcal{I}$ such that $x \in I \subset (u(c), u(b))$. Hence $(u(c), u(b))$ and $(u(c'), u(b'))$ are disjoint for all distinct $(b, c), (b', c') \in K$, and so K is countable. It follows that

$$B := \{b \in X/_\sim \mid \text{there exists } c \in X/_\sim \text{ s.t. } (b, c) \in K \text{ or } (c, b) \in K\}$$

[3] My convention is to assume $0 \in \mathbb{N}$. I use $\mathbb{N}_+$ to denote the positive integers.

is countable. Thus $B \cup C$ is countable. Now, if $b, c \in X/_\sim - (A \cup B)$ then there exists $a \in A \cup B$ such that $b \succ' a \succ' c$, and so $A \cup B$ is $\succ'$-order dense in $X/_\sim$.

Conversely, suppose $\succ$ is a weak order on X and so $\succ'$ is a weakly connected weak order on $X/_\sim$.

For $E \subseteq X/_\sim$, we define $\sup_{\succ'} E$ and $\inf_{\succ'} E$ to be the most and least preferred elements of E under $\succ'$.

Assume A is countable and $\succ'$-order dense in $X/_\sim$, and that A contains $\sup_{\succ'} X/_\sim$ and $\inf_{\succ'} X/_\sim$, if they exist. Define

$$B := \{b \in X/_\sim - A \mid \sup\{a \in A \mid b \succ' a\} \text{ exists, or } \inf\{a \in A \mid a \succ' b\} \text{ exists}\}.$$

For each $b \in B$, define $\bar{A}_b = \sup\{a \in A \mid b \succ' a\}$ and $\underline{A}_b = \inf\{a \in A \mid a \succ' b\}$, and define $\bar{a}_b = \inf \bar{A}_b$ and $\underline{a}^b = \sup \underline{A}_b$. Note $\bar{A}_b$ and $\underline{A}_b$ are disjoint for each $b \in B$, with $\bar{A}_b \cup \underline{A}_b = A$. Hence for each $a \in A$, $a = \bar{a}_b$ for at most one $b \in X/_\sim - A$ and similarly, $a = \underline{a}_b$ for at most one $b \in X/_\sim - A$. Thus B is countable, and so $C := A \cup B$ is countable.

Lemma 2. *Neither*

(i) $\sup_{\succ'}\{c \in C \mid b \succ' c\}$*, nor*

(ii) $\inf_{\succ'}\{c \in C \mid c \succ' b\}$

exist for any $b \in X/_\sim - C$*.*

Proof. Suppose (i) does not hold. Then there is some $\underline{c}^b = \sup\{c \in C \mid b \succ' c\}$ for some $b \in X/_\sim - C$. If $\underline{c}^b \in A$ then $b \in B$, yielding a contradiction, so $\underline{c}^b \notin A$. But it follows that $a \succ' b \succ' \underline{c}^b \succ' c$ for all $c \in \{c \in A \mid c \succ' b\}$ and $a \in \{a \in A \mid b \succ' a\}$. Yet there is no $b' \in A$ such that $b \succ' b' \succ' \underline{c}^b$, contradicting $\succ'$-order density. Hence (i) holds. The proof for (ii) is symmetric. □

Now by the proof of Theorem 1, there is a function $u : C \to \mathbb{R}$ such that $a \succ' c$ iff $u(a) > u(c)$ for all $a, c \in C$. Now for each $b \in X/_\sim - C$, put

$$U^b := \{u(c) \mid c \in C \text{ and } c \succ' b\},$$
$$U_b := \{u(c) \mid c \in C \text{ and } b \succ' c\},$$

and define $u(b) = \frac{1}{2}(\sup U_b + \inf U^b)$.

Since $u(a) < u(c)$ for all $a \in U_b$ and $c \in U^b$, we have $\sup U_b \leq \inf U^a$. By Lemma 2, for each $b \in X/_\sim - C$, we have $u(a) < \sup U_b$ for all $u(a) \in U_b$ and $u(c) > \inf U^b$ for all $u(c) \in U^b$, and thus $u(a) < u(b) < u(c)$ for all $a \in \{c \in C \mid b \succ' c\}$ and all $c \in \{c \in C \mid c \succ' b\}$. Thus the ordering of $b \in X/_\sim - C$ and $c \in C$ is preserved by this extension of u.

Now suppose $a, b \in X/_\sim - C$. If $a \succ' b$ then $a \succ' c \succ' b$ for some $c \in C$, so $u(c) > u(b)$, $u(a) > u(c)$, and thus $u(a) > u(b)$. On the other hand, if $u(b) > u(a)$ then there is some $u(c) \in U^b$ such that $u(a) > u(c) > u(b)$, giving $c \succ' b$ and $a \succ' c$, and, by transitivity, $a \succ' b$. Thus $a \succ' b$ iff $u(a) > u(b)$ for all $a, b \in X/_\sim$. Proof now follows by defining $u : X \to \mathbb{R}$ by $u(x) = u(a)$ for all $x \in [a]$ and all $a \in X/_\sim$. □

Given a utility function $u : X \to \mathbb{R}$ representing a preference relation $\succ$, the numbers $u(x)$ and $u(y)$ themselves carry no interpretation whatsoever. They certainly do not correspond to "utils" in the sense that a utilitarian might mean. This is obvious when we observe that any order-preserving transformation of u is also a utility function representing $\succ$:

Proposition 4. *If $\succ$ is a weak order on X admitting a utility representation $u : X \to \mathbb{R}$, and if $f : \mathbb{R} \to \mathbb{R}$ is an order-preserving function (that is, f is strictly increasing), then $f \circ u$ is a utility function representing $\succ$.*

Conversely, if $u : X \to \mathbb{R}$ and $v : X \to \mathbb{R}$ are both utility functions representing the weak order $\succ$, then there is an order preserving function $f : \mathbb{R} \to \mathbb{R}$ such that $v = f \circ u$.

Proof. Let $\succ$ be a weak order on X. Then $x \succ y$ iff $u(x) > u(y)$. If f is order-preserving, then $f(u(x)) > f(u(y))$ iff $u(x) > u(y)$, and so $f(u(x)) > f(u(y))$ iff $x \succ y$. Thus $f \circ u$ represents $\succ$.

Now suppose u and v represent $\succ$. Define $g : u(X) \to \mathbb{R}$ to be some function such that $v(x) = g(u(x))$ for each $x \in X$. Since the following are equivalent: (i) $x \succ y$, (ii) $u(x) > u(y)$ and (iii) $v(x) > v(y)$, and since $g(u(x)) = v(x)$ for all $x \in X$, we have that for any $a, b \in u(X)$, if $a > b$ then $g(a) > g(b)$. Now, let $f : \mathbb{R} \to \mathbb{R}$ be such that $f(a) = g(a)$ for each $a \in u(X)$. Define $\bar{a} = \sup u(X)$ and $\underline{a} = \inf u(X)$, which are guaranteed to exist in the extended reals. For $a \geq \bar{a}$, let $f(a) = g(a) + a - \bar{a}$, and for $a \leq \underline{a}$, let $f(a) = g(a) + a - \underline{a}$. Now, fix any $a \in (\underline{a}, \bar{a}) - u(X)$ and define $b_a = \sup\{b < a \mid b \in u(X)\}$, $g_a = \sup\{g(b) \mid b < a, b \in u(X)\}$, $b^a = \inf\{b > a \mid b \in u(X)\}$ and $g^a = \inf\{g(b) \mid b > a, b \in u(X)\}$. Since g is strictly increasing on $u(X)$, $g^a > g_a$. Now let $f(a) = g_a + \frac{g^a - g_a}{b^a - b_a}(a - b_a)$. Then f is an order preserving function s.t. $v = f \circ u$. $\square$

Proposition 4 implies that utility representations of preference relations, where they exist, are unique up to order-preserving transformations.

1.1.2 Preferences on $\mathbb{R}^n$

In many situations, preferences depend on many different factors, and so it is convenient to consider the set of alternatives to be some subset of $\mathbb{R}^n$ (or even $\mathbb{R}^{\mathbb{N}}$ or $\mathbb{R}^I$ for some real interval I):

Example 4.

(a) *Bundles of goods.* Consumers have preferences over bundles $x = (x_1, \dots, x_n)$ of n different goods, with x_k being the quantity of good k in the bundle.

(b) *Dynamic choice.* In settings where outcomes take place in discrete time, an alternative $x = (x_0, x_1, \dots)$ lists outcomes x_t for each period t.

(c) *Multidimensional management problems.* In managerial settings, a decisionmaker might care about multiple performance metrics. For example, imagine the chief executive officer of a company who cares about both her firm's profits and its market share. Then the set of alternatives is $X = \mathbb{R} \times [0,1]$, each alternative $x = (x_1, x_2)$ consisting of a value for profit $x_1 \in \mathbb{R}$ and for market share $x_2 \in [0,1]$.

In these settings, if preferences can be represented by a utility function, then the utility function defines a *utility surface* over the n-dimensional plane. The equivalence classes in $X/_\sim$ in this case are *indifference curves.*

We will restrict to rectangular subsets of $\mathbb{R}^n$, i.e. subsets $X = X_1 \times \cdots \times X_n$ such that each X_i is an interval (possibly unbounded). Define the strict partial order $>$ on X so that $x > y$ if $x_i \geq y_i$ for all $i = 1, \ldots, n$ and $x \neq y$.

Theorem 3. *Let $X \subseteq \mathbb{R}^n$ be a rectangular subset. Suppose*

(i) *$\succ$ is a weak order on X;*

(ii) *$x > y$ implies $x \succ y$, and*

(iii) *if $x \succ y$ and $y \succ z$ then there are $\alpha, \beta \in (0,1)$ s.t. $\alpha x + (1-\alpha)z \succ y$ and $y \succ \beta x + (1-\beta)z$.*

Then there is a function $u : X \to \mathbb{R}$ such that

$$x \succ y \quad \textit{iff} \quad u(x) > u(y) \quad \textit{for all } x, y \in X.$$

Proof. First we prove a useful lemma:

Lemma 3. *Under the hypotheses of Theorem 3, if $x, y, z \in X$ and $x > y > z$ then there is a unique $\alpha \in (0,1)$ such that $y \sim \alpha x + (1-\alpha)z$.*

Proof. First suppose there is no $\alpha \in (0,1)$ s.t. $y \sim \alpha x+(1-\alpha)z$. Then by the hypotheses, there is some $\beta \in (0,1)$ for which

(a) $y \succ \alpha x + (1-\alpha)z$ for all $\alpha < \beta$;

(b) $\alpha x + (1-\alpha)z \succ y$ for all $\alpha > \beta$;

(c) either $y \succ \beta x + (1-\beta)z$ or $\beta x + (1-\beta)z \succ y$,

Suppose in (c) that the latter applies. Then by (ii), $\beta x + (1-\beta)z \succ y \succ z$, and by (iii), there is some $\alpha \in (0,1)$ s.t. $\alpha[\beta x + (1-\beta)z] + (1-\alpha)z = \alpha\beta x + (1-\alpha\beta)z \succ y$. Now since $\alpha\beta < \alpha$, by (ii) we have $y \succ \alpha\beta x + (1-\alpha\beta)z$, yielding a contradiction. By a symmetric argument, if in (c) the former applies, we also have a contradiction.

Therefore $y \sim \alpha x+(1-\alpha)z$ for some $\alpha \in (0,1)$. Now suppose that for $\alpha_1, \alpha_2 \in (0,1)$, we have $y \sim \alpha_1 x + (1-\alpha_1)z$ and $y \sim \alpha_2 x + (1-\alpha_2)z$. Then transitivity of $\sim$ implies $\alpha_1 x+(1-\alpha_1)z \sim \alpha_2 x+(1-\alpha_2)z$. Since $x > z$, this holds only if $\alpha_1 = \alpha_2$, for if $\alpha_1 \neq \alpha_2$ then either $\alpha_1 x + (1-\alpha_1)z > \alpha_2 x + (1-\alpha_2)z$ or $\alpha_1 x + (1-\alpha_1)z < \alpha_2 x + (1-\alpha_2)z$ which by (ii) imply a contradiction. $\square$

We mean to apply Theorem 2, so must show there is a countable $\succ'$-order dense subset $Z \subset X/_\sim$.

For each interval X_i, define a_i, b_i as the lower and upper endpoints of X_i. For each i, define $Y_i = \mathbb{Q} \cup (\{a_i, b_i\} \cap \mathbb{R})$, i.e. Y_i is the set of rational numbers together with the endpoints of X_i if those endpoints are finite. Then Y_i is countable and so defining $Z_i = X_i \cap Y_i$, we have that Z_i is countable. Letting $W_i = \{\alpha x_i + (1-\alpha) y_i \mid \alpha \in [0,1] \cap \mathbb{Q} \text{ and } x_i, y_i \in Z_i\}$, we have that W_i is countable, and so if we define $W = \bigtimes_{i=1}^n W_i$, then W is countable. Define A to be the set of elements of $X/_\sim$ that contain at least one point in W. Then A is countable given for each $x \in W$, there is exactly one $a \in X/_\sim$ s.t. $x \in [a]$. Take $a, b \in X/_\sim - A$ s.t. $a \succ' b$. We mean to show there is some $c \in A$ for which $a \succ' c \succ' b$. This is equivalent to showing that for any $x, y \in X - W$ with $x \succ y$, there is some $z \in W$ for which $x \succ z \succ y$. We divide into two cases:

(a) $x > y$. Then we have some $z, z' \in Z := \bigtimes_{i=1}^n Z_i$ s.t. $z > x$ and $y > z'$. From Lemma 3 and (ii), we have that there are $\alpha, \beta \in (0,1)$ with $\alpha < \beta$ s.t. $x \sim \alpha z + (1-\alpha) z'$ and $y \sim \beta z + (1-\beta) z'$. Take any $\gamma \in (\alpha, \beta) \cap \mathbb{Q}$. By (ii), $x \succ \gamma z + (1-\gamma) z' \succ y$, and given $z, z' \in Z$ and γ is rational, $\gamma z + (1-\gamma) z' \in W$.

(b) $x \not> y$. Define $v_i = x_i \wedge y_i$ and $w_i = x_i \vee y_i$.[4] Then $v < x < w$ and $v < y < w$. Hence by (iii), there are $\alpha, \beta \in (0,1)$ with $\alpha < \beta$ s.t. $x \sim \alpha v + (1-\alpha) w$ and $y \sim \beta v + (1-\beta) w$. Now, $\alpha v + (1-\alpha) w > \beta v + (1-\beta) w$ so by case (a), there is a $z \in W$ s.t. $\alpha v + (1-\alpha) w \succ z \succ \beta v + (1-\beta) w$, i.e. s.t. $x \succ z \succ y$.

□

1.1.3 Continuity

Frequently, we want that the strict preference relation $\succ$ and the utility function representing $\succ$ has convenient properties. An obvious convenient property is continuity – it allows us to apply Brouwer's fixed point theorem, it is a necessary condition for differentiability, and so on. For this we need a new axiom:

Axiom. Let $\succ$ be a weak order on topological space $(X, \mathcal{T})$.

(**C1**) For every $x, y \in X$ with $x \succ y$, there exist open neighbourhoods T_x of x and T_y of y such that $x' \succ y$ for all $x' \in T_x$ and $x \succ y'$ for all $y' \in T_y$.

It is common to call preferences $\succ$ *continuous* if they satisfy (**C1**). Theorem 4 justifies why:

Theorem 4. *Suppose $(X, \mathcal{T})$ is a topological space and $\succ$ is a weak order on X for which there is some utility function representation. Then there is a continuous function $u : X \to \mathbb{R}$ representing $\succ$ iff $\succ$ satisfies* (**C1**).

Proof. We claim (**C1**) is equivalent to:

[4]Note we denote the meet of x and y by $x \wedge y = \inf\{x, y\}$ and the join of x and y by $x \vee y = \sup\{x, y\}$.

(**C1′**) For every $y \in X$, both $\{x \in X \mid x \succ y\}$ and $\{x \in X \mid y \succ x\}$ are open.

It is immediate that (**C1′**) implies (**C1**), so we need only prove the converse. First consider $U_y := \{x \in X \mid x \succ y\}$. If there is no $x \in X$ for which $x \succ y$ then $U_y = \varnothing$ and so $U_y \in \mathcal{T}$. If $x \succ y$ for some $x \in X$, then there is a $T_x \in \mathcal{T}$ with $x \in T_x$ and $x' \succ y$ for all $x' \in T_x$. The family $\mathcal{T}_y = \{T \in \mathcal{T} \mid x \succ y \text{ for all } x \in T\}$ is the family of all such open neighbourhoods for all $x \succ y$. Hence $U_y = \bigcup_{T \in \mathcal{T}_y} T$, and so $U_y \in \mathcal{T}$ since $\mathcal{T}$ is closed under arbitrary union. A symmetric argument shows that $L_y := \{x \in X \mid y \succ x\}$ lies in the topology $\mathcal{T}$.

Next, we mean to show (**C1′**) implies there is some continuous $u : X \to \mathbb{R}$ representing $\succ$. By the hypotheses of the theorem, there is some $v : X \to \mathbb{R}$ representing $\succ$. Call a nonempty interval I of $\mathbb{R}$ a *gap* of $v(X)$ (the range of v) if $I \cap v(X) = \varnothing$ and I has form

$$I = \{b \mid v(x) < b < v(y)\}$$

for all $v(x) \in \{v(x) < a \mid x \in X\}$ and $v(y) \in \{v(y) > a \mid y \in X\}$, for some $a \in I$.

Lemma 4 (Debreu, 1964)**.** *If $v : X \to \mathbb{R}$ represents $\succ$ and $\succ$ satisfies (**C1′**), then there is a function $u : X \to \mathbb{R}$ representing $\succ$ such that all the gaps of $u(X)$ are open.*

Proof. See Debreu (1964) for a proof (it is 4 pages long). □

Now let $u : X \to \mathbb{R}$ be a function representing $\succ$ such that all the gaps of $u(X)$ are open. Fix any $a \in \mathbb{R}$. If $a = u(y)$ for some $y \in X$, then $\{x \in X \mid u(x) \in (a, \infty)\} = \{x \in X \mid x \succ y\}$, which is open by (**C1′**). If $a \notin u(X)$ but a lies in some gap of $u(X)$, then the gap has the form $(a_1, a_2) \ni a$ where $a_1, a_2 \in u(X)$, and so $\{x \in X \mid u(x) \in (a, \infty)\} = \{x \in X \mid x \succ z\}$ is open for $a_1 = u(z)$. The remaining case has $a \notin u(X)$ and a lying in no gap of $u(X)$. Then either

(i) $a \leq \inf X$. Then $\{x \in X \mid u(x) \in (a, \infty)\} = X \in \mathcal{T}$.

(ii) $a \geq \sup X$. Then $\{x \in X \mid u(x) \in (a, \infty)\} = \varnothing \in \mathcal{T}$.

(iii) $a = \inf\{u(x) > a \mid x \in X\}$. For each $y \in X$ let $U_y = \{x \in X \mid \succ y\}$. Then $\{x \in X \mid v(x) \in (a, \infty)\} = \bigcup_{y \in X : u(y) > a} U_y \in \mathcal{T}$ since it is a union of open sets.

Hence $\{x \in X \mid u(x) \in (a, \infty)\}$ is open for all $a \in \mathbb{R}$. A symmetric argument shows $\{x \in X \mid u(x) \in (-\infty, b)\} \in \mathcal{T}$ for all $b \in \mathbb{R}$. Moreover, for any $a, b \in \mathbb{R}$, $\{x \in X \mid u(x) \in (a, b)\} = \{x \in X \mid u(x) \in (a, \infty)\} \cap \{x \in X \mid u(x) \in (-\infty, b)\}$ is a finite intersection of open sets so is open. Now any open set $A \subseteq \mathbb{R}$ can be written as an arbitrary union of open intervals, so for any such set, we have that $\{x \in X \mid u(x) \in A\} \in \mathcal{T}$ since it is an arbitrary union of sets in $\mathcal{T}$. This establishes sufficiency of (**C1**). □

For separable metric spaces, we can rephrase Theorem 4 since metric spaces are associated with topological spaces induced by the given metric. In a separable metric space, we can replace (**C1**) with:

Axiom. Let $\succ$ be a weak order on a separable metric space (X, ρ).

(**C2**) For every sequence $\{x_n\} \subseteq X$ with $x = \lim_{n\to\infty} x_n$ and for every $y \in X$:

(i) if $x \succ y$ then there is an $N \in \mathbb{N}$ such that $x_n \succ y$ for all $n \geq N$, and

(ii) if $y \succ x$ then there is an $N \in \mathbb{N}$ such that $y \succ x_n$ for all $n \geq N$.

This gives us the following corollary:[5]

Corollary 1. *Suppose (X, ρ) is a separable metric space and $\succ$ is a weak order on X for which there is some utility function representation. Then there is a continuous function $u : X \to \mathbb{R}$ representing $\succ$ iff $\succ$ satisfies* (**C2**).

Proof. We need only show (**C2**) implies (**C1**) holds for the topology induced by the metric ρ. Fix any $x, y \in X$ with $x \succ y$. Under (**C2**), for every sequence $\{x_n\}$ with $x_n \to x$, there is some N such that $x_n \succ y$ for all $n \geq N$. For each such sequence, define $M_{\{x_n\}} := \max\{N \in \mathbb{N} \mid y \succsim x_n\}$ and $r_{\{x_n\}} := \rho(x_{M_{\{x_n\}}}, x)$. Define $r := \inf\{r_{\{x_n\}} \mid x_n \to x\}$. Since $\rho \geq 0$, we have $r \geq 0$. Towards contradiction, suppose $r = 0$. Then there exist sequences $\{x_n^m\} : m \in \mathbb{N}$ such that $r_{\{x_n^m\}} < \frac{1}{m}$ for each m. Now construct the sequence $\{x_m\}$ so that $x_m = x_{M_{\{x_n^m\}}}$ for each $m \in \mathbb{N}$. Then $y \succ x_m$ for all $m \in \mathbb{N}$, and since $\rho(x_m, x) < \frac{1}{m}$, we have $x_m \to x$, contradicting (**C2**). Hence $r > 0$. It follows that there is some open ϵ-ball B_x of x with $\epsilon \in (0, r)$ such that $x' \succ y$ for all $x' \in B_x$. A symmetric argument applies for sequences $\{y_n\}$ with $y_n \to y$ to show there is some open ϵ-ball B_y of y with $x \succ y'$ for all $y' \in B_y$. Hence (**C1**) holds. Applying Theorem 4 completes the proof. □

1.2 Preferences without transitive indifference

As noted in Example 2, transitive indifference is a problematic assumption in many circumstances because we cannot necessarily distinguish between sufficiently similar alternatives. If you cannot discern small quality differences, then you might be indifferent between each pair of similar objects in a row of such objects, but have a clear strict preference between the objects at either end of the row, one of which is clearly inferior to the other. This motivates relaxations of transitive indifference. Naturally, this requires we define strict preference relations in some way other than as a weak order.

1.2.1 Strict partial orders

To allow non-transitive indifference, we will instead assume that the strict preference relation $\succ$ is a strict partial order.

Definition 6 (Strict partial order)**.** A *strict partial order* $\succ$ on a set X is an irreflexive and transitive binary relation on X.[6]

[5]This corollary is the version familiar in most microeconomics textbooks, such as Kreps (1988, 2013) or Mas-Colell, Whinston & Green (1995).

[6]Note this implies a strict partial order is asymmetric, since if $x \succ y$ and $y \succ x$ then transitivity implies $x \succ x$, contradicting irreflexivity.

This is weaker than the notion of a weak order. Indeed, by Proposition 1, *any weak order is a strict partial order.*

We define the indifference relation $\sim$ analogously to Definition 2, i.e. $x \sim y$ if $x \not\succ y$ and $y \not\succ x$. To see we can now have non-transitive indifference, suppose $X = \{a, b, c\}$ and $a \succ c$ only. Then $\succ$ is irreflexive and transitive, so is a strict partial order, and we have $a \sim b$, $b \sim c$ but $a \succ c$, so $\sim$ is not transitive.

This said, we can obtain a similar result to Proposition 2 if we define the equivalence relation $\approx$ on X so that $x \approx y$ if for all $z \in X$, we have $x \sim z$ iff $y \sim z$. This implies $x \sim y$ since $\sim$ is reflexive.

Proposition 5. *Suppose $\succ$ is a strict partial order on X (that is, $\succ$ is irreflexive and transitive). Then*

(i) *for each $x, y \in X$, precisely one of*

  (a) $x \succ y$,

  (b) $y \succ x$,

  (c) $x \approx y$, *or*

  (d) $x \sim y$ *and* $x \not\approx y$

  holds;

(ii) *$\approx$ is an equivalence relation;*

(iii) *for all $x, y \in X$, $x \approx y$ iff for all $z \in X$, $x \succ z$ iff $y \succ z$ and $z \succ x$ iff $z \succ y$;*

(iv) *$x \succ y$ and $y \approx z$ implies $x \succ z$, and $x \approx y$ and $y \succ z$ implies $x \succ z$;*

(v) *if we define $\succ'$ on $X/_{\approx}$ (the set of equivalence classes on X under $\approx$) so that $a \succ' b$ iff $x \succ y$ for some $x \in [a]$ and $y \in [b]$, then $\succ'$ is a strict partial order on $X/_{\approx}$.*

Proof. (i) For any $x, y \in X$, if $x \succ y$ and $y \succ x$ then transitivity would imply $x \succ x$, contradicting irreflexivity. Thus $x \succ y$ implies $y \not\succ x$ so $\succ$ is asymmetric, and thus at most one of (a) and (b) can hold. If neither (a) nor (b) holds, then $x \sim y$ and so precisely one of (c) or (d) holds, given $x \sim y$ is necessary for $x \approx y$.

(ii) By definition, $\approx$ is reflexive and symmetric. Moreover, if $x \approx y$ and $y \approx z$, then from the definition, for any $t \sim x$ we have $y \sim t$ and thus $z \sim t$, so $x \sim t$ implies $z \sim t$. Likewise, for any $t \sim z$, we have $y \sim t$ and thus $x \sim t$ by definition of $\approx$, so $z \sim t$ implies $x \sim t$, and thus we conclude $x \approx z$, so $\approx$ is transitive.

(iii) Suppose $x \approx y$. Then if $x \succ z$, either $y \sim z$ or $y \succ z$ (if $z \succ y$ then transitivity of $\succ$ would imply $x \succ y$, yielding a contradiction); now if $y \sim z$, then the definition of $\approx$ gives us $x \sim z$, yielding a contradiction. Therefore $x \succ z$ implies $y \succ z$. That $y \succ z$ implies $x \succ z$ is symmetric, and the proof that $z \succ x$ iff $z \succ y$ is similar. This also completes the proof of (iv).

Conversely, suppose the rhs of (iii) holds. Suppose $x \sim t$. Then by definition, we do not have $x \succ t$ and so do not have $y \succ t$ and we do not have $t \succ x$ so do not have $y \succ t$, so $x \sim t$ implies $y \sim t$. Similarly, $y \sim t$ implies $x \sim t$. Thus $x \approx y$.

(iv) See proof of (iii).

(v) Take $a \in X/_{\approx}$. If $a \succ' a$ then $x \succ y$ for some $x, y \in [a]$, which satisfy $x \approx y$ by definition, yielding a contradiction. Hence $\succ'$ is irreflexive. Now fix $a, b, c \in X/_{\approx}$ and suppose $a \succ' b$ and $b \succ' c$. Then $x \succ y$, $y \approx y'$ and $y' \succ z$ for $x \in [a]$, $y, y' \in [b]$ and $z \in [c]$. By (iv), if $x \succ y$ and $y \approx y'$ then $x \succ y'$, and transitivity of $\succ$ thus implies $x \succ z$, so $a \succ' c$. Hence $\succ'$ is transitive.

□

To make progress in representing a strict partial order by a utility function, we need to make use of an extension theorem due to Szpilrajn (1930). Briefly, we need to introduce some notation. For some partial orders $\succ$ and $\succ'$ on a set X, we write $\succ \subseteq \succ'$ if $x \succ y$ implies $x \succ' y$ for all $x, y \in X$. We write $\succ \subset \succ'$ if $\succ \subseteq \succ'$ but there is also some pair $x, y \in X$ such that $x \succ' y$ and $x \not\succ y$. When $\succ \subseteq \succ'$, we say that $\succ'$ *includes* $\succ$ or is an *extension* of $\succ$.

Theorem 5 (Szpilrajn's extension theorem)**.** *If $\succ$ is a strict partial order on a set X, then there exists a strict total order $\succ^*$ on X such that $\succ^*$ is an extension of $\succ$: that is, such that*

$$x \succ y \quad \textit{implies} \quad x \succ^* y.$$

Proof. Suppose $\succ$ is a strict partial order but not a strict total order (otherwise setting $\succ^* = \succ$ gives the conclusion immediately). Then there is some pair $x, y \in X$ such that $x \neq y$, $x \not\succ y$ and $y \not\succ x$. Let $a \succeq b$ denote "$a \succ b$ or $a = b$." Define $\succ'$ on X so that for all $s, t \in X$:

(i) if $s \succ t$ then $s \succ' t$; else

(ii) if $s \succeq x$ and $y \succeq t$ then $s \succ' t$.

(i) ensures $\succ'$ is a (strict) inclusion of $\succ$ i.e. $\succ' \supset \succ$.

Suppose $t \succ' t$ for some $t \in X$. Since $t \succ t$ is false by irreflexivity of $\succ$ and we cannot have both $t = x$ and $t = y$, it must be that $t \succeq x$ and $y \succeq t$ with one holding without equality, yet this implies $y \succ x$ by transitivity of $\succ$, contradicting that x, y are incomparable under $\succ$. Hence we cannot have $t \succ' t$ for any t, so $\succ'$ is irreflexive.

Next, suppose $r \succ' s$ and $s \succ' t$. If $r \succ s$ and $s \succ t$, then transitivity of $\succ$ implies $r \succ t$ so $r \succ' t$. If $r \succ s$ but $s \not\succ t$, then $s \succeq x$ and $y \succeq t$, and so $r \succ x$, giving $r \succ' t$. Similarly if $r \not\succ s$ but $s \not\succ t$, then $r \succeq x$ and $y \succeq s$, so $y \succ t$, and hence $r \succ' t$. Finally if $r \not\succ s$ and $s \not\succ t$, then $r \succeq x$ and $y \succeq s$ and $s \succeq x$ and $y \succeq t$. But $s \succeq x$ and $y \succeq s$ imply $y \succ x$, contradicting that x, y are incomparable. Thus this final case is an impossibility. We have shown that $\succ'$ is transitive. Thus we have a procedure that takes any non-total strict partial order and yields an extension that compares two previously incomparable elements.

Let $\mathcal{R}$ be the set of all strict partial orders on X that include $\succ$, ordered by the inclusion relation $\supset$. Clearly $\supset$ is a strict partial order.

We claim $\mathcal{R}$ has a maximal element. Take any chain $\mathcal{C} \subseteq \mathcal{R}$.[7] Denote by the union $\bigcup \mathcal{C}$ the relation defined by $s \left(\bigcup C\right) t$ iff $s \succ' t$ for some $\succ' \in \mathcal{C}$.

Since $\succ \subseteq \succ'$ for all $\succ' \in \mathcal{C}$, we have that $\succ$ is contained in $\bigcup \mathcal{C}$. Moreover, $\bigcup \mathcal{C}$ is irreflexive, since each $\succ' \in \mathcal{C}$ is irreflexive. Next, $\bigcup \mathcal{C}$ is transitive, since if $r \left(\bigcup \mathcal{C}\right) s$ and $s \left(\bigcup \mathcal{C}\right) t$ then there are partial orders $\succ', \succ'' \in \mathcal{C}$ such that $r \succ' s$ and $s \succ'' t$; since $\mathcal{C}$ is a chain, we must have $\succ' \subseteq \succ''$ or $\succ'' \subseteq \succ'$, so $r \succ' s \succ' t$ or $r \succ'' s \succ'' t$, and so by transitivity, $r \succ' t$ or $r \succ'' t$, and thus $r \left(\bigcup \mathcal{C}\right) t$. Thus $\bigcup \mathcal{C}$ is a strict partial order. Finally, $\bigcup \mathcal{C}$ is an upper bound for $\mathcal{C}$, since by definition $\succ' \subset \bigcup \mathcal{C}$ for every $\succ' \in \mathcal{C}$. Since every chain $\mathcal{C}$ has an upper bound, Zorn's lemma gives that there is a maximal element $\succ^*$ of $\mathcal{R}$. Towards contradiction, suppose $\succ^*$ is not a total order. Then there are some incomparable elements x, y under $\succ^*$, but then applying the procedure previously outlined, we can construct a strict partial order $\succ'$ that compares x, y with $\succ' \supset \succ^*$, contradicting that $\succ^*$ is maximal. Hence $\succ^*$ is a strict total order. □

A version of the theorem holds for weak partial orders, with similar proof. Note we needed Zorn's lemma, which is equivalent to the axiom of choice.[8]

Using Szpilrajn's extension theorem, we can get something close to a representation result for countable sets of alternatives, as with Theorem 1. However, we cannot generally obtain a representation of $\succ$ in the sense of Definition 3. For example, suppose $X = \{x, y, z\}$ and $x \sim y$, $y \sim z$ and $x \succ z$. For a utility function $u : X \to \mathbb{R}$ to represent $\succ$, we need that $u(x) = u(y) = u(z)$ and $u(x) > u(z)$, which is a contradiction. Instead the best we can do is the following:

Theorem 6. *Suppose $\succ$ is a strict partial order on a set X and suppose $X/_{\approx}$ is countable. Then there is a function $u : X \to \mathbb{R}$ such that, for all $x, y \in X$,*

$$\begin{aligned} x \succ y &\text{ implies } u(x) > u(y), \text{ and} \\ x \approx y &\text{ implies } u(x) = u(y). \end{aligned}$$

Proof. Let $\succ'$ be defined on $X/_{\approx}$ so that $a \succ' b$ iff $x \succ y$ for some $x \in [a]$ and $y \in [b]$. Then by Proposition 5(v), $\succ'$ is a strict partial order on $X/_{\approx}$. Applying Szpilrajn's extension theorem (Theorem 5), we have that there is some strict total order $\succ^*$ on $X/_{\approx}$ that is an extension of $\succ'$. Now, the proof of Theorem 1 implies there is a function $u : X/_{\approx} \to \mathbb{R}$ such that $a \succ^* b$ iff $u(a) > u(b)$, for all $a, b \in X/_{\approx}$. For every $a \in X/_{\approx}$, define $u(x) = u(a)$ for each $x \in [a]$. Then u is defined on X and $x \succ y$ implies $u(x) > u(y)$. Finally, if $x \approx y$, then $x, y \in [a]$ for some $a \in X/_{\approx}$, and so $u(x) = u(y)$. □

Similarly, we can obtain something a bit like a representation theorem for uncountable sets, as with Theorem 2:

[7] Recall a chain is a set that is totally ordered under $\supset$.

[8] If you do not believe in the axiom of choice, then I am sorry.

Theorem 7. *Let $\succ$ be a strict partial order on X and let $\succ'$ be the strict partial order induced on $X/_{\approx}$ by $\succ$. Suppose there is a countable subset $Z \subset X/_{\approx}$ such that Z is $\succ'$-order dense in $X/_{\approx}$. Then there is a function $u : X \to \mathbb{R}$ such that, for all $x, y \in X$,*

$$x \succ y \textit{ implies } u(x) > u(y), \textit{ and}$$
$$x \approx y \textit{ implies } u(x) = u(y).$$

Proof. Fix the $\succ'$-order dense countable subset $Z \subseteq X/_{\approx}$. By Szpilrajn's extension theorem (Theorem 5), there is a strict total order $\succ^*$ on $X/_{\approx}$ that extends $\succ'$.

Now define $\cong$ on $X/_{\approx}$ so that $a \cong b$ iff

(i) $a = b$, or

(ii) $a, b \notin Z$ and there is no $c \in Z$ s.t. $a \succ^* c \succ^* b$ or $b \succ^* c \succ^* a$.

By definition, $\cong$ is reflexive and symmetric. Suppose $a \cong b$ and $b \cong c$ for distinct $a, b, c \in X/_{\approx}$ (if a, b, c are not all distinct, then $a \cong c$ holds trivially). Then either (a) $a \succ^* b$ and $a \succ^* c$ or (b) $a \succ^* b$ and $c \succ^* b$ or (c) $b \succ^* a$ and $b \succ^* c$ or (d) $b \succ^* a$ and $c \succ^* b$. In each case, by (ii) there is no $d \in Z$ s.t. $a \succ^* d \succ^* c$ or $c \succ^* d \succ^* a$. Hence $a \cong c$. Thus *cong* is transitive, and we can conclude it is an equivalence relation on $X/_{\approx}$.

Take $r, s, t \in (X/_{\approx})/_{\cong}$ [equivalence classes of $X/_{\approx}$ under $\cong$.] Define the relation $\succ''$ on $(X/_{\approx})/_{\cong}$ so that $s \succ'' t$ iff $s \neq t$ and $a \succ^* b$ for some $a \in [s]$ and $b \in [t]$. Now $\succ''$ is a strict total order on $(X/_{\approx})/_{\cong}$ given $\succ^*$ is a strict total order and $\cong$ is an equivalence relation on $X/_{\approx}$. Define $T := \{t \in (X/_{\approx})/_{\cong} \mid a \in [t] \text{ for some } a \in Z\}$. If $s, t \notin T$ and $s \succ'' t$ then there are $a, b \in X/_{\approx}$ s.t. $a \succ^* b$, $a \in [s]$, $b \in [t]$ and $a, b \notin Z$. Since $s \succ'' t$, $s \neq t$ and thus $a \ncong b$, we have that $a \succ^* c \succ^* b$ for some $c \in Z$. Now c lives in some equivalence class $[r]$ under $\cong$, and we have $r \in T$. Thus T is $\succ''$-order dense in $(X/_{\approx})/_{\cong}$. By Theorem 2, there is a function $f : (X/_{\approx})/_{\cong} \to \mathbb{R}$ such that $s \succ'' t$ iff $f(s) > f(t)$ for all $s, t \in (X/_{\approx})/_{\cong}$. Now, if $a \succ' b$ for $a \in [s]$ and $b \in [t]$, then $a \succ^* b$ and either $s = t$ or $s \succ'' t$. If $a \in Z$ or $b \in Z$, then $a \ncong b$ and so $s \neq t$. Hence suppose $a, b \notin Z$. If $s = t$ then there is no $c \in Z$ s.t. $a \succ^* c \succ^* b$, which contradicts that Z is $\succ^*$-order dense in $X/_{\approx}$. Thus $a \succ' b$ and $a, b \notin Z$ together imply $a \succ' c \succ' b$ for some $c \in Z$, and so $a \succ' b$ implies $s \succ'' t$.

For every $t \in (X/_{\approx})/_{\cong}$, let $u(a) = f(t)$ for all $a \in [t]$. From the definition of f, we have that $a \succ' b$ implies $u(a) > u(b)$. For every $a \in X/_{\approx}$, take $u(x) = u(a)$ for all $x \in [a]$. If $x \succ y$ and $x \in [a], y \in [b]$, then $a \succ' b$ so $x \succ y$ implies $u(x) > u(y)$. If $x, y \in [a]$ then $u(x) = u(y)$, so $x \approx y$ implies $u(x) = u(y)$. $\square$

This theorem appears in Fishburn (1970) based on Richter (1966). Note that the converse of the theorem is not true. If $\succ$ is defined on $\mathbb{R}$ so that $x \succ y$ iff $x > y$ and $y = x - n$ for some integer n. Then $u(x) = x$ is a function such that $x \succ y$ implies $u(x) > u(y)$ and $x \approx y$ implies $u(x) = u(y)$. For any countable subset $Z \subset \mathbb{R}$, we can find some $x \in \mathbb{R}$ so that $x \notin Z$ and $(x - 1) \notin Z$. But then there is no $z \in Z$ such that $x \succ z \succ x - 1$. Thus there is no countable $\succ'$-order dense subset of $\mathbb{R}/_{\approx}$.

As you might imagine by now, we can also prove an analogue of Theorem 3, under other conditions. Fishburn (1970, p.34) gives the theorem with proof – he also discusses how it links to upper semicontinuity (pp.38-9).

1.3 Revealed preference theory

So far, we have taken preferences as given. However, in practice, we never observe preferences. Rather, we observe choices. If we assume decisionmakers are making choices that they most prefer according to some underlying preference relation, then we can gather data on the preference relation by observing their choices. For example, if given a choice between a vanilla ice cream v and a lemon curd tart t, a decisionmaker opts for v, then we can conclude $v \succsim t$. This relies, of course, on the assumption that the decisionmaker is choosing an option that they most prefer. In many settings, we might not expect this assumption to hold. Decisionmakers might make mistakes, particularly if they have to make decisions quickly, face complex choice situations, or have false beliefs. In other cases, decisionmakers might not even be aware of their own preferences because they lack information about some of the alternatives. More on these caveats later.

When it comes to revealed preference, *menus* become important. Decisionmakers are generally presented with a limited menu of feasible alternatives. If a dessert shop does not sell chocolate mousse c, then the fact that a decisionmaker picks vanilla ice cream tells us nothing about that decisionmaker's preferences over chocolate mousse.[9]

Given a set of alternatives X, we call any nonempty subset of X a *menu*, and we denote the set of all menus in X by $\mathscr{M}(X)$. The first thing that comes to mind from "menu" is a list of food items available to order, but a menu might instead be a list of political candidates one could vote for, for example. The menu $M \subseteq X$ the decisionmaker actually chooses from in a given choice problem is the set of *feasible* alternatives. In a restaurant context (to avoid issues, imagine someone else is paying), this is the set of dishes the restaurant actually offers. In elections, this would be the list of candidates (or parties) on the ballot. In classic demand theory, the menu the decisionmaker faces is their budget set.

Definition 7 (Choice correspondence)**.** Let X be a set of alternatives. Let $\mathscr{M} \subseteq \mathscr{M}(X)$ be a family of nonempty subsets of X. A function $c : \mathscr{M} \to \mathscr{M}(X)$ is called a *choice correspondence* if $c(M) \subseteq M$ for each menu $M \in \mathscr{M}$.

If c is a choice correspondence such that $c(M)$ is a singleton for each $M \in \mathscr{M}$, then we call c a *choice function*.

Because choice correspondences map into $\mathscr{M}(X)$, they are nonempty-valued. We thus assume the decisionmaker always chooses something from the menu. This is not a particularly taxing assumption, because we can incorporate any action (including inaction) into the set of alternatives – choosing not to order anything at the restaurant is on the notional menu a diner chooses from. However, the assumption does rule out that a decisionmaker might be frozen in indecision or otherwise "refuse to decide".

If decisionmakers choose alternatives to maximize their own preferences, then it is natural to think of a decisionmaker's preference relation $\succ$ as inducing a choice correspondence.

[9] At least to a point – if there is a café selling next door to the dessert shop, then the fact that the decisionmaker went to the dessert shop is potentially informative.

Definition 8. Let X be a set of alternatives, let $\mathscr{M} \subseteq \mathscr{M}(X)$, and let $\succ$ be a strict preference relation $\succ$ on X such that every menu $M \in \mathscr{M}$ has a maximal element with respect to $\succ$. The *choice correspondence induced by* $\succ$ is the choice correspondence $c_\succ : \mathscr{M} \to \mathscr{M}(X)$ defined by

$$c_\succ(M) := \{x \in M \mid \text{there is no } y \in M \text{ such that } y \succ x\}, \quad \text{for all } M \in \mathscr{M}.$$

The induced choice correspondence $c_\succ$ would not be well-defined if there were some menu $M \in \mathscr{M}$ that has no maximal element with respect to $\succ$, since this would imply $c_\succ(M)$ is empty. Nonemptiness of $c_\succ(M)$ is guaranteed if M is finite. However, if M is infinite, we must be more careful. For example, take $X = \mathbb{N}$ and suppose $x \succ y$ iff $x > y$. Then $c_\succ(X)$ would be empty, since for any integer $n \in c_\succ(X)$, we have $n+1 \succ n$, contradicting the definition of $c_\succ$.

The conditions under which we can obtain a choice correspondence from a preference relation are straightforward. The converse – obtaining a preference relation from a choice correspondence (or choice data) – is much more interesting and conditions under which we can obtain a "reasonable" strict preference relation from a choice correspondence are less trivial.

Definition 9. We say a choice correspondence $c : \mathscr{M} \to \mathscr{M}(X)$ is *rationalized* by a strict preference relation $\succ$ if $c = c_\succ$. We say c is rationalizable if it is rationalized by some strict preference relation $\succ$.

Note that throughout this section, we have not taken a stance on what type of binary relation the strict preference relation $\succ$ is. The definitions above apply regardless of our stance, though if we say a particular choice correspondence is rationalizable in a context where we have taken the stance that strict preference relations are weak orders, say, then we mean that there is some weak order rationalizing c.

1.3.1 Choices from a complete set of finite menus

We begin with the case where, although the set of alternatives X is arbitrary, the menus a decisionmaker chooses from are always finite. This is simpler than the general case, because if a decisionmaker has an acyclic preference relation $\succ$, there will always be some choice in any finite menu M that is maximal under $\succ$.

Axioms. Let X be an arbitrary set of alternatives and assume $\mathscr{M} \subseteq \mathscr{M}(X)$ is a collection of finite menus in X in that is "complete" in the sense that it contains all pairs and triples. For Sen's γ, we further assume $\mathscr{M}$ is closed under finite union. Let $c : \mathscr{M} \to \mathscr{M}(X)$ be a choice correspondence.

(**R1**) *Sen's* α. For all $x \in X$ and $M_1, M_2 \in \mathscr{M}$, if $x \in M_1 \subseteq M_2$ and $x \in c(M_2)$, then $x \in c(M_1)$.

(**R2**) *Sen's* β. For all $x, y \in X$ and $M_1, M_2 \in \mathscr{M}$ with $M_1 \subseteq M_2$, if $x, y \in c(M_1)$ and $y \in c(M_2)$ then $x \in c(M_2)$.

(**R3**) *Sen's γ.* For all $x \in X$ and any $M_1, M_2 \in \mathscr{M}$, if $x \in c(M_1) \cap c(M_2)$, then $x \in c(M_1 \cup M_2)$.

(**R4**) *Arrow's C4.* For all $M_1, M_2 \in \mathscr{M}$, if $M_1 \subseteq M_2$ and $M_1 \cap c(M_2)$ is nonempty, then $c(M_1) = M_1 \cap c(M_2)$.

(**R5**) *Weak axiom of revealed preference (WARP).* For all $M_1, M_2 \in \mathscr{M}$ and all $x, y \in M_1 \cap M_2$, if $x \in c(M_1)$ and $y \in c(M_2)$ then $x \in c(M_2)$.[10]

The first three of these axioms derive their names from Sen (1971). Sen's α was originally proposed by Chernoff (1954), and is also known as *Chernoff's axiom* or Arrow's *independence of irrelevant alternatives* (after Arrow (1959)). To explain more wordily, imagine you are at an ice cream shop. Sen's α states that if you would be willing to choose chocolate ice cream over every other flavour, then when you have the choice of chocolate, vanilla, or lavender, you'd better be willing to pick chocolate. Sen's α thus captures a notion of menu independence – it requires that altering the menu in an irrelevant way does not change the decisionmaker's choices.

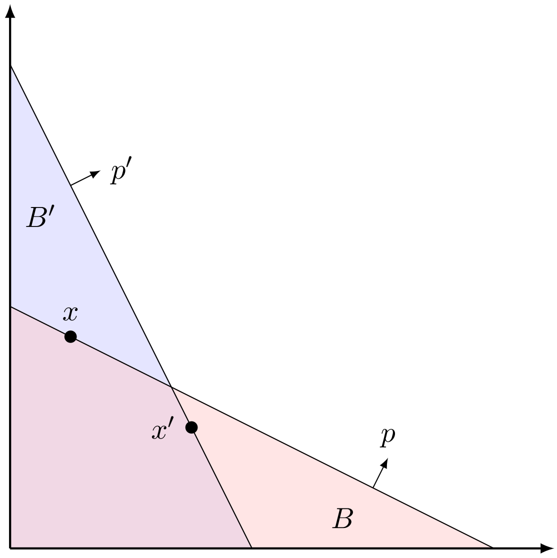
Sen's β, on the other hand, states that if given the choice between vanilla and chocolate, you pick chocolate, but given the choice between vanilla, chocolate and lavender, you pick vanilla, then you should be happy with chocolate in the second case too.

Arrow's C4 is due to Arrow (1959). Suppose you enter the shop, peruse the main and specials menus, and pick a scoop of chocolate ice cream from the main menu and coffee banana fudge from the specials menu. Arrow's C4 says if you had entered the shop and the specials menu were unavailable, you would still have picked chocolate.

The most famous of these axioms, WARP, was probably first stated in the form above by Arrow (1959), but can trace its history in other forms back to Samuelson (1938). Suppose, had you entered the ice cream shop and looked only at the main menu, you would have seen both chocolate and lavender on the menu and picked chocolate. As it happens, the specials board is up, and you have picked lavender. WARP implies you could just as easily have chosen chocolate in this case.

Example 5 (A WARP violation in consumer theory)**.**

[10] We can equivalently write (**R5**) as: for all $M_1, M_2 \in \mathscr{M}$ and all $x, y \in M_1 \cap M_2$, if $x \in c(M_1)$ and $y \notin c(M_1)$ then $y \notin c(M_2)$ (an axiom Kreps (2013) labels "choice coherence"). To see this is equivalent, first suppose (**R5**) holds, $x, y \in M_1 \cap M_2$ and $x \in c(M_1)$. By definition, if $y \in c(M_2)$ then $y \in c(M_1)$. Hence $y \notin c(M_1)$ necessitates $y \notin c(M_2)$, or else we have a contradiction. Conversely, if $y \in c(M_2)$, then "choice coherence" implies $y \in c(M_1)$, or else we have a contradiction.



B and B' are the budget sets induced by price vectors p and p', and decisionmakers choose a bundle in $\mathbb{R}^2_+$. We have $x \in c(B)$ but $x \notin c(B')$, and $x' \in c(B')$ but $x' \notin c(B)$, despite $x, x' \in B \cap B'$.

Proposition 6. *Let X be a set of alternatives and suppose $c : \mathscr{M} \to \mathscr{M}(X)$ is a choice correspondence, where $\mathscr{M}$ is a collection of finite menus in X including all pairs and triples. Then the following are equivalent:*

(i) *c satisfies both Sen's α (**R1**) and Sen's β (**R2**);*

(ii) *c satisfies Arrow's C4 axiom (**R4**);*

(iii) *c satisfies WARP (**R5**), and*

(iv) *c is rationalized by some weak order $\succ$.*

Proof. First, suppose (i) holds. Take $M_1, M_2 \in \mathscr{M}$ with $M_1 \subseteq M_2$ and suppose $M_1 \cap c(M_2)$ is nonempty. For all $x \in M_1 \cap c(M_2)$, (**R1**) implies $x \in c(M_1)$. Thus $c(M_1) \subseteq M_1 \cap c(M_2)$. Since $M_1 \cap c(M_2)$ is nonempty, there is some $y \in c(M_1)$ such that $y \in c(M_2)$, and (**R2**) thus implies that for any other $z \in c(M_1)$, we have $z \in c(M_2)$. Thus $c(M_1) \supseteq M_1 \cap c(M_2)$. Hence we have shown (i) implies (ii).

Next, suppose (ii) holds. Take $M_1, M_2 \in \mathscr{M}$. If $M_1 \cap c(M_2)$ and $M_2 \cap c(M_1)$ are empty, then the statement of WARP holds for M_1, M_2 vacuously. Hence suppose at least one of these sets is nonempty, say $M_2 \cap c(M_1)$ (the other case is symmetric). Then (ii) implies $c(M_2) = M_2 \cap c(M_1)$. Hence for $x \in M_1 \cap M_2$, $x \in c(M_1)$ implies $x \in c(M_2)$. Thus (ii) implies (iii).

Now suppose (iii) holds. Construct $\succ$ by the following procedure: for each menu $M \in \mathscr{M}$, whenever $x, y \in M$, $x \in c(M)$ and $y \notin c(M)$, put $x \succ y$. We need to verify $\succ$ is a weak order, i.e. asymmetric and negatively transitive. If $\succ$ is not asymmetric, then there are $x, y \in X$ such that $x \succ y$ and $y \succ x$, and hence there must be menus

$M_1, M_2 \in \mathscr{M}$ such that (a) $x, y \in M_1 \cap M_2$, (b) $x \in c(M_1)$ but $y \notin c(M_1)$ and (c) $y \in c(M_2)$ but $x \notin c(M_2)$. But this would immediately contradict (iii). Hence $\succ$ is asymmetric. Now, suppose towards contradiction that $\succ$ is not negatively transitive. Then there are $x, y, z \in X$ with $x \succ z$ but neither $x \succ y$ nor $y \succ z$ (Lemma 1). It follows that there is some menu $M_1 \in \mathscr{M}$ with $x, z \in M_1$, $x \in c(M_1)$ and $z \notin M_1$. If $y \in M_1$ then we are done because $y \notin c(M_1)$ implies $x \succ y$ and $y \in c(M_1)$ implies $y \succ z$. Hence suppose $y \notin M_1$. Since $\mathscr{M}$ contains all triples, it contains the triple $M_2 = \{x, y, z\}$. Since c is nonempty-valued, $c(M_2)$ is nonempty, and by asymmetry, $z \notin c(M_2)$. Hence $x \in c(M_2)$ or $y \in c(M_2)$. If $y \in c(M_2)$ then $y \succ z$, and if $y \notin c(M_2)$, we must have $x \in c(M_2)$, so $x \succ y$. Hence $\succ$ is negatively transitive.

Finally, suppose (iv) holds, so $c = c_\succ$ for a weak order $\succ$. Take $M_1, M_2 \in \mathscr{M}$ with $M_1 \subseteq M_2$. For any $x \in c(M_2)$, by definition there is no $y \in M_2$ with $y \succ x$, and thus no $y \in M_1 \subseteq M_2$ with $y \succ x$, so $x \in c(M_1)$. Therefore c satisfies (**R1**). Now take any $x, y \in c(M_1)$ with $y \in c(M_2)$. Then $y \not\succ x$, since otherwise $x \notin c(M_1)$, and $x \not\succ y$, since otherwise $y \notin c(M_1)$. Thus $y \sim x$. If $x \notin c(M_2)$, then there is some $z \in M_2$ with $z \succ x$, which by Proposition 2 would imply $z \succ y$. But since $y \in c(M_2)$, there is no $z \in M_2$ such that $z \succ y$, yielding a contradiction. Hence c satisfies (**R2**). Thus (i) follows from (iv), completing the proof. □

Proposition 6 applies to weak orders. Sen (1971) also asks the more general question of when there might be *any* preference relation rationalizing a choice correspondence. Note we impose no properties on strict preference relations beyond asymmetry.

Proposition 7. *Let X be a set of alternatives and suppose $c : \mathscr{M} \to \mathscr{M}(X)$ is a choice correspondence, where $\mathscr{M}$ is a collection of finite menus in X including all pairs that is closed under finite union.*[11] *Then c is rationalized by some strict preference relation $\succ$ iff c satisfies both Sen's α* (**R1**) *and Sen's γ* (**R3**).

Proof. First, suppose (**R1**) and (**R3**) hold. Construct $\succ$ by the same procedure as in the proof of Proposition 6: for each menu $M \in \mathscr{M}$, whenever $x, y \in M$, $x \in c(M)$ and $y \notin c(M)$, put $x \succ y$.

Clearly $\succ$ is irreflexive. Take $x, y \in X$ and suppose $x \succ y$. Since $\mathscr{M}$ contains all pairs, it contains $M_1 = \{x, y\}$. Since $x \succ y$, there is some menu $M_2 \in \mathscr{M}$ with $M_2 \supseteq M_1$, $x \in c(M_2)$ and $y \notin c(M_2)$. By (**R1**), we therefore have $x \in c(M_1)$. If we also have $y \succ x$, then there must be another menu $M_3 \supset M_1$ with $y \in c(M_3)$ and $x \notin c(M_3)$, and then (**R1**) implies $y \in c(M_1)$. By hypothesis, M_3 is finite. Suppose M_3 does not contain any proper subset $M_3' \in \mathscr{M}$ such that $M_1 \subseteq M_3'$, $y \in c(M_3')$ but $x \notin c(M_3')$ (this is without loss, for otherwise M_3 a set with these properties). Then we can write M_3 as the union of two distinct menus $M_4, M_5 \supseteq M_1$ such that $x \in c(M_4)$ and $x \in c(M_5)$. By (**R3**), it follows that $x \in c(M_3)$, yielding a contradiction. Hence $\succ$ is asymmetric, and so is a strict preference relation.

Conversely, suppose c is rationalized by strict preference relation $\succ$. Take menus $M_1 \subseteq M_2$ and $x \in c(M_2)$. Then there is no $y \in M_2$ such that $y \succ x$, and so there is

[11] Note this implies $\mathscr{M}$ contains all nonempty, non-singleton finite sets.

no $y \in M_1 \subseteq M_2$ for which $y \succ x$. Thus $x \in c(M_1)$, so (**R1**) is satisfied. Next take $M_1, M_2 \in \mathscr{M}$ and suppose $x \in c(M_1) \cap c(M_2)$. Since $\mathscr{M}$ is closed under finite union, $M_1 \cup M_2 \in \mathscr{M}$. Since $x \in c(M_1)$, there is no $y_1 \in M_1$ for which $y_1 \succ x$ and since $x \in c(M_2)$, there is no $y_2 \in M_2$ for which $y_2 \succ x$. Since any $y \in M_1 \cup M_2$ must lie in M_1 or M_2, there is thus no $y \in M_1 \cup M_2$ for which $y \succ x$, so $x \in c(M_1 \cup M_2)$. Thus (**R3**) is satisfied. $\square$

In Proposition 7, the strict preference relation $\succ$ rationalizing a choice correspondence c needs to be acyclic, since a cycle $x \succ y$, $y \succ z$, $z \succ x$ would imply $c(\{x, y, z\})$ is empty. However, it does not need to be transitive.

Since weak orders are strict preference relations, we can conclude from Proposition 6 that Sen's α and Sen's β must together imply Sen's γ.[12] This is also easy to see via WARP, noting (**R1**)-(**R2**) are together equivalent to (**R5**) by Proposition 6:

Corollary 2. *If choice correspondence $c : \mathscr{M} \to \mathscr{M}(X)$ satisfies WARP* (**R5**)*, then it satisfies Sen's γ* (**R3**).

Proof. Since any choice correspondence c must be nonempty-valued, if we take menus $M_1, M_2 \in \mathscr{M}$, then there is some $y \in c(M_1 \cup M_2)$. Now $y \in M_1$ or $y \in M_2$. Wlog suppose the former. Since $y \in M_1$ and $y \in c(M_1 \cup M_2)$, (**R1**) implies $y \in c(M_1)$. Now, if $x \in c(M_1)$, then (**R5**) and $y \in c(M_1)$ imply $x \in c(M_1 \cup M_2)$, which implies (**R3**). $\square$

The lack of transitivity in the strict preference relations of Proposition 7 is a little disturbing. Previously we worked with strict partial orders, allowing us to capture intransitive indifference without sacrificing the transitivity of strict preference relations. For this, we will make use of the following:

Axiom. Let X be an arbitrary set of alternatives and assume $\mathscr{M} \subseteq \mathscr{M}(X)$ is a collection of finite menus in X containing all pairs and triples. Let $c : \mathscr{M} \to \mathscr{M}(X)$ be a choice correspondence.

(**R6**) For all $M_1, M_2 \in \mathscr{M}$ with $M_1 \subseteq M_2$, if $x, y \in M_1$, $x \in c(M_2)$ and $y \notin c(M_2)$, then $y \notin c(M_1)$.

(**R6′**) *Sen's δ.* For all $M_1, M_2 \in \mathscr{M}$ with $M_1 \subseteq M_2$, if $x, y \in c(M_1)$ for $x \neq y$, then $c(M_2) \neq \{x\}$.

We will be using (**R6**), which I find more intuitive. Sen's δ (**R6′**) is the version in Sen (1971).

Proposition 8. *Let X be a set of alternatives and suppose $c : \mathscr{M} \to \mathscr{M}(X)$ is a choice correspondence, where $\mathscr{M}$ is a collection of finite menus in X including all pairs and triples. Then c is rationalized by some strict partial order $\succ$ iff c satisfies both Sen's α* (**R1**) *and condition δ* (**R6**).

[12] Note Sen's β does not by itself imply Sen's γ.

Proof. First suppose (**R1**) and (**R6**) hold. Construct $\succ$ by the usual procedure: for each menu $M \in \mathscr{M}$, if $x, y \in M$, $x \in c(M)$ and $y \notin c(M)$, put $x \succ y$.

Clearly $\succ$ is irreflexive. Take $x, y \in X$ and suppose $x \succ y$. Then there is some menu M with $x, y \in M$, $x \in c(M)$ and $y \notin c(M)$. Now (**R1**) implies $x \in c(\{x, y\})$ and (**R6**) implies $y \notin c(\{x, y\})$. If $y \succ x$ then there is another menu $M' \supset \{x, y\}$ with $y \in c(M')$, $x \notin M'$. But then (**R6**) implies $y \in c(\{x, y\})$, a contradiction. Hence $\succ$ is asymmetric. Significantly for the next part of the proof, we have also shown that $x \succ y$ implies $c(\{x, y\}) = \{x\}$.

Now take $x, y, z \in X$ and suppose $x \succ y$ and $y \succ z$. Then $c(\{x, y\}) = \{x\}$ and $c(\{y, z\}) = \{y\}$. If $y \in c(\{x, y, z\})$, then we would have $y \in c(\{x, y\})$ by (**R1**), yielding a contradiction. If $z \in c(\{x, y, z\})$, we would have $z \in c(\{y, z\})$, again yielding a contradiction. Finally, nonemptiness of c implies $c(\{x, y, z\}) = \{x\}$. Hence $x \succ z$. Thus $\succ$ is transitive, so we have that $\succ$ is a strict partial order.

Conversely, suppose c is rationalized by strict partial order $\succ$. Since (**R1**) holds per the proof given in Proposition 7, we need only show that (**R6**) also holds. Take menus $M_1 \subseteq M_2$. Fix $x, y \in M_1$ and suppose $x \in c(M_2)$. Then $y \not\succ x$, so either $x \succ y$ or $x \sim y$. If $x \succ y$ then $y \notin c(M_2)$ and $y \notin c(M_1)$, so (**R6**) is not violated. If $y \sim x$ then $y \in c(M_2)$ and $y \in c(M_1)$, so again (**R6**) holds. Hence c satisfies both (**R1**) and (**R6**). □

1.3.2 Bowing to empiricists

The propositions in the previous section are beautiful as they are. However, more empirically-minded people like to try the hopeless task of estimating preference relations from observations. The framework above then runs into some problems:

(1) When a decisionmaker chooses from a given menu M, we do not typically observe the entire set $c(M)$. Usually, we observe just one $x \in c(M)$.

(2) In empirical settings, we do not typically observe choices for a "complete" set of menus $\mathscr{M}$.

Issue (1) is easily resolved. Given a choice correspondence $c_\succ$ induced by $\succ$, we can always make a selection to yield a choice function. Conversely, if we observe only singleton choices for each menu, we can rationalize the observed choice function c by taking $x \sim y$ for all $x, y \in X$.

Not so fast! That would be too easy (and pretty useless for the empiricists). Rather, we usually rule out indifference by assumption, so we take the observed choice $x \in M$ to be the *only* choice in $c(M)$. This might not always be reasonable but it forces us to construct a preference relation that is not trivial. In this case, we have a simple corollary of Proposition 6:

Corollary 3. *Let X be a set of alternatives and suppose $c : \mathscr{M} \to \mathscr{M}(X)$ is a singleton-valued choice correspondence, where $\mathscr{M}$ is a collection of finite menus in X including all pairs and triples. Then c is rationalized by some weak order $\succ$ iff c satisfies Sen's α* (**R1**)*.*

Proof. By Proposition 6, if $c : \mathscr{M} \to \mathscr{M}(X)$ is rationalized by weak order $\succ$, then it satisfies (**R1**).

Conversely, suppose c is singleton-valued and satisfies (**R1**). Note c satisfies (**R2**) trivially. Define $\succ$ by $x \succ y$ if there exists some menu $M \in \mathscr{M}$ with $x, y \in M$ such that $x \in c(M)$ (since c is singleton-valued, this implies $y \notin c(M)$). By the proof of Proposition 6, $\succ$ is a weak order. □

Note the induced weak preference relation $\succsim$ corresponding to a weak order $\succ$ in Corollary 3 is antisymmetric. Since $c(\{x, y\})$ selects either x or y, we always have $x \succ y$ or $y \succ x$ for distinct $x, y \in X$. Thus $x \sim y$ iff $x = y$. Assuming away indifference like this is hardly satisfying – can we can do better? The answer is yes if we are prepared to make further assumptions (or have other data). For example, in classic consumer theory, there are ways to resolve both issues (1) and (2) while allowing the possibility of indifference. Similarly, if we assume there is some procedure by which decisionmakers resolve indifferences, then we can allow indifferences – Rubinstein & Salant (2006) consider the role of order in a list in resolving indifferences, for example [see Section **??**].

Let us turn to the more interesting issue (2), once again choosing to neglect issue (1). In a very carefully controlled experiment with a tiny number of alternatives, it might be possible to observe choices for all pairs and triples. However, in most settings, we have a more limited set of choice data, and are unlikely to have data on a complete set of menus. The propositions above then fail:

Example 6. The dessert shop has a pop-up stall at the market selling only their three best-selling desserts: apple strudel a, banoffee pie b and chocolate mousse c. The set of alternatives is thus $X = \{a, b, c\}$. You really like their desserts and are very hungry, so you keep going back. However, each time they are sold out of one of the products, so you face a menu of only two of the three desserts, i.e. $\mathscr{M} = \{\{a, b\}, \{b, c\}, \{a, c\}\}$. Your choices are as follows:

$$c(\{a, b\}) = \{a\},$$
$$c(\{b, c\}) = \{b\},$$
$$c(\{a, c\}) = \{c\}.$$

Then the choice correspondence c trivially satisfies (**R1**), (**R2**) and (**R4**), since none of the menus in $\mathscr{M}$ are subsets of each other. WARP (**R5**) is also trivially satisfied because all the intersections of menus in $\mathscr{M}$ are singletons.

Despite this, there is no weak order $\succ$ rationalizing c. From $c(\{a, b\}) = \{a\}$ we have $a \succ b$, from $c(\{b, c\}) = \{b\}$, we have $b \succ c$, and from $c(\{a, c\}) = \{c\}$, we have $c \succ a$. But this violates transitivity and thus negative transitivity.

Given more limited choice data, it is interesting (and useful for empiricists) to ask what conditions we need to ensure the choice correspondence is rationalizable.

Definition 10 (Principle of revealed preference)**.** Let $c : \mathscr{M} \to \mathscr{M}(X)$ be a choice correspondence defined on a set of finite menus $\mathscr{M}$ in a set of alternatives X.

(a) *Direct revealed preference.* For $x, y \in X$, we say x is *directly revealed preferred* to y, denoted $xR^d y$, if there is a menu $M \in \mathscr{M}$ with $x, y \in M$ and $x \in c(M)$.

(b) *Weak revealed preference.* For $x, y \in X$, we say x is (weakly) *revealed preferred* to y, denoted xRy if there is a sequence $x_1, \dots, x_n$ with

$$xR^d x_1 R^d \cdots R^d x_n R^d y.$$

(c) *Strict revealed preference.* For $x, y \in X$, we say x is *strictly revealed preferred* to y, denoted xSy, if there is a menu $M \in \mathscr{M}$ with $x, y \in M$, $x \in c(M)$ and $y \notin c(M)$.

Clearly, $xR^d y$ implies xRy. We will be needing to make restrictions on the revealed preference relations R and S:

Axiom. Let X be an arbitrary set of alternatives and assume $\mathscr{M} \subseteq \mathscr{M}(X)$ is a collection of finite menus in X. Let $c : \mathscr{M} \to \mathscr{M}(X)$ be a choice correspondence, and let R and S be the weak and strict revealed preference relations for c.

(**R7**) *Generalized axiom of revealed preference (GARP).* For all $x, y \in X$, if xRy then not ySx.

GARP ensures that Sen's α and Sen's β hold:

Lemma 5. *If the choice correspondence $c : \mathscr{M} \to \mathscr{M}(X)$ satisfies* (**R7**) *then c satisfies* (**R1**) *and* (**R2**).[13]

Proof. Suppose c satisfies (**R7**), and take menus $M_1, M_2 \in \mathscr{M}$ with $M_1 \subseteq M_2$.

If $x \in c(M_2)$, then xRy for all $y \in M_2$. Since c is nonempty-valued, there is some $y \in c(M_1) \subseteq M_2$. By (**R7**), not ySx, so $y \in c(M_1)$ implies $x \in c(M_1)$. Thus c satisfies (**R1**).

Next, suppose $x, y \in c(M_1)$ and $y \in c(M_2)$. Then xRy and yRx. If $x \notin c(M_2)$, we would have ySx, but this contradicts (**R7**). Hence c satisfies (**R2**). □

Proposition 9. *Let X be a set of alternatives and let $\mathscr{M} \subseteq \mathscr{M}(X)$ be a connected set of finite menus such that each $x \in X$ lies in at least one non-singleton menu $M \in \mathscr{M}$. Suppose $c : \mathscr{M} \to \mathscr{M}(X)$ is a choice correspondence. Then c is rationalized by some weak order $\succ$ iff c satisfies GARP* (**R7**).

Proof. Suppose c is rationalized by a weak order $\succ$. Fix $x, y \in X$ and consider any menu $M \in \mathscr{M}$ with $x, y \in M$ (if there is no such menu, the condition in (**R7**) holds trivially for x, y). If $x \succ y$, then $x \in c(M)$ and $y \notin c(M)$. Since this holds for all such menus M, there is no menu M such that $x, y \in M$, $y \in c(M)$ and $x \notin c(M)$. Hence not ySx.

Conversely, suppose (**R7**) holds. We will extend $\mathscr{M}$ by adding in the "missing" pairs and triples, and then extend c to this new set of menus. First extend to the missing triples. Take $\mathscr{M}_1^* := \mathscr{M} \cup \{\{x, y, z\} \subseteq X\}$. Construct $c_1^* : \mathscr{M}_1^* \to \mathscr{M}$ as follows:

[13] When we defined (**R1**) and (**R2**), we stated we were doing so on a set of menus containing all pairs and triples in X. Obviously we are relaxing that when we talk of c satisfying these axioms here. We maintain this relaxation throughout this section.

(i) if $M \in \mathscr{M}$, put $c^*(M) := c(M)$;

(ii) if $M \in \mathscr{M}_1^* - \mathscr{M}$ and there is a $M' \in \mathscr{M}$ with $M' \supset M$, if $x \in c(M')$ put $x \in c(M')$;

(iii) if $M \in \mathscr{M}_1^* - \mathscr{M}$ and there is no $M' \in \mathscr{M}$ with $M' \supset M$, but there is an $M'' \in \mathscr{M}$ with $M'' \subset M$, put $x \in c(M)$ if $x \notin c(M'')$;

(iv) for any remaining $M \in \mathscr{M}_1^* - \mathscr{M}$, put $c(M) = M$.

We claim this operation is such that c_1^* satisfies (**R1**) and (**R2**). Take $M_1, M_2 \in \mathscr{M}_1^*$ such that $M_1 \subseteq M_2$. Note we will never have $M_1, M_2 \in \mathscr{M}^* - M$ unless $M_1 = M_2$, since $\mathscr{M}^* - M$ contains only different triples. Thus there are only three cases to check:

(a) $M_1, M_2 \in \mathscr{M}$. Then by Lemma 5, we have that the conditions of (**R1**) and (**R2**) hold for these sets, since $c_1^* = c$ on these sets.

(b) $M_1 \in \mathscr{M}_1^* - \mathscr{M}$ and $M_2 \in \mathscr{M}$. Then by (ii), if $x \in c(M_2)$, $x \in c(M_1)$ so the condition in (**R1**) holds for these sets. Now suppose $y \in\in c(M_2)$, and $x, y \in c(M_1)$. If $x \notin c(M_2)$, then ySx. But $x \in c(M_1)$ implies there is some $M' \in \mathscr{M}$ with $x, y \in M'$ and $x \in c(M')$, so xRy, and thus ySx contradicts (**R7**). Thus the condition in (**R2**) holds for these sets.

(c) $M_1 \in \mathscr{M}$ and $M_2 \in \mathscr{M}_1^* - \mathscr{M}$. Then M_2 is covered either by (ii) or (iii). First suppose M_2 is covered by (ii). Suppose $x \in c(M_2)$. Then $x \in M'$ for some $M' \in \mathscr{M}$ with $M' \supset M_2$, and since c satisfies (**R1**) over $\mathscr{M}$ and $M_1 \subset M'$, $x \in c(M_1)$. If $x, y \in M_1$ and $y \in c(M_2)$, then there must be a $M' \in \mathscr{M}$ with $M' \supseteq M_2$ with $y \in M'$. By Lemma 5, we have $x \in c(M')$, and thus $x \in c(M_2)$. Hence the condition in (**R2**) holds for these sets.

Suppose M_2 is instead covered by (iii). Then there is no $x \in c(M_2)$ such that $x \in M_1$ so the statements of (**R1**) and (**R2**) hold trivially for such sets.

Next extend to the missing pairs. Take $\mathscr{M}^* := \mathscr{M}_1^* \cup \{\{x, y\} \subseteq X\}$. Construct $c^* : \mathscr{M}^* \to \mathscr{M}$ by following identical steps to (i)-(iv) with $\mathscr{M}^*$ in place of $\mathscr{M}_1^*$ and $\mathscr{M}_1^*$ in place of $\mathscr{M}$. By a similar proof to that for c_1^* above, this operation ensures that c^* satisfies (**R1**) and (**R2**). Now by Proposition 6, there is a weak order $\succ$ rationalizing c^*. Since $c(M) = c^*(M)$ for all $M \in \mathscr{M}$, it follows that c is also rationalized by $\succ$. □

1.3.3 Behavioural problems I: violations of basic revealed preference theory

"Show me the axiom and I'll design the experiment that refutes it" – Amos Tversky.

Revealed preference theory provides us with a way to test whether decisionmakers actually obey the nice axioms we dreamt up from our ivory towers. It turns out that in real life, it is very easy to get people to violate even our most basic modelling assumptions. Here are some sources of WARP violations, which we will delineate by which Sen-type axioms they violate.

(i) *Indiscernibles.* If the decisionmaker cannot distinguish between sufficiently similar alternatives, then we can generate violations of transitive indifference, as Luce's famous sugar example (Example 2) illustrated. In this case, Sen's β (**R2**) does not hold. Say $x \sim y$ and $y \sim z$ but $x \succ z$. Then $c(\{x, y, z\}) = \{x, y\}$ but $c(\{y, z\}) = \{y, z\}$, and (**R2**) would imply $z \in c(\{x, y, z\})$. Section 1.2 discusses options for modelling this.

(ii) *Menu dependence.* Sen's α (**R1**) captures a notion of menu independence that is easy to violate in real life.

Example 7 (Epistemic value of a menu)**.** One source of menu dependence is that the menu itself conveys information about the choice situation.

(a) *Dinner (Luce and Raiffa, 1957).* A diner visits a new restaurant, unsure about the quality of its chef. Told ahead of time that she has a choice between chicken p and steak s, she says she strictly prefers to order chicken, i.e. $c(\{p, s\}) = \{p\}$. However, on arriving at the restaurant and seeing the restaurant also serves frog legs f, she revises her order and now strictly prefers steak s over the other options, i.e. $c(\{p, s, f\}) = \{s\}$.

Following the Rubinstein approach to econometrics, you simply ask the diner why she changed her decision. She responds that she in her opinion, the quality of a chicken dish does not vary much with the quality of the chef but steak is much more variable. She prefers a good steak to chicken to a mediocre steak. On seeing frog legs on the menu, she concluded the chef is skilled, and thus that her steak would be well-prepared.

(b) *A cocaine offer (Sen, 1993).* An acquaintance you recently met invites you to their house for tea. You have the option of going for tea x or staying away y. You are eager to make new friends, so you are inclined to go for tea, i.e. $c(\{x, y\}) = \{x\}$. The acquaintance then quietly mentions they have a large parcel of cocaine they would like to move and they would pay you quite well if you helped them to deliver it, so you now have the option of doing this (z). At this point, you very much prefer to stay away, i.e. $c(\{x, y, z\}) = \{y\}$.

It is easy to see these cases are violations of Sen's α, *if* we accept how we have defined alternatives. One could argue that the definition of alternatives here conflates different alternatives together – going for tea, x, with the acquaintance is quite a different prospect when they reveal themselves to be in the cocaine trade, so we could treat it as a different alternative altogether. Sen (1993) takes issue with this argument, on the grounds that inter-menu consistency is impossible if the definition of alternatives are generally menu-dependent. I'm not sure I agree with Sen. It seems to me that a reasonable model here is to distinguish between the decisionmaker's tastes over a granular set of alternatives (that treats well-prepared steak and mediocre steak as distinct alternatives, say) and the decisionmaker's beliefs about which menu she is being presented with. Inter-menu consistency is ensured by being selective about the level of granularity of alternatives.

Example 8 (Decoy and compromise effects)**.** The decoy and compromise effects are common in the marketing literature. Suppose you are choosing between alternatives that differ along several dimensions. For example, say you are buying a shirt and you care about both price and quality. There is a dominance relation $\succ$ on the set of alternatives, i.e. you strictly prefer a higher quality, lower price shirt to a lower quality, higher price shirt. Alternatives that cannot be ranked via the dominance relation are incomparable. Take alternatives x, y that are incomparable and alternatives $x' \prec x$ and $y' \prec y$ where x' (y') is similar to x (y).

The *decoy effect* captures the idea that including a dominated alternative that is similar to one of the alternatives can highlight the advantages of that alternative. Thus we can have $c(\{x, x', y\}) = \{x\}$ and $c(\{x, y, y'\}) = \{y\}$, violating Sen's α.

Now take alternatives x, y, z that are incomparable, where y has attributes that place it 'in between' x and z (for example, x is a very expensive, high quality shirt, z is very cheap but poor quality and y is a reasonable price and quality). The *compromise effect* captures the idea that including items with extreme attributes might increase the attractiveness of options with a moderate profile, so we can have $c(\{x, y\}) = \{x\}$ but $c(\{x, y, z\}) = \{y\}$, again violating Sen's α.

(iii) *Framing.* Even without altering the menu, the manner in which alternatives are presented can alter choice. Framing effects attack the notion that decisionmakers have stable, context-independent preferences.

Example 9 (Framing effects)**.**

(a) *Status quo effects.* Highlighting one option as a default option or the status quo in a menu can result in different choices. For example, opt-out organ donor registers tend to have much higher registration rates than opt-in donor registers.

(b) *Endowment effects.* In consumer theory, the menu is characterized by a budget set that depends only on the consumer's wealth (determined by their endowment) and prices. However, there is plenty of evidence that choices are not invariant to endowment – the price a decisionmaker is willing to sell an item they own can be much higher than the price they would be willing to pay for the same object if they did not own it.

(c) *Order effects.* The order in which alternatives are presented and considered can affect a decisionmaker's choices. We discuss this in more detail in section **??**.

(d) *Choice overload.* Increasing the number of alternatives in a menu can have a paralysing effect on choice. For example, a decisionmaker might be willing to buy one of three jams but faced with 33 jams, will prefer not to buy anything.

Unlike some, I would argue none of these examples are necessarily evidence of *irrationality* on the part of decisionmakers. Assuming that people are irrational if they don't follow our armchair assumptions about how they should behave is a sign of intellectual arrogance. Human decisionmaking is complicated and subject to many constraints that we do not typically model. This is easy to see in the case of indiscernibles – clearly, even if I have strictly monotone preferences regarding how much sugar is in my coffee, my taste receptors cannot perfectly measure sweetness. Even if I ran lab tests on every cup of coffee in Example 2, I could not perfectly estimate the concentration of glucose in each cup. But it is also plausible for the more cognitive phenomena, such as the decoy, compromise and framing effects. The way the brain values alternatives and makes decisions is very complicated – we are constantly learning and updating how we value alternatives in whatever situation we face. We would do well to note that the human brain achieves all the computational, memory and other tasks required for decisionmaking with about 12 watts of power (roughly the same as a dim light bulb), most of which is not even dedicated towards executive function. If the benchmark is an idealized unconstrained agent with fixed preferences, then of course humans will fall short.

This said, there are normative reasons in support of WARP/GARP. If you have a good idea of your preferences, then WARP might be a helpful guide to improving the consistency of your decisionmaking. The best normative argument for WARP is that if either $\succ$ or $\sim$ exhibit intransitivity, you can be made the victim of a money pump. Here it is natural to think of a consumer theory context, where the decisionmaker chooses from a budget set $B(p)$ that depends on vector of prices p. A money pump is a situation where you would voluntarily accept a sequence of trades that guarantee you lose money overall.

Definition 11 (Money pump)**.** Let $\succ$ be a strict preference relation over a set of bundles $X \subseteq \mathbb{R}^n_+$, let $B : \mathbb{R}^n_{++} \to \mathscr{M}(X)$ be a budget set correspondence, mapping each price vector p to a menu $B(p) \in \mathscr{M}(X)$. For a finite sequence of price vectors $\{p^n\}_{n=1}^m \subseteq$, the *money pump cost* associated with corresponding choices $\{x^n\}_{n=1}^m$ with $x^n, x^{n+1} \in c(B(p^n))$ for each n is defined by

$$\mathrm{mp}(\{p^n, x^n\}) = \sum_{n=1}^{m} p^n \cdot (x^n - x^{n+1}),$$

where $x^{m+1} := x^1$.

We say the sequence $\{p^n, x^n\}_{n=1}^m$ constitutes a *money pump* if $\mathrm{mp}(\{p^n, x^n\}) > 0$.

This has a nice story behind it. Suppose $n = 2$, and we have the WARP violation given in Example 5, so we have alternative bundles x, x', and both $x, x' \in B(p)$ and $x, x' \in B(p')$, but $x \in c(B(p))$ and $x' \in c(B(p'))$. Say you have purchased bundle x at prices p. If an arbitrageur knows your preferences, they can offer to buy bundle x and sell you bundle x' at prices p'. Given your preferences, you would be willing to accept this trade. But then the arbitrageur can offer to sell you back bundle x and buy back bundle x', this time at prices p. Again, you would be willing to accept such a trade. But if $p \cdot (x - x') + p' \cdot (x' - x) > 0$, then the arbitrageur has extracted positive profits from you, while leaving you in exactly the position you started with! Definition 11 extends this logic to GARP violations with longer sequences of trades.

There is a clear sense in which you would be being cheated if you were to fall victim to a money pump, although like other arbitrage opportunities, it is not obvious that in real life, arbitrageurs would be in a position to exploit such gaps if they are small. Echenique, Lee & Sum (2011) develop a *money pump index*,

$$\text{mpi}(\{p^n, x^n\}_{n=1}^m) = \frac{\text{mp}(\{p^n, x^n\})}{\sum_{n=1}^m p^n \cdot x^n},$$

to measure the magnitude of GARP violations. Using data from supermarket purchases, they find GARP violations are common, but such violations are small – and supermarkets lack the capacity to act as an arbitrageur in any case, since you cannot really resell your (usually perishable) goods to a supermarket. In a financial trade setting, being victim to a money pump is more plausible.

This said, I can think of phenomena where money-pump-type effects matter. In particular, "shrinkflation" and "skimpflation" are very real phenomena, whereby inflation manifests in the form of either reducing the size of a given product (reducing a chocolate bar from 150g to 140g, for instance) or reducing the quality of ingredients or production processes, rather than through price changes. *The Economist* goes as far as saying this wreaks "havoc with economists' models." Yet it is plausible that while you will probably notice a price change from \$1.99 to \$2.19, you are probably less likely to notice a slight reduction in the weight of a chocolate bar, or a slight reduction in the quality of the ingredients, particularly when you do not have the two objects side-by-side and when prices are extremely stable (for the extreme stability of prices, see e.g. Aparicio & Rigabon, 2023). Manufacturers are potentially taking advantage of the difficulty of distinguishing between similar objects here.

1.4 Separability

In many decision situations, the set of alternatives X available to a decisionmaker are multidimensional, i.e. it takes the form $X = \bigtimes_{i \in I} X_i$, where I is some index set. Section 4 dealt with a special case of this. In such situations, decisionmakers make choices across a number of different choice domains. An important question is whether decisionmaker's preferences are such that they make tradeoffs between choices within some choice subdomains independently of others. We will consider only the setting where the index set is countable, and most of the theorems are for finite index sets.

Definition 12 (Separable representations of preferences)**.** Let I be a countable index set, and consider the set of alternatives $X = \bigtimes_{i \in I} X_i$, a product set. Let $\succ$ be a weak order on X.

(a) *Ordinally separable representation.* We say $\succ$ has an *ordinally separable utility representation* $U : X \to \mathbb{R}$ if there are functions $u_i : X_i \to \mathbb{R}$ for each $i \in I$ and a coordinate-wise strictly increasing function $W : \bigtimes_{i \in I} u_i(X_i) \to \mathbb{R}$ such that

$$U(\{x_i\}_{i \in I}) = W(\{u_i(x_i)\}_{i \in I}).$$

(b) *Additively separable representation.* We say $\succ$ has an *additively separable utility representation* $U : X \to \mathbb{R}$ if there are functions $u_i : X_i \to \mathbb{R}$ for each $i \in I$ such that

$$U(\{x_i\}_{i\in I}) = \sum_{i\in I} u_i(x_i).$$

Clearly, if U is an additively separable utility representation of $\succ$, it is also an ordinally separable one, with $W(\{y_i\}_{i\in I}) := \sum_{i\in I} y_i$. Additive separability implies that how the decisionmaker makes tradeoffs between the components in any pair X_i, X_j is independent of how the decisionmaker makes tradeoffs in $X_{-i,-j} := \bigtimes_{k\neq i,j} X_k$. Example 11 illustrates this in the context of a dynamic consumption problem.

Ordinal separability is much weaker than additive separability, and we might ask why such an apparently weak concept is useful. If each X_i is, say, a subset of $\mathbb{R}$, then it is true that ordinal separability does not buy us much. However, if the X_is are multidimensional, then ordinal separability is more intuitive: it implies that the tradeoffs between components within each X_i are made independently from how the decisionmaker makes tradeoffs outside of X_i.

Example 10.

(a) *Dynamic choice.* In discrete time dynamic choice settings, we have either a finite or infinite horizon, and decisionmakers make a choice x_t for each time period t. An additively separable utility representation in this case takes the form

$$U(x) = \sum_{t=0}^{T} \delta^t u_t(x_t), \qquad \text{or} \qquad U(x) = \sum_{t=0}^{\infty} \delta^t u_t(x_t),$$

depending on whether the horizon is finite (with terminal period $T \in \mathbb{N}$) or infinite. In the infinite horizon case, $\delta \in (0, 1)$. Typically we also impose that $u_t = u$ for all t, i.e. the utility does not depend on the period.

(b) *Uncertainty.* Consider a model of uncertainty, with some countable state space S. A decisionmaker makes choices that yield consequence x_s in each state s. An additively separable utility representation in this setting takes the form

$$U(x) = \sum_{s\in S} \mu(s) u_s(x_s),$$

where μ is a probability measure on S. Typically, we also impose that $u_s = u$ for all s, i.e. the utility does not depend on state.

(c) *Quasilinear utility representations.* Suppose a decisionmaker has preferences $\succ$ on set of alternatives $X = \bigtimes_{i=1}^{n} X_i$. If $X_1 \subseteq \mathbb{R}$ and if there is a utility function $U : X \to \mathbb{R}$ representing $\succ$ that takes the form

$$U(x_1, \ldots, x_n) = x_1 + v(x_2 \ldots, x_n),$$

for some function $v : \bigtimes_{i=2}^{n} X_i \to \mathbb{R}$, then we say $\succ$ is *quasilinear* in x_1. If we cannot write v in the form $v(x_2, \dots, x_n) = \sum_{i=2}^{n} w_i(x_i)$, then this is an example of an ordinally separable utility representation that is not additively separable. Quasilinear preferences are a common assumption in many applications, particularly in mechanism design.

Example 11 (Dynamic consumption choice)**.** In a simple finite horizon discrete time consumption choice problem, a decisionmaker faces the problem of choosing a stream of consumption $\{c_t\}_{t=0}^{T}$, subject to a budget constraint $\sum_{t=0}^{T} p_t c_t \leq w$, where w is her wealth and p_t is the price of the period t consumption good. If the decisionmaker's preferences have a utility representation

$$U(c_1, \dots, c_t) = \sum_{t=0}^{T} \delta^t u(c_t),$$

where u is a strictly increasing, concave function, then the necessary and sufficient conditions for utility maximization are the *Euler condition*

$$\frac{u'(c_t)}{\delta^{t'-t} u'(c_{t'})} = \frac{p_t}{p_{t'}}$$

for each $t = 0, \dots, T-1$ and $t' > t$, along with terminal condition

$$c_T = \frac{w - \sum_{t=0}^{T-1} p_t c_t}{p_T}.$$

Consumption in periods other than t, t' do not enter into the Euler condition that captures the tradeoff between consumption in periods t, t'. This is the consequence of the additive separability of the decisionmakers preferences.

There are two main approaches to separability. One approach relies on topological notions, and the other approach is algebraic, relying on group theory. We survey the background mathematics for both approaches in the mathematical appendix. Which approach is more convenient really depends on the application – sometimes topology can get in the way, sometimes it makes proving things very neat.

1.4.1 Conditions for ordinal separability

Under what conditions will a weak order $\succ$ have an ordinally separable representation? Debreu (1960) gives an answer under some topological assumptions on the set of alternatives X. There is also an algebraic approach (Wakker, 1988) that is slightly more general.

Assume the index set I is finite. For each pair of alternatives $x, y \in X$ and each set $E \subseteq I$, define the alternative $x_E y \in X$ by

$$(x_E y)_i := \begin{cases} x_i & \text{if } i \in E, \\ y_i & \text{if } i \notin E. \end{cases}$$

This takes components from x for indices in E and components from y for all the other indices. For brevity, we write $x_i y$ in place of $x_{\{i\}} y$.

Definition 13. Let $\succ$ be a weak order on set of alternatives $X = \bigtimes_{i \in I} X_i$.

(a) *Separable set.* We call a subset $E \subseteq I$ *separable* if for all $x, y, a, b \in X$, $x_E a \succ y_E a$ iff $x_E b \succ y_E b$.

(b) *Separable index.* We call an index $i \in I$ *separable* if $\{i\}$ is a separable set.

(c) *Null indices.* We call an index i *null* if for all $x, y, z \in X$, $x_i z \sim y_i z$.

An index i is null if the decisionmaker does not care about the component corresponding to that index at all. Separability of a set E is not quite as easy to explain in words. Take an alternative, and modify it in two different ways by replacing only the components indexed by E, generating a first option and a second option. Then separability says that if we took any other alternative and modified them in the same way, the decisionmaker would have the same preferences over the first and second options in both cases. For separability of an index i, we modify only the ith component. For example, say I will eat a fruit for breakfast for the next three days. Then we can think of the set of alternatives as bundles of fruit, indexed by time $I = \{1, 2, 3\}$. Say for each fruit bundle, I consider swapping day 2's fruit to either an apple or a banana. If day 2 is separable and for some fruit bundle, I prefer the banana-swapped version to the apple-swapped version, then I must prefer banana-swapped versions to apple-swapped versions for all the bundles. This rules out that, say, a love of variety means that I would prefer to have a banana on day 2 if I had apples every other day and would prefer to have an apple on day 2 if I had bananas every other day.

Axioms. Let I be an index set, let $X = \bigtimes_{i \in I} X_i$ be a set of alternatives and let $\succ$ be a weak order on X.

(**S1**) *Singleton separability.* Each index $i \in I$ is separable.

Moreover, assume each X_i is a connected and separable topological space, and X is equipped with the product topology.

(**S2**) *Continuity.* For all $x \in X$, the sets $\{y \in X \mid y \succ x\}$ and $\{y \in X \mid x \succ y\}$ are open in X.

Note (**S2**) is equivalent to (**C1**) (it is a restatement of (**C1′**), which we proved was equivalent in the proof of Theorem 4). Hence (**S2**) states that $\succ$ are continuous preferences.

Theorem 8 (Debreu, 1960)**.** *Suppose $\succ$ is a weak order on product set of alternatives $X = \bigtimes_{i \in I} X_i$, where I is a finite index set and each X_i is a connected and separable topological space. Then $\succ$ satisfies* (**S1**) *and* (**S2**) *iff it has an ordinally separable utility representation $U = W(\{u_i\}_{i \in I})$, where W and each u_i are continuous functions.*

Proof. Omitted. □

1.4.2 Conditions for additive separability

Now we'll turn to the conditions under which $\succ$ has an additively separable representation (generally a more useful notion). In this case we need something stronger than singleton separability (**S1**), which simply tells us that tradeoffs within X_i do not depend on what is happening elsewhere. For additive separability, we need that tradeoffs *between* X_i and X_j do not depend on what is happening elsewhere. Rather, we will need to consider separability of all index subsets:

Axiom. Let I be an index set, let $X = \bigtimes_{i \in I} X_i$ be a set of alternatives and let $\succ$ be a weak order on X.

(**S3**) *Separability.* Each subset $E \subseteq I$ is separable.

However, we need to build up to the general case. First, let's consider the case of 2 indices. For this, we will use another axiom, originally due to Thomsen.

Axiom. Let $X = X_1 \times X_2$ and let $\succ$ be a weak order on X.

(**S4**) *Thomsen's condition.* For all $x, y, z \in X$, if $(x_1, y_2) \sim (y_1, x_2)$ and $(y_1, z_2) \sim (z_1, y_2)$, then $(x_1, z_2) \sim (z_1, x_2)$.

Theorem 9 (Debreu, 1960)**.** *Let $\succ$ be a weak order on set of alternatives $X = X_1 \times X_2$, where each X_i is a connected and separable topological space and X is equipped with the product topology. Then $\succ$ satisfies continuity* (**S2**) *and Thomsen's condition* (**S4**) *iff there are continuous utility functions $u_i : X_i \to \mathbb{R}$ such that*

(i) *u_i is constant if i is null;*

(ii) *$U(x) := u_1(x_1) + u_2(x_2)$ is an additively separable utility representation of $\succ$.*

Moreover, if $V(x) = v_1(x_1) + v_2(x_2)$ is an additively separable utility representation of $\succ$, then there are real numbers $\alpha > 0$ and $\beta_1, \beta_2 \in \mathbb{R}$ such that $v_i = \alpha u_i + \beta_i$ for $i = 1, 2$.

Proof. Omitted. □

Theorem 9 can then be used to prove the more general case:

Theorem 10 (Debreu, 1960)**.** *Let $\succ$ be a weak order on set of alternatives $X = \bigtimes_{i \in I} X_i$ where I is a finite index set, each X_i is a connected and separable topological space, and X is equipped with the product topology. Suppose there are at least three null indices. Then $\succ$ satisfies continuity* **S2** *and separability* (**S3**) *iff there are continuous utility functions $u_i : X_i \to \mathbb{R}$ such that*

(i) *u_i is constant if i is null;*

(ii) *$U(x) = \sum_{i \in I} u_i(x_i)$ is an additively separable utility representation of $\succ$.*

Moreover, if $V(x) = v_1(x_1) + v_2(x_2)$ is an additively separable utility representation of $\succ$, then there are real numbers $\alpha > 0$ and $\{\beta_i\}_{i \in I}$ such that $v_i = \alpha u_i + \beta_i$ for all $i \in I$.

Proof. Omitted. □

2 Decisions under uncertainty

Decision theory really takes off once we admit the world is uncertain: decisionmakers frequently have to make decisions not knowing for sure what the consequences of their actions are. While decisionmaking under certainty opened up some interesting questions, such as issues around what it means to be indifferent, decisionmaking under uncertainty opens up more interesting issues.

The main issue is how to go about modelling uncertainty in the first place. There are two interesting dimensions to this. One is which *uncertainty domain* we choose to use, and the other is our stance on what probabilities represent.

One approach – the most restrictive – is to treat uncertainty as *objective*. This was the approach of von Neumann and Morgenstern (1947). In this setting, there is a set of *consequences* (or *prizes*) and the probability over consequences is known and external to the decisionmaker. The basic object we are concerned with, defined by the probability measure over consequences, is a *lottery*. For example, a fair coin, by definition, turns up heads with probability $\frac{1}{2}$, so if I play a game where a fair coin is flipped and I win \$1 if it turns up heads and lose \$1 if it turns up tails, then I can model this as a lottery. This interpretation is somewhat intuitive in settings where uncertainty is easily quantifiable, such as actual lotteries, games of chance such as blackjack, and other settings where the prizes and odds are known.[14] The domain of uncertainty here is one of *risk* – acts (i.e. choosing a lottery) translate directly into a given distribution over consequences.

Treating uncertainty as objective has two problems. Restricting to lotteries with known distributions is very restrictive, because most forms of uncertainty cannot be easily quantified. Suppose you are about to travel from sunny, beautiful northern California to dank, grim New York. You are choosing whether to use the subway or not. Your decision might depend on your assessment of how safe the subway is. This assessment is subjective – it will be influenced by what you have read in the news about the subway, the advice of friends with experience of New York, your own experiences, preexisting stereotypes, and so on. Probability here serves as a quantification of *personal* uncertainty, i.e. uncertainty is taken to be *subjective*. This motivates Savage's (1954) subjective expected utility approach, which is an elegant, complete theory of Bayesianism. Savage takes as basic a set of states of the world and a set of consequences. Savage acts map states into consequences. The domain of uncertainty here is thus quite different from the risk domain of von Neumann & Morgenstern – rather than uncertainty being loaded onto consequences, it is loaded onto states. From this basic setting, Savage builds up his theory from a set of normatively quite attractive (but empirically falsified) axioms. These axioms imply that the decisionmaker's beliefs about the probability of events in the state space occurring can be separated from the decisionmaker's tastes over lotteries of consequences.

[14] Of course, even in such settings, there is some subjective component. In blackjack, you might suspect that the dealer is cheating in some way, or when it comes to flipping the coin you might believe the coin is not fair. These suspicions would alter how you assess the lottery. Anscombe and Aumann (1963) make the point that even the "objective" probabilities related to events like coin tosses, roulette wheel spins and so forth are subjective – it is just that everyone agrees on the (subjective) probabilities.

The way Savage went about proving the representation theory for his approach is (needlessly) complicated, which has historically prevented his theory from being modified in interesting ways. Rather, developments have often followed the flexible extraneous probabilities approach introduced by Anscombe & Aumann (1963), and simplified by Fishburn (1970). This augments subjective probabilities with extraneous probabilities associated with the consequences of "objective" devices such as fair coins, roulette wheels, or fair dice. This setting is more general than Savage's, because Anscombe-Aumann acts map states into lotteries, whereas Savage acts map into consequences.[15] Hence Savage acts can be viewed as special cases of Anscombe-Aumann acts where the lotteries are degenerate. The Anscombe-Aumann framework has proven a very popular setting for interesting extensions to expected utility theory, such as for models of ambiguity aversion.

2.1 Risk and expected utility theory

We start with the von Neumann-Morgenstern (1947) approach to expected utility, further developed by Friedman & Savage (1948, 1952), Marschak (1950), Herstein & Milnor (1953), Blackwell & Girshik (1954), and others. In this approach, the primitives are a measurable space of *consequences* X, the set $\Delta(X)$ of probability measures on X, and a binary relation $\succ$ on $\Delta(X)$. Hence $\Delta(X)$ is the set of alternatives here. In the context of von Neumann-Morgenstern expected utility theory, we refer to the measures $\mu \in \Delta(X)$ as (simple) *lotteries* over X.

Note that $\Delta(X)$ is convex (Proposition **??**). This is convenient since we can consider compound lotteries, i.e. convex combinations $\alpha\mu_1 + (1-\alpha)\mu_2$ where $\alpha \in [0,1]$ and $\mu_1, \mu_2 \in \Delta(X)$. This is a "compound" lottery since it has the interpretation of a two stage lottery: first, one performs a Bernoulli experiment with probability $\alpha \in [0,1]$ of selecting the lottery μ_1 and complementary probability of selecting lottery μ_2; second, the selected lottery is performed. Since such compound lotteries lie in $\Delta(X)$, any compound lottery is equivalent to some simple lottery over X.

Example 12 (Simple and compound lotteries)**.** Suppose you are at a charity event and the charity is holding a lottery. The charity is not very imaginative and instead of interesting prizes, the prizes they offer are different dollar amounts of cash. They offer two different lotteries, μ_1 and μ_2. Lottery μ_1 has the following story: you can buy a ticket for \$1 and the prize if you win is \$99. There are 100 tickets for sale, so $\mu_1(98) = \frac{1}{100}$ and $\mu_1(-1) = \frac{99}{100}$. Lottery μ_2, on the other hand, has two prizes: there are again 100 tickets and one will win \$80 and a second will win \$19, and tickets again cost \$1, so $\mu_1(18) = \mu_1(79) = \frac{1}{100}$ and $\mu_1(-1) = \frac{49}{50}$.

You are complaining to your friend that the two lotteries seem pretty boring. You were hoping the prizes might be a nice bottle of wine/equally attractive non-alcoholic

[15]This is the Fishburn (1970) characterization. The original Anscombe-Aumann primitive is in fact lotteries over Savage acts. Similarly, our presentation of the von Neumann-Morgenstern is a simplification compared to their original setting, in which they explicitly modelled compound lotteries through decision trees – an unnecessary complication if we assume decisionmakers understand Bayes' rule.

beverage, or maybe a concert ticket or something. Your friend proposes to make things a bit more exciting: you pay them \$1, they will flip a fair coin ($\alpha = \frac{1}{2}$), and if it lands heads they will buy the first ticket (corresponding to lottery μ_1) and if it lands tails they buy the other ticket (i.e. μ_2), and they'll give you all the winnings if they win. Let μ_3 be this lottery. Now $\mu_3(98) = \mu_3(79) = \mu_3(18) = \frac{1}{200}$ and $\mu_3(-1) = \frac{197}{200}$. You complain that this isn't any more exciting at all: if the charity had originally offered 200 tickets for a \$1 each with 3 tickets paying out \$99, \$80 and \$19 and the rest losing, then you would be facing exactly the same problem.[16] The two-stage process is a compound lottery, and we see it reduces to a simple lottery.

2.1.1 Mixture sets and the von Neumann-Morgenstern axioms

The classic approach to von Neumann-Morgenstern expected utility theory relies heavily on the concept of *mixture sets* (also called *mixture spaces*), developed by Herstein & Milnor (1953).[17]

Definition 14 (Mixture set)**.** A *mixture set* is a set $\mathcal{M}$ endowed with an operation $\oplus : \mathcal{M} \times \mathcal{M} \times [0,1] \to \mathcal{M}$, assigning to each triple (μ_1, μ_2, α) with $\mu_1, \mu_2 \in \mathcal{M}$ and $\alpha \in [0,1]$ an element $\mu_1 \oplus_\alpha \mu_2 \in \mathcal{M}$, satisfying

(i) $\mu_1 \oplus_1 \mu_2 = \mu_1$,

(ii) $\mu_1 \oplus_\alpha \mu_2 = \mu_2 \oplus_{1-\alpha} \mu_1$, and

(iii) $(\mu_1 \oplus_\alpha \mu_2) \oplus_\beta \mu_2 = \mu_1 \oplus_{\beta\alpha} \mu_2$,

for all $\mu_1, \mu_2 \in \mathcal{M}$ and all $\alpha, \beta \in [0,1]$.

Proposition 10. *In Definition 14,* (i)-(iii) *imply:*

(iv) $\mu_1 \oplus_0 \mu_2 = \mu_2$, *and*

(v) $(\mu_1 \oplus_\alpha \mu_2) \oplus_\beta (\mu_1 \oplus_\gamma \mu_2) = \mu_1 \oplus_{\beta\alpha+(1-\beta)\gamma} \mu_2$

for all $\mu_1, \mu_2 \in \mathcal{M}$ *and* $\alpha, \beta, \gamma \in [0,1]$.

Proof. (iv) follows since $\mu_1 \oplus_0 \mu_2 = \mu_2 \oplus_1 \mu_1 = \mu_2$, where the first equality is by (ii) and the second is by (i). (v) follows from (i) and (ii): take $\alpha, \gamma \in (0,1)$ and suppose wlog that $\alpha \leq \gamma$. Then

$$\begin{aligned}
\mu_1 \oplus_{\alpha\beta+(1-\beta)\gamma} \mu_2 &= \mu_1 \oplus_{(\beta\alpha/\gamma+1-\beta)\gamma} \mu_2 \\
&= (\mu_1 \oplus_\gamma \mu_2) \oplus_{\beta\alpha/\gamma+1-\beta} \mu_2 \\
&= \mu_2 \oplus_{\beta(1-\alpha/\gamma)} (\mu_1 \oplus_\gamma \mu_2) \\
&= (\mu_2 \oplus_{1-\alpha/\gamma} [\mu_1 \oplus_\gamma \mu_2]) \oplus_\beta (\mu_1 \oplus_\gamma \mu_2) \\
&= ([\mu_1 \oplus_\gamma \mu_2] \oplus_{\alpha/\gamma} \mu_2) \oplus_\beta (\mu_1 \oplus_\gamma \mu_2) \\
&= (\mu_1 \oplus \alpha\mu_2) \oplus_\beta (\mu_1 \oplus_\gamma \mu_2),
\end{aligned}$$

[16] Fishburn (1970) is like the friend – he mentions that while the two lotteries are exactly the same, there might be psychologically different

[17] For a deeper overview of the properties of mixture sets, see Mongin (2001).

as desired. The second, fourth and final equalities are by (iii) and the third and fifth equalities follow by (ii). □

Proposition 10 implies (i) can be replaced with the condition that $\mu \oplus_\alpha \mu = \mu$ for all $\mu \in \mathcal{M}$ and all $\alpha \in [0,1]$ in Definition 14.

Example 13 (Convex sets are mixture sets)**.** Take C to be any convex set in a vector space. We claim C is a mixture set when endowed with the operation $\mu_1 \oplus_\alpha \mu_2 = \alpha\mu_1 + (1-\alpha)\mu_2$, i.e. $\mu_1 \oplus_\alpha \mu_2$ is a convex combination of $\mu_1, \mu_2 \in C$ with weight $\alpha \in [0,1]$. Clearly, $\mu_1 \oplus_\alpha \mu_2 \in C$ since C is convex. Note (i) is obviously satisfied, (ii) is satisfied since $\mu_1 \oplus_\alpha \mu_2 = \alpha\mu_1 + (1-\alpha)\mu_2 = (1-\alpha)\mu_2 + \alpha\mu_1 = \mu_2 \oplus_{1-\alpha} \mu_1$, and (iii) is satisfied since $(\mu_1 \oplus_\alpha \mu_2) \oplus_\beta \mu_2 = \beta(\alpha\mu_1 + (1-\alpha)\mu_2) + (1-\beta)\mu_2 = \beta\alpha\mu_1 + [\beta - \beta\alpha + 1 - \beta]\mu_2 = \beta\alpha\mu_1 + (1-\beta\alpha)\mu_2 = \mu_1 \oplus_{\beta\alpha} \mu_2$.

There are several useful special cases to note for expected utility theory:

(a) *Spaces of probability measures.* If X is an arbitrary measurable space and $\Delta(X)$ is the space of all probability measures on X, then $\Delta(X)$ is a mixture set when endowed with the operation $\mu_1 \oplus_\alpha \mu_2 = \alpha\mu_1 + (1-\alpha)\mu_2$. This follows since $\Delta(X)$ is convex (Proposition **??**).

(b) *Spaces of simple probability measures.* Let X be an arbitrary measurable space, and call a probability measure μ on X *simple* if μ has finite support. Let $\Delta_S(X)$ be the space of simple probability measures on X. We claim $\Delta_S(X)$ is convex, and hence a mixture set when endowed with the convex combination operation. To see this, note that if $\mu_1, \mu_2 \in \Delta_S(X)$, then $\alpha\mu_1 + (1-\alpha)\mu_2$ is a probability measure (by convexity of $\Delta(X)$) and the support of $\alpha\mu_1 + (1-\alpha)\mu_2$ lies in the union of the supports of μ_1 and μ_2, which is finite, so $\alpha\mu_1 + (1-\alpha)\mu_2 \in \Delta_S(X)$.

Whenever dealing with convex sets that are mixture sets, the convex combination operation is the natural mixture operation so we will take it as obvious that $\mu_1 \oplus_\alpha \mu_2 = \alpha\mu_1 + (1-\alpha)\mu_2$ for all μ_1, μ_2 in the set and all $\alpha \in [0,1]$.

I have used the notation $\oplus$ for the mixture set operation to emphasize that the operation in general does not need to be a convex combination operation.[18] Most expositions (including Herstein & Milnor, 1953) jump straight to denoting the mixture set operation as a convex combination $\alpha\mu_1 + (1-\alpha)\mu_2$. However, there are plenty of counterexamples where mixture sets are *not* convex sets. We give one such example:

Example 14 (Mongin, 2001)**.** Let $\mathcal{M} = \{x, y, z\}$ for three distinct elements x, y, z. Endow $\mathcal{M}$ with the operation $\oplus$ so that: for all $\alpha \in [0,1]$, $y \oplus_\alpha y = y$ and $z \oplus_\alpha z = z$; $x \oplus_1 y = y \oplus_0 x = z \oplus_0 y = y \oplus_1 z = y$ and $x \oplus_1 z = z \oplus_0 x = y \oplus_0 z = z \oplus_1 y = z$, and $a \oplus b = x$ for $a, b \in \mathcal{M}$ otherwise.

Then $\mathcal{M}$ is a mixture set, but it is clearly not convex.

[18]Kreps (1988) represented the operation via a family of functions h_α to make the point that this operation is abstract, though he dropped this in Kreps (2013).

Definition 15 (Mixture-preserving function)**.** Given a mixture set $\mathcal{M}$, we say that a function $f : \mathcal{M} \to \mathbb{R}^n$ is *mixture-preserving* if for every $\mu_1, \mu_2 \in \mathcal{M}$ and every $\alpha \in [0,1]$, we have

$$u(\mu_1 \oplus_\alpha \mu_2) = \alpha f(\mu_1) + (1-\alpha) f(\mu_2).$$

An important example of a mixture-preserving function is the expectation operator:

Example 15. Given an arbitrary set X, fix a function $f : X \to \mathbb{R}$. Define the functional $T : \Delta(X) \to \mathbb{R}$ by

$$T(\mu) := \int_X f \,\mathrm{d}\mu$$

for each probability measure $\mu \in \Delta(X)$. Then T is a mixture-preserving function, since

$$\begin{aligned} T(\alpha\mu_1 + [1-\alpha]\mu_2) &= \int_X f \,\mathrm{d}(\alpha\mu_1 + [1-\alpha]\mu_2) \\ &= \alpha \int_X f \,\mathrm{d}\mu_1 + (1-\alpha) \int_X f \,\mathrm{d}\mu_2 = \alpha T(\mu_1) + (1-\alpha) T(\mu_2). \end{aligned}$$

The following extends a result from Coulhon & Mongin (1989):[19]

Proposition 11. *Let $\mathcal{M}$ be a convex subset of a vector space and consider a function $f : \mathcal{M} \to \mathbb{R}^n$. Then f is mixture-preserving iff f is affine on $\mathcal{M}$, i.e. iff there is some linear form $\varphi : \mathcal{M} \to \mathbb{R}^n$ and some constant vector $c \in \mathbb{R}^n$ such that $f(x) = \varphi(x) + c$ for all $x \in \mathcal{M}$.*

Proof. First suppose f is mixture-preserving, i.e. for all $\mu_1, \mu_2 \in \mathcal{M}$ and all $\alpha \in [0,1]$, we have $f(\alpha\mu_1 + (1-\alpha)\mu_2) = \alpha f(\mu_1) + (1-\alpha) f(\mu_2)$.

Now, define $g(\mu_1 - \mu_2) = f(\mu_1) - f(\mu_2)$ for $\mu_1, \mu_2 \in \mathcal{M}$. Define $E = \{\mu_1 - \mu_2 \mid \mu_1, \mu_2 \in \mathcal{M}\}$. Extend g to $\operatorname{span}\mathcal{M}$ by fixing a basis $e_1, \ldots, e_n$ of $\operatorname{span}\mathcal{M}$ such that each $e_i \in E$ and defining $g(x) = \sum_{i=1}^n \lambda_i g(e_i)$ for each $x = \sum_{i=1}^n \lambda_i e_i \notin E$. To show f is affine, we need only show g is linear. We claim $g(\alpha x + (1-\alpha)y) = \alpha g(x) + (1-\alpha) g(y)$ for all $x, y \in V$ and $\alpha \in [0,1]$. Now $g(\alpha(\mu_1 - \mu_1') + (1-\alpha)(\mu_2 - \mu_2')) = f(\alpha\mu_1 + (1-\alpha)\mu_2) - f(\alpha\mu_1' + (1-\alpha)\mu_2') = \alpha[f(\mu_1) - f(\mu_1')] + (1-\alpha)[f(\mu_2) - f(\mu_2')] = \alpha g(\mu_1 - \mu_1') + (1-\alpha) g(\mu_2 - \mu_2')$, so $g(\alpha x + (1-\alpha)y) = \alpha g(x) + (1-\alpha) g(y)$ for all $x, y \in E$. Since we defined the extension of g to $\operatorname{span}\mathcal{M}$ to ensure linearity on $(\operatorname{span}\mathcal{M}) - E$, it follows that this also holds for all $x, y \in \operatorname{span}\mathcal{M}$.

Lemma 6. *If V is a vector space and $g : V \to \mathbb{R}^n$ is a function such that*

(i) *$g(\alpha x + (1-\alpha)y) = \alpha g(x) + (1-\alpha) g(y)$ for all $\alpha \in [0,1]$ and $x, y \in C$, and*

[19] The result has been known for far longer, and was first stated less rigorously by Harsanyi (1955) and with an incorrect proof by Harsanyi (1977). But do not despair – there are plenty of correct proofs! For example, Domotor (1979), Fishburn (1984), Border (1985), and so on. The result is interesting in social choice theory where it implies that any von Neumann-Morgenstern social utility function satisfying Pareto indifference is affine in individual's von Neumann-Morgenstern utility functions.

(ii) $g(0) = 0$,

then g is linear, i.e. $g(x + y) = g(x) + g(y)$ and $g(\alpha x) = \alpha g(x)$ for all $\alpha \in \mathbb{R}$ and all $x, y \in V$.

Proof. Suppose (i) and (ii) hold. Taking $y = 0$, we have that $f(\alpha x) = f(\alpha x + (1 - \alpha)0) = \alpha f(x) + (1 - \alpha)f(0) = \alpha f(x)$. Next, take any $x, y \in V$. Now, we can write $x + y = \alpha\hat{x} + (1-\alpha)\hat{y}$ by fixing $\alpha \in (0, 1)$ and defining $\hat{x} = x/\alpha$ and $\hat{y} = y/(1-\alpha)$. Then $f(x+y) = f(\alpha\hat{x} + (1-\alpha)\hat{y}) = \alpha f(\hat{x}) + (1-\alpha)f(\hat{y}) = f(x) + f(y)$. Thus f is linear. $\square$

Now, g satisfies the hypotheses of Lemma 6 and so g is linear. Hence f is affine.

Conversely, suppose f is affine, i.e. $f(x) = \varphi(x) + c$ for some linear form φ. Now, $f(\alpha\mu_1 + (1 - \alpha)\mu_2) = \varphi(\alpha\mu_1 + (1 - \alpha)\mu_2) + c = \alpha[\varphi(\mu_1) + c] + (1 - \alpha)[\varphi(\mu_2) + c] = \alpha f(\mu_1) + (1 - \alpha)f(\mu_2)$. $\square$

Let's relate this back to our actual goal: describing decisionmaking under risk. As we noted in Example 13, spaces of probability distributions and convex subsets of such spaces are mixture sets. A simple *lottery* over a space of consequences X is a probability measure $\mu \in \Delta(X)$, and thus the set of simple lotteries over X is a mixture space. Because the extra generality has no cost at all, let's take any arbitrary mixture space $\mathcal{M}$ as basic (in the von Neumann-Morgenstern setting, this will always be a space of lotteries). As usual, we take as basic a strict preference relation $\succ$ over $\mathcal{M}$, which we assume is a weak order.[20] Von Neumann & Morgenstern (1947) propose allegedly reasonable axioms for $\succ$ to satisfy:

Axioms (von Neumann-Morgenstern axioms)**.** Let $\mathcal{M}$ be a mixture set and let $\succ$ be a weak order over $\mathcal{M}$.

(V1) *Independence of irrelevant alternatives.* For all $\mu_1, \mu_2, \mu_3 \in \mathcal{M}$, we have $\mu_1 \succ \mu_2$ iff $\mu_1 \oplus_\alpha \mu_3 \succ \mu_2 \oplus_\alpha \mu_3$ for all $\alpha \in (0, 1]$.

(V2) *Archimedean continuity.* If $\mu_1, \mu_2, \mu_3 \in \mathcal{M}$ with $\mu_1 \succ \mu_2$ and $\mu_2 \succ \mu_3$, then $\mu_1 \oplus_\alpha \mu_3 \succ \mu_2$ and $\mu_2 \succ \mu_1 \oplus_\beta \mu_3$ for some $\alpha, \beta \in (0, 1)$.

These axioms only really make sense when applied to lotteries (expected utility, after all, is a question of *expectation* which requires probability measures). As usual, $\succ$ being a weak order implies that the induced indifference relation $\sim$ is transitive. Again, it is pretty plausible that indifference need not be transitive. Suppose lotteries μ_1 and μ_2 assign $\mu_1(35) = 1$ and $\mu_2(36) = 1$, while $\mu_3(0) = \mu_3(80) = \frac{1}{2}$. Then it seems obvious that any reasonable decisionmaker's preferences are such that $\mu_2 \succ \mu_1$. Yet it is plausible that the decisionmaker would be indifferent between each of these lotteries and μ_3: $\mu_1 \sim \mu_3$ and $\mu_2 \sim \mu_3$, which would violate transitivity of $\sim$. This is plausible because we are rarely have very precise preferences about lotteries, just as we rarely have very precise preferences about other continuous (or very fine) alternatives (as in Example 2).

[20] Recall that binary relation $\succ$ is a weak order if it is asymmetric and negatively transitive.

The axiom (**V1**), *independence of irrelevant alternatives* is "the core of expected-utility theory," as Fishburn (1970) puts it. It says that if we have any two lotteries μ_1, μ_2 and the decisionmaker strictly prefers μ_1 to μ_2, then the decisionmaker should also prefer any mixture $\alpha\mu_1 + (1-\alpha)\mu_3$ of μ_1 with a third lottery μ_3 over the mixture $\alpha\mu_2 + (1-\alpha)\mu_3$ of μ_2 with μ_3, for $\alpha > 0$. Recall that these mixtures $\alpha\mu_1 + (1-\alpha)\mu_3$ have two possible interpretations. One is as a simple lottery where $\alpha\mu_1 + (1-\alpha)\mu_3$ is just another distribution over consequences. The other is as a compound lottery where μ_1 is chosen with probability α and μ_3 with probability $1-\alpha$. The latter interpretation justifies the axiom – consider the payoff matrix:

$$\begin{array}{ccc} & \alpha & 1-\alpha \\ A & \mu_1 & \mu_3 \\ B & \mu_2 & \mu_3 \end{array}$$

The decisionmaker can choose option A or option B, and after making her choice, a coin is flipped and selects the left column with probability α and the right with probability $1-\alpha$. If $\mu_1 \succ \mu_2$, then A weakly dominates B, so we would reasonably expect that $\alpha\mu_1 + (1-\alpha)\mu_3 \succ \alpha\mu_2 + (1-\alpha)\mu_3$.

The *Archimedean continuity* axiom (**V2**) ensures that if $\mu_1 \succ \mu_2 \succ \mu_3$, then there is are nontrivial mixtures of μ_1 and μ_3 that are preferred to μ_2 and that μ_2 is preferred to. To motivate this, consider the following version of Russian roulette. There are n revolvers (each has 6 chambers) and precisely one of the chambers in precisely one revolver is loaded. You are offered one of the revolvers uniformly at random, and if you put the gun to your head and pull the trigger, you will be paid \$10 unless you die. The other option is not to play, in which case you do not get anything. Now, the probability of dying if you pull the trigger is $\frac{1}{6n}$. You prefer the \$10 over receiving nothing and you prefer to receive nothing over death. The continuity axiom says that as n grows larger (so the probability of death becomes ever smaller), you eventually would opt to play the game for the \$10. If there are a trillion revolvers, then it is plausible you would opt to take the bet. Now, you might complain that you would not play under any circumstances, but in practice, you take calculated risks about your life all the time – crossing a busy road, for instance, always carries some risk that a reckless speeding driver might hit you.

Before we state the mixture set theorem – the main result – we first need an important (if somewhat lengthy) lemma.

Lemma 7. *Suppose $\mathcal{M}$ is a mixture set and that $\succ$ is a weak order on $\mathcal{M}$ satisfying* (**V1**) *and* (**V2**). *Then for all $\mu_1, \mu_2, \mu_3, \mu_4 \in \mathcal{M}$:*

(a) *if $\mu_1 \succ \mu_2$ then $\mu_1 \oplus_\beta \mu_2 \succ \mu_1 \oplus_\alpha \mu_2$ for all $0 \leq \alpha < \beta \leq 1$;*

(b) *if $\mu_1 \succsim \mu_2$, $\mu_2 \succsim \mu_3$ and $\mu_1 \succ \mu_3$ then $\mu_2 \sim \mu_1 \oplus_\alpha \mu_3$ for precisely one $\alpha \in [0,1]$;*

(c) *if $\mu_1 \succ \mu_2$ and $\mu_3 \succ \mu_4$ then $\mu_1 \oplus_\alpha \mu_3 \succ \mu_2 \oplus_\alpha \mu_4$ for all $\alpha \in [0,1]$;*

(d) *if $\mu_1 \sim \mu_2$ then $\mu_1 \oplus_\alpha \mu_2 \sim \mu_1$ for all $\alpha \in [0,1]$, and*

(e) *if* $\mu_1 \sim \mu_2$ *then* $\mu_1 \oplus_\alpha \mu_3 \sim \mu_2 \oplus_\alpha \mu_3$ *for all* $\alpha \in [0,1]$.

Proof. (a) If $\beta < 1$ then (**V1**) and Proposition 10(iv) imply $\mu_1 \oplus_\beta \mu_2 \succ \mu_2 \oplus_\beta \mu_2 = \mu_2$, and if $\beta = 1$ then $\mu_1 \oplus_\beta \mu_2 = \mu_1 \succ \mu_2$. Now, if $\alpha > 0$ then

$$\mu_1 \oplus_\beta \mu_2 = (\mu_1 \oplus_\beta \mu_2) \oplus_{\alpha/\beta} (\mu_1 \oplus_\beta \mu_2) \succ (\mu_1 \oplus_\beta \mu_2) \oplus_{\alpha/\beta} \mu_2 = \mu_1 \oplus_\alpha \mu_2$$

by (**V1**), Proposition 10(iv) and Definition 14(iii). If $\alpha = 0$ then $\mu_1 \oplus_\beta \mu_2 \succ \mu_1 \oplus_\alpha \mu_2 = \mu_2$ by Definition 14(i)-(ii).

(b) Suppose $\mu_2 \sim \mu_1$. Then Definition 14(i)-(ii) and (a) together imply $\mu_2 \sim \mu_1 \oplus_1 \mu_3 \succ \mu_1 \oplus_\beta \mu_3$ for all $\beta < 1$. Hence $\alpha = 1$ is the unique $\alpha \in [0,1]$ for which $\mu_2 \sim \mu_1 \oplus_\alpha \mu_3$. By symmetry, if $\mu_2 \sim \mu_3$ then $\alpha = 0$ is the unique $\alpha \in [0,1]$ for which $\mu_2 \sim \mu_1 \oplus_\alpha \mu_3$. Hence suppose $\mu_1 \succ \mu_2 \succ \mu_3$. By (**V2**), we have $\mu_1 \oplus_\alpha \mu_3 \succ \mu_2 \succ \mu_1 \oplus_\beta \mu_3$ for some $\alpha, \beta \in (0,1)$, so the sets $\bar{B} = \{\alpha \in (0,1) \mid \mu_1 \oplus_\alpha \mu_3 \succ \mu_2\}$ and $\underline{B} = \{\beta \in (0,1) \mid \mu_2 \succ \mu_1 \oplus_\beta \mu_3\}$ are both nonempty. Moreover, (i) and Definition 14(iii) imply that $\alpha \in \bar{B}$ for all $\alpha > \inf \bar{B}$ and $\beta \in \underline{B}$ for all $\beta < \sup \underline{B}$ We claim there is an $\alpha \in (0,1)$ s.t. $\mu_2 \sim \mu_1 \oplus_\alpha \mu_3$. Suppose otherwise. Then $\underline{B} \cup \bar{B} = (0,1)$ and by (a), $\alpha > \beta$ for all $\alpha \in \bar{B}$ and $\beta \in \underline{B}$. Hence $\inf \bar{B} = \sup \underline{B}$. Let $\alpha := \inf \bar{B} = \sup \underline{B}$. Now take a monotonically decreasing sequence $\{\alpha_n\} \subset \bar{B}$ and a monotonically increasing sequence $\{\beta_n\} \subset \underline{B}$ s.t. $\alpha_n \to \alpha$ and $\beta_n \to \alpha$. By (a), $\mu_1 \oplus_{\alpha_m} \mu_3 \succ \mu_1 \oplus_{\alpha_n} \mu_3 \succ \mu_2$ for all $n > m$, and thus $\mu_1 \oplus_\alpha \mu_3 \succsim \mu_2$. By a symmetric argument $\mu_2 \succsim \mu_1 \oplus_\alpha \mu_3$. Hence $\mu_2 \sim \mu_1 \oplus_\alpha \mu_3$, yielding a contradiction. Moreover, α is unique since for all $\beta > \alpha$ we have $\mu_1 \oplus_\beta \mu_3 \succ \mu_2$ and for all $\beta < \alpha$ we have $\mu_2 \succ \mu_1 \oplus_\beta \mu_3$.

(c) If $\alpha \in (0,1)$, then (**V1**) implies $\mu_1 \oplus_\alpha \mu_3 \succ \mu_2 \oplus_\alpha \mu_3$ and $\mu_3 \oplus_{1-\alpha} \mu_2 \succ \mu_4 \oplus_{1-\alpha} \mu_2$. By Definition 14(ii), it follows that $\mu_1 \oplus_\alpha \mu_3 \succ \mu_2 \oplus_\alpha \mu_4$.

(d) If $\mu_1 \sim \mu_2$ and $\mu_1 \oplus_\alpha \mu_2 \succ \mu_1$, then $\mu_1 \oplus_\alpha \mu_2 \succ \mu_2$. Hence $(\mu_1 \oplus_\alpha \mu_2) \oplus_\alpha (\mu_1 \oplus_\alpha \mu_2) \succ \mu_1 \oplus_\alpha \mu_2$ by (c) or $\mu_1 \oplus_\alpha \mu_2 \succ \mu_1 \oplus_\alpha \mu_2$ by Proposition 10(iv), both of which are contradictory. A similar argument shows that $\mu_1 \sim \mu_2$ and $\mu_1 \succ \mu_1 \oplus_\alpha \mu_2$ together yield a contradiction. (d) follows.

(e) Definition 14(i)-(ii) imply (e) if $\alpha \in \{0,1\}$. Hence suppose $\alpha \in (0,1)$ and $\mu_1 \sim \mu_2$. If $\mu_3 \sim \mu_1$, then (d) implies $\mu_1 \oplus_\alpha \mu_3 \sim \mu_1 \sim \mu_2 \sim \mu_2 \oplus_\alpha \mu_3$. Suppose instead that $\mu_3 \succ \mu_1$ (the argument is symmetric for $\mu_1 \succ \mu_3$). Then (a) and Proposition 10(iv) imply $\mu_3 \succ \mu_1 \oplus_\alpha \mu_3$. Moreover, suppose $\mu_1 \oplus_\alpha \mu_3 \succ \mu_2 \oplus_\alpha \mu_3$. Then by (b) and Definition 14(ii)-(iii), $\mu_1 \oplus_\alpha \mu_3 \sim \mu_3 \oplus_{1-\beta} (\mu_2 \oplus_\alpha \mu_3) = \mu_2 \oplus_{\alpha\beta} \mu_3$ for a unique $\beta \in (0,1)$. Since $\mu_3 \succ \mu_2$ (given $\mu_3 \succ \mu_1$ and $\mu_1 \sim \mu_2$), we also have $\mu_3 \oplus_{1-\beta} \mu_2 \succ \mu_2 \sim \mu_1$ by (a) and Proposition 10(iv). Thus $\mu_2 \oplus_\beta \mu_3 \succ \mu_1$ by Definition 14(ii), and so $(\mu_2 \oplus_\beta \mu_3) \oplus_\alpha \mu_3 \succ \mu_1 \oplus_\alpha \mu_3$ by (**V1**). Together with Definition 14(iii), this implies $\mu_2 \oplus_{\alpha\beta} \mu_3 \succ \mu_1 \oplus_\alpha \mu_3$, which contradicts the hypothesis, so we cannot have $\mu_1 \oplus_\alpha \mu_3 \succ \mu_2 \oplus_\alpha \mu_3$. A similar argument shows we cannot have $\mu_2 \oplus_\alpha \mu_3 \succ \mu_1 \oplus \alpha \mu_3$. Hence (e) follows.

□

There are several other versions of continuity that can be used in place of (**V2**). For example, (**V2′**): for all $\mu_1, \mu_2, \mu_3 \in \mathcal{M}$ with $\mu_1 \succ \mu_2 \succ \mu_3$, there is some $\alpha \in (0,1)$ with $\mu_2 \sim \mu_1 \oplus_\alpha \mu_3$. This is known as *calibration continuity* or *solvability*, and is implied by Lemma 7(b). Moreover, we could use (**V2″**): for all $\mu_1, \mu_2, \mu_3 \in \mathcal{M}$ with $\mu_1 \succ \mu_2 \succ \mu_3$, we have $A = \{\alpha \in (0,1) \mid \mu_1 \oplus_\alpha \mu_3 \succ \mu_2\}$ and $B = \{\beta \in (0,1) \mid \mu_2 \succ \mu_1 \oplus_\beta \mu_3\}$ are open in $(0,1)$. This is *mixture continuity* and follows from Lemma 7(a)-(b). Note (**V2′**) and (**V2″**) are implied by (**V1**) and (**V2**) jointly – they are in fact weaker than (**V2**) alone.

We can now turn to the much-hyped theorem:

Theorem 11 (Mixture set theorem; Herstein & Milnor, 1953)**.** *Suppose $\mathcal{M}$ is a mixture set and $\succ$ is a weak order on $\mathcal{M}$. Then $\succ$ satisfies axioms* (**V1**) *and* (**V2**) *iff there is a mixture-preserving function $u : \mathcal{M} \to \mathbb{R}$ such that*

$$\mu_1 \succ \mu_2 \quad \textit{iff} \quad u(\mu_1) > u(\mu_2) \quad \textit{for all } \mu_1, \mu_2 \in \mathcal{M}.$$

Moreover, if u is a mixture-preserving function representing $\succ$ in the sense that $\mu_1 \succ \mu_2$ iff $u(\mu_1) > u(\mu_2)$, then any other mixture-preserving function $v : \mathcal{M} \to \mathbb{R}$ represents $\succ$ iff v is a positive affine transformation of u, that is, there are numbers $a, b \in \mathbb{R}$ with $a > 0$ such that $v(\mu) = au(\mu) + b$ for all $\mu \in \mathcal{M}$.

Proof. First, we claim that (**V1**) and (**V2**) are necessary for the conclusion. First suppose (**V1**) does not hold but that there is some mixture-preserving u with $\mu_1 \succ \mu_2$ iff $u(\mu_1) > u(\mu_2)$. Take $\mu_1 \succ \mu_2$ and $\mu_1 \oplus_\alpha \mu_3 \precsim \mu_2 \oplus_\alpha \mu_3$ with $\alpha \in (0,1)$. If the conclusion holds then $u(\mu_1) > u(\mu_2)$ and $u(\mu_1 \oplus_\alpha \mu_3) \le u(\mu_2 \oplus_\alpha \mu_3)$. Now if u is mixture-preserving, $u(\mu_1 \oplus_\alpha \mu_3) = \alpha u(\mu_1) + (1-\alpha)u(\mu_3) > \alpha u(\mu_2) + (1-\alpha)u(\mu_3) = u(\mu_2 \oplus_\alpha \mu_3)$, yielding a contradiction. Second, suppose (**V2**) does not hold but the conclusion holds. Then there are $\mu_1, \mu_2, \mu_3 \in \mathcal{M}$ with $\mu_1 \succ \mu_2 \succ \mu_3$ s.t. there is no $\alpha \in (0,1)$ with $\mu_1 \oplus_\alpha \mu_3 \succ \mu_2$ or there is no $\beta \in (0,1)$ with $\mu_2 \succ \mu_1 \oplus_\beta \mu_3$. Then $u(\mu_1) > u(\mu_2) > u(\mu_3)$. By the Archimedean property of $\mathbb{R}$, we have some $\alpha, \beta \in (0,1)$ s.t. $\alpha u(\mu_1) + (1-\alpha)u(\mu_3) > u(\mu_2)$ and $\beta u(\mu_1) + (1-\beta)u(\mu_3) < u(\mu_2)$. Since u is mixture-preserving, it follows that $u(\mu_1 \oplus_\alpha \mu_3) > u(\mu_2)$ and $u(\mu_1 \oplus_\beta \mu_3) < u(\mu_2)$. But then $\mu_1 \oplus_\alpha \mu_3 \succ \mu_2$ and $\mu_2 \succ \mu_1 \oplus_\beta \mu_3$, yielding a contradiction.

Next we tackle sufficiency. We first prove the following lemma.

Lemma 8. *Suppose $\succ$ is a weak order on a mixture set $\mathcal{M}$ and that* (**V1**) *and* (**V2**) *hold. For any $\mu_1, \mu_2 \in \mathcal{M}$ with $\mu_1 \succ \mu_2$, define $M(\mu_1, \mu_2) = \{\mu \in \mathcal{M} \mid \mu_1 \succ \mu \succ \mu_2\}$. Then there is a mixture-preserving function $u : M(\mu_1, \mu_2) \to \mathbb{R}$ such that $\mu \succ \mu'$ iff $u(\mu) > u(\mu')$ for all $\mu, \mu' \in M(\mu_1, \mu_2)$.*

Proof. Suppose the hypotheses of the lemma hold and $\mu_1 \succ \mu_2$. By Lemma 7(b), we can assign a unique number $f(\mu) \in [0,1]$ to each $\mu \in M(\mu_1, \mu_2)$ so that

$$\mu \sim \mu_1 \oplus_{f(\mu)} \mu_2$$

with $f(\mu_1) = 1$ and $f(\mu_2) = 0$. Now take $\mu, \mu' \in M(\mu_1, \mu_2)$ s.t. $f(\mu) > f(\mu')$. By Lemma 7(a), $\mu \oplus_{f(\mu)} \mu'' \succ \mu' \oplus_{f(\mu')} \mu''$, which by transitivity and the definition of f

imply that $\mu \succ \mu'$. Now if instead $f(\mu) = f(\mu')$, then transitivity of $\sim$ and the definition of f imply $\mu \sim \mu'$. Thus for all $\mu, \mu' \in M(\mu_1, \mu_2)$, $\mu \succ \mu'$ iff $f(\mu) > f(\mu')$.

Next, note if $\mu, \mu' \in M(\mu_1, \mu_2)$, then $\mu \oplus_\alpha \mu' \in M(\mu_1, \mu_2)$ for all $\alpha \in [0, 1]$. This follows by Definition 14(i)-(ii) for $\alpha \in \{0, 1\}$; for $\alpha \in (0, 1)$, $\mu_1 = \mu_1 \oplus_\alpha \mu_1 \succsim \mu \oplus_\alpha \mu_1 = \mu_1 \oplus_{1-\alpha} \mu \succsim \mu' \oplus_{1-\alpha} \mu = \mu \oplus_\alpha \mu' \succsim \mu_2 \oplus_\alpha \mu' = \mu' \oplus_{1-\alpha} \mu_2 \succsim \mu_2 \oplus_{1-\alpha} \mu_2 = \mu_2$. Thus if $\mu, \mu' \in M(\mu_1, \mu_2)$, it follows by definition of f that $\mu \oplus_\alpha \mu' \sim \mu_1 \oplus_{f(\mu_1 \oplus_\alpha \mu_2)} \mu_2$ for all $\alpha \in [0, 1]$. Now, applying Lemma 7(e)twice gives

$$\mu \oplus_\alpha \mu' \sim (\mu_1 \oplus_{f(\mu)} \mu_2) \oplus_\alpha (\mu_1 \oplus_{f(\mu')} \mu_2),$$

and thus by Proposition 10(v),

$$\mu \oplus_\alpha \mu' \sim \mu_1 \oplus_{\alpha f(\mu) + (1-\alpha) f(\mu')} \mu_2,$$

which with transitivity and Lemma 7(a) implies that f is mixture-preserving. $\square$

Now, we extend the lemma to conclude that there is a mixture preserving function $u : \mathcal{M} \to \mathbb{R}$ s.t. $\mu \succ \mu'$ iff $u(\mu) > u(\mu')$ for all $\mu, \mu' \in \mathcal{M}$. Take $\mu_1 \succ \mu_2$ and define M as in Lemma 8. For $i = 1, 2$, choose μ_1^i, μ_2^i so that $\mu_1^i \succ \mu_2^i$ and $M(\mu_1, \mu_2) \subset M(\mu_1^i, \mu_2^i)$. Let $f_i^* : M(\mu_1^i, \mu_2^i) \to \mathbb{R}$ be mixture-preserving functions such that $\mu \succ \mu'$ iff $f_i^*(\mu) \succ f_i^*(\mu')$ for all $\mu, \mu' \in M(\mu_1^i, \mu_2^i)$. Define $f_i : M(\mu_1^i, \mu_2^i) \to \mathbb{R}$ to be increasing linear transformations of f_i^* so that $f_i(\mu_1) = 1$ and $f_i(\mu_2) = 0$. Since increasing linear transformations are order-preserving, $\mu \succ \mu'$ iff $f_i(\mu) \succ f_i(\mu')$ for all $\mu, \mu' \in M(\mu_1^i, \mu_2^i)$.

Take $\mu \in M(\mu_1^1, \mu_2^1) \cap M(\mu_1^2, \mu_2^2)$. Since we reuse it, I will call the following argument the "f_i-equivalence argument". By definition, if $\mu \sim \mu_1$ or $\mu \sim \mu_2$ then $f_1(\mu) = f_2(\mu)$. Of the remaining possibilities, Lemma 7(b) and strict preference guarantee that the $\alpha, \beta, \gamma \in (0, 1)$ is unique:

$$\begin{array}{lll} \mu \succ \mu_1 \succ \mu_2 & \text{implies} & \mu_1 \sim \mu \oplus_\alpha \mu_2, \\ \mu_1 \succ \mu \succ \mu_2 & \text{implies} & \mu \sim \mu_1 \oplus_\beta \mu_2, \\ \mu_1 \succ \mu_2 \succ \mu & \text{implies} & \mu_2 \sim \mu_1 \oplus_\gamma \mu. \end{array}$$

Now applying the definition of f_i to each gives

$$\begin{aligned} f_i(\mu_1) = 1 = \alpha f_i(\mu) \quad & = \alpha f_i(\mu) + (1-\alpha) f_i(\mu_2), \\ f_i(\mu) = \beta \quad & = \beta f_i(\mu_1) + (1-\beta) f_i(\mu_2), \\ f_i(\mu_2) = 0 = \gamma + (1-\gamma) f_i(\mu) \quad & = \gamma f_i(\mu_1) + (1-\gamma) f_i(\mu). \end{aligned}$$

In each case, we have $f_1(\mu) = f_2(\mu)$. Since for all collections of intervals $\{M(\mu_1^i, \mu_2^i)\}$ containing $\mu_1, \mu_2, \mu \in M(\mu_1^i, \mu_2^i)$, we have that $f_i(\mu) = f_j(\mu)$, we can take $u(\mu) := f_i(\mu)$ for any such i. Now, every $\mu, \mu' \in \mathcal{M}$ lies in some such interval $M(\mu_1^i, \mu_2^i)$, and so u is well-defined on $\mathcal{M}$ and satisfies $\mu \succ \mu'$ iff $u(\mu) > u(\mu')$.

Finally, we establish the claim that if u represents $\succ$ then any other mixture-preserving function v represents $\succ$ iff it is an affine transformation of u. Suppose u

represents $\succ$. If $v(\mu) = au(\mu) + b$ for all $\mu \in \mathcal{M}$ with $a > 0$, then clearly v represents $\succ$, for $u(\mu_1) > u(\mu_2)$ iff $v(\mu_1) > v(\mu_2)$ and $v(\mu_1 \oplus_\alpha \mu_2) = au(\mu_1 \oplus_\alpha \mu_2) + b = a[\alpha u(\mu_1)+(1-\alpha)u(\mu_2)]+b = \alpha(au(\mu_1)+b)+(1-\alpha)(au(\mu_2)+b) = \alpha v(\mu_1)+(1-\alpha)v(\mu_2)$. For the converse, suppose $u, v : \mathcal{M} \to \mathbb{R}$ are mixture-preserving and both represent $\succ$. First, if u is constant on $\mathcal{M}$, then $\mu_1 \sim \mu_2$ for all $\mu_1, \mu_2 \in \mathcal{M}$, so v is constant on $\mathcal{M}$, and $v(\mu) = u(\mu) + (c' - c)$ where $c := u$ and $c' := v$, and so v is an affine transformation of u. Now suppose $\mu_1 \succ \mu_2$ for some $\mu_1, \mu_2 \in \mathcal{M}$. For all $\mu \in \mathcal{M}$, define

$$f_1(\mu) = \frac{u(\mu) - u(\mu_2)}{u(\mu_1) - u(\mu_2)} \quad \text{and} \quad f_2(\mu) = \frac{v(\mu) - v(\mu_2)}{v(\mu_1) - v(\mu_2)}.$$

Then f_1 and f_2 are both mixture-preserving functions representing $\succ$ since they are linear transformations of u and v. Moreover, $f_1(\mu_1) = f_2(\mu_1) = 1$ and $f_1(\mu_2) = f_2(\mu_2) = 0$. By the f_i-equivalence argument above, $f_1 = f_2$. Now, by definition of f_1 and f_2,

$$v(\mu) = \frac{v(\mu_1) - v(\mu_2)}{u(\mu_1) - u(\mu_2)}u(\mu) + v(\mu_2) - \frac{v(\mu_1) - v(\mu_2)}{u(\mu_1) - u(\mu_2)}u(\mu_2) =: au(\mu) + b,$$

concluding the proof. □

Note that under the hypotheses of Theorem 11, there are other utility functions representing $\succ$ that are not mixture-preserving. In particular, if u is a mixture-preserving function representing $\succ$, any strictly monotone transformation of u also represents $\succ$, but by the theorem, such a transformation will only be mixture-preserving if it is an affine transformation. For example, $\log u$ also represents $\succ$, but given $\mu_1, \mu_2 \in \mathcal{M}$ and $\alpha \in (0, 1)$, $\log u(\mu_1 \oplus_\alpha \mu_2) = \log(\alpha u(\mu_1) + (1-\alpha)u(\mu_2)) \neq \alpha \log u(\mu_1) + (1-\alpha) \log u(\mu_2)$, so $\log u$ is clearly not mixture-preserving.

2.1.2 Von Neumann-Morgenstern expected utility representations

The mixture set theorem (Theorem 11) is a substantial generalization of von Neumann and Morgenstern's representation theorem. Indeed, von Neumann and Morgenstern's theorem turns out to be a fairly straightforward corollary of the mixture set theorem, applied to simple probability measures (that is, probability measures with finite support, see Example 13(b)).

However, what the mixture set theorem tells us in the von Neumann-Morgenstern setting is that there is some mixture-preserving function of lotteries that represents preferences satisfying the von Neumann-Morgenstern axioms. However, an expected utility representation is not quite this. Rather it is the following:

Definition 16 (Von Neumann-Morgenstern expected utility)**.** Let X be a space of consequences and let $L \subseteq \Delta(X)$ be a set of probability measures on X. If $\succ$ is a weak order on L, then we say that a function $u : X \to \mathbb{R}$ is a *Bernoulli utility function* representing $\succ$ if

$$\mu_1 \succ \mu_2 \quad \text{iff} \quad \int_X u(x)\mu_1(\mathrm{d}x) > \int_X u(x)\mu_2(\mathrm{d}x) \qquad \text{for actually } \mu_1, \mu_2 \in L.$$

We call the function $U : L \to \mathbb{R}$ defined by

$$U(\mu) = \int_X u(x)\mu(\mathrm{d}x) \qquad \text{for each } \mu \in L$$

a (von Neumann-Morgenstern) *expected utility representation* of $\succ$.

Naturally, if μ is a simple probability distribution, $\int_X u(x)\mu(\mathrm{d}x)$ can be written as $\sum_{x \in \operatorname{supp} \mu} u(x)\mu(x)$.

In an expected utility representation, we have an intuitive interpretation for u. Decisionmakers have preferences over consequences, and u encodes those preferences. If L includes all the degenerate lotteries over X, it is easy to see how $\succ$, a weak order over lotteries, implies a weak order $\tilde{\succ}$ over consequences – we have $x \,\tilde{\succ}\, y$ iff $\delta_x \succ \delta_y$, where δ_x, δ_y are Dirac measures (and thus degenerate lotteries).[21] However, u encodes even more. That $x \,\tilde{\succ}\, y \,\tilde{\succ}\, z$, say, tells us that the decisionmaker prefers x to y to z, but it does not tell us whether the decisionmaker prefers a lottery that places equal probability $\frac{1}{2}$ on x and z to a lottery that places certainty on y. The Bernoulli utility function u, however, does tell us this. Since expectations are linear operators, the *values* of u are informative about the decisionmaker's preferences over lotteries (up to scale). This insight is often phrased as the Bernoulli utility function having a cardinal interpretation.

Proposition 12 (Cardinal uniqueness)**.** *Let $\Delta_D(X)$ denote the set of all Dirac measures on X. Consider any convex subset $L \subseteq \Delta(X)$ such that $\Delta_D(X) \subseteq L$, and let $\succ$ be a weak order on L. Then $u : X \to \mathbb{R}$ and $v : X \to \mathbb{R}$ are Bernoulli utility representations of $\succ$ iff v is a positive affine transformation of u.*

Proof. Let $U(\mu) = \int_X u \,\mathrm{d}\mu$ and $V(\mu) = \int_X v \,\mathrm{d}\mu$. Since L is convex, it is a mixture set when endowed with the convex combination operation $\mu_1 \oplus_\alpha \mu_2 = \alpha\mu_1 + (1-\alpha)\mu_2$ for $\mu_1, \mu_2 \in L$. By Example 15, if U and V are expected utility representations, then they are mixture-preserving functions.

First suppose u is a Bernoulli utility representation of $\succ$. Take $a, b \in \mathbb{R}$ with $a > 0$. If $v(x) = au(x) + b$ for all $x \in X$ then by linearity of the integral, $V(\mu) = aU(\mu) + b$ for each $\mu \in L$, and the second part of Theorem 11 implies V is an expected utility representation of $\succ$ (and thus v is a valid Bernoulli utility representation of $\succ$).

Conversely, if U and V are expected utility representations of $\succ$, then by the second part of Theorem 11, $V(\mu) = aU(\mu) + b$ for some $a > 0$ and $b \in \mathbb{R}$. Now since $\Delta_D(X) \subseteq L$, we have $v(x) = V(\delta_x) = aU(\delta_x) + b = au(x) + b$ for each $x \in X$. $\square$

Theorem 12 (Von Neumann-Morgenstern representation theorem)**.** *Let X be a space of consequences, let $\Delta_S(X)$ denote the space of simple probability measures on X, and let $\succ$ be a weak order on $\Delta_S(X)$. Then $\succ$ satisfies axioms* (**V1**) *and* (**V2**) *iff there is a function $u : X \to \mathbb{R}$ such that*

$$\mu_1 \succ \mu_2 \quad \textit{iff} \quad \sum_{x \in \operatorname{supp} \mu_1} u(x)\mu_1(x) > \sum_{x \in \operatorname{supp} \mu_2} u(x)\mu_2(x)$$

[21] See Definition **??**. A *Dirac measure* δ_x concentrated at x takes value $\delta_x(A) = 1$ whenever $x \in A$ and $\delta_x(A) = 0$ whenever $x \notin A$.

for all $\mu_1, \mu_2 \in \Delta_S(X)$.

Proof. Note that $\Delta_S(X)$ is convex (Example 13(b)), and thus is a mixture set when endowed with the convex combination operation $\mu_1 \oplus_\alpha \mu_2 = \alpha\mu_1 + (1-\alpha)\mu_2$ for all $\mu_1, \mu_2 \in \Delta_S(X)$ and $\alpha \in [0,1]$. Applying Theorem 11 immediately gives that there is some mixture-preserving function $f : \Delta_S(X) \to \mathbb{R}$. For any such f, define $u : X \to \mathbb{R}$ by $u(x) = f(\delta_x)$ for each $x \in X$. For each $k \in \mathbb{N}$, define $\Delta_k(X) := \{\mu \in \Delta_S(X) \mid |\operatorname{supp}\mu| \leq k\}$. We proceed by induction. Fix $k > 1$ and suppose we have that $f(\mu) = \sum_{x\in\operatorname{supp}\mu} u(x)\mu(x)$ for all $\mu \in \Delta_{k-1}(X)$ (by definition of u, f can be decomposed in this form for $k = 1$.) Take any $\mu \in P_k(X)$, and note we have that $\mu = \alpha\mu_1 + (1-\alpha)\mu_2$ for some $\mu_1, \mu_2 \in \Delta_{k-1}(X)$. Since f is mixture-preserving,

$$\begin{aligned}
f(\mu) &= \alpha f(\mu_1) + (1-\alpha) f(\mu_2) \\
&= \alpha \sum_{x\in\operatorname{supp}\mu_1} u(x)\mu_1(x) + (1-\alpha) \sum_{x\in\operatorname{supp}\mu_2} u(x)\mu_2(x) \\
&= \sum_{x\in\operatorname{supp}\mu} u(x)[\alpha\mu_1(x) + (1-\alpha)\mu_2(x)] \qquad = \sum_{x\in\operatorname{supp}\mu} u(x)\mu(x).
\end{aligned}$$

This establishes the theorem. □

Thus for lotteries that correspond to simple probability measures, the mixture set theorem immediately gives us a von Neumann-Morgenstern expected utility representation. However, more generally, lotteries can be non-simple probability measures. In this case, the mixture set theorem still tells us that there is some mixture-preserving function f on the set of lotteries that represents $\succ$. But such a function f does not generally have an expected utility interpretation. When the set of lotteries contains non-simple probability measures, we need a slightly different approach to ensure that $\succ$ has an expected utility representation.

2.1.3 Expected utility with general probability measures

There are a few ways of extending the von Neumann-Morgenstern representation theorem to general probability measures. The way that Fishburn (1970) does it is very inelegant. Grandmont (1972), who was concerned with conditions under which a *continuous* Bernoulli utility function representing $\succ$ exists, requires far fewer technical axioms. This is elegant but requires dealing with the weak* topology (see the mathematical appendix, Section **??** and possibly Section **??**).

We need a new continuity axiom in place of (**V2**):

Axiom. Let X be a separable, metrizable space of consequences and let $\succ$ be a weak order over $\Delta(X)$.

(**V3**) *Topological continuity under the weak* topology.* For any $\mu \in \Delta(X)$, both $\{\nu \in \Delta(X) \mid \nu \succ \mu\}$ and $\{\nu \in \Delta(X) \mid \mu \succ \nu\}$ are open in the weak* topology.

The independence axiom (**V1**), meanwhile, is sufficient as is.

Theorem 13 (Grandmont, 1972). *Let X be a separable, metrizable space of consequences and let $\succ$ be a weak order on $\Delta(X)$. Then $\succ$ satisfies axioms* (**V1**) *and* (**V3**) *iff there is a bounded, continuous function $u : X \to \mathbb{R}$ such that*

$$\mu_1 \succ \mu_2 \quad \textit{iff} \quad \int_X u(x)\mu_1(\mathrm{d}x) > \int_X u(x)\mu_2(\mathrm{d}x)$$

for all $\mu_1, \mu_2 \in \Delta(X)$.

Proof. The proof boils down to two steps, neither of which I will prove. First:

Lemma 9. *$\succ$ satisfies axioms* (**V1**) *and* (**V3**) *iff there is a linear and weak*-continuous function $U : \Delta(X) \to \mathbb{R}$ representing $\succ$.*

Proof. Omitted. □

Given Lemma 9, we now need only show that the linear, weak*-continuous function is in fact an expected utility representation.

Lemma 10. *The function $U : \Delta(X) \to \mathbb{R}$ is linear and weak*-continuous iff there is a bounded, continuous function $u : X \to \mathbb{R}$ such that*

$$U(\mu) = \int_X u \, \mathrm{d}\mu$$

for each $\mu \in \Delta(X)$.

Proof. Omitted. □

Now the theorem follows immediately from Lemma 9 and Lemma 10. □

2.1.4 Monetary consequences and risk aversion

A common setting for much thinking about risk is when the consequences are monetary payoffs, i.e. the set of consequences is $X = \mathbb{R}$ (or $X = \mathbb{R}_+$). Such settings arise naturally in finance, firm investment problems, labour economics (e.g. acquiring education has uncertain payoff in terms of salary), and so on.

We will equip $\mathbb{R}$ with its Borel σ-algebra $\mathcal{B}$, and refer to the lotteries in $\Delta(\mathbb{R})$ as *monetary lotteries*. Since monetary lotteries are probability measures on $\mathbb{R}$, it can be convenient to associate monetary lotteries with a random variable $f : \Omega \to \mathbb{R}$ (where $(\Omega, \Sigma, \mathbb{P})$ is some probability space). The monetary lottery μ is then the law of f, that is, $\mu(B) = \mathbb{P}(\{\omega \in \Omega \mid f(\omega) \in B\})$ for every Borel set $B \in \mathcal{B}$. Thus all the standard objects we are familiar with for random variables can be defined for monetary lotteries.

In particular, given a monetary lottery μ, we can define the *cumulative distribution function* (cdf) of μ as

$$F_\mu(x) = \mu((-\infty, x)) \qquad \text{for all } x \in \mathbb{R}.$$

Let Φ be the set of all cdfs, i.e. the set of all nondecreasing right-continuous functions F that have $\lim_{t\downarrow-\infty} F(t) = 0$ and $\lim_{t\uparrow\infty} F(t) = 1$. Now, any cdf $F \in \Phi$ corresponds to a monetary lottery $\mu \in \Delta(\mathbb{R})$, and any monetary lottery has a corresponding cdf. Moreover, equipped with mixture operation $F_{\mu_1 \oplus_\alpha \mu_2} = \alpha F_{\mu_1} + (1-\alpha)F_{\mu_2}$ for all $F_1, F_2 \in \Phi$ and $\alpha \in [0,1]$. Thus it is no loss to represent the monetary lotteries with Φ instead of $\Delta(\mathbb{R})$.

Second, we can use random variables more explicitly. We mentioned that we can represent a monetary lottery $\mu \in \Delta(\mathbb{R})$ via a random variable Y on some probability space $(\Omega, \Sigma, \mathbb{P})$. This probability space is essentially arbitrary provided it is "rich enough" to represent μ (for example, a space with trivial sigma-algebra $\Sigma = \{\varnothing, \Omega\}$ cannot represent a non-degenerate lottery μ). An easy way to guarantee $(\Omega, \Sigma, \mathbb{P})$ is rich enough is to take $\Omega = (0,1)$, take Σ to be the Borel σ-algebra in $(0,1)$, and $\mathbb{P} = \lambda$ where λ is the Lebesgue measure restricted to $(0,1)$ (making $\mathbb{P}$ a uniform probability measure). Given this probability space, we can take the quantile function as the random variable representing the monetary lottery. Given a monetary lottery $\mu \in \Delta(\mathbb{R})$, we define the *quantile function* f_μ of μ by

$$f_\mu(t) = \inf\{x \in \mathbb{R} \mid t \leq F(x)\} \qquad \text{for } t \in (0,1).$$

Any nondecreasing, right-continuous function f is a quantile function for some monetary lottery $\mu \in \Delta(X)$. Letting $\mathscr{F}$ be the set of all quantile functions, we can thus represent the set of monetary lotteries by $\mathscr{F}$. However, $f_{\mu_1 \oplus_\alpha \mu_2} \neq \alpha f_{\mu_1} + (1-\alpha) f_{\mu_2}$ in general.

All three formulations of a monetary lottery reduce to equivalent expected utility representations:

Proposition 13. *Let F denote the cdf of a probability measure $\mu \in \Delta(\mathbb{R})$ and let f denote the quantile function of μ. Then for any bounded, continuous function $u : X \to \mathbb{R}$,*

$$\int_{\mathbb{R}} u(x)\mu(\mathrm{d}x) = \int_{\mathbb{R}} u(x)\,\mathrm{d}F(x) = \int_0^1 u(f(\omega))\,\mathrm{d}\omega.$$

Monetary consequences $\mathbb{R}$ provide a setting where we can precisely define the notion of risk attitudes. All of the following definitions also hold if we replace $\mathbb{R}$ with a convex subset of vector space, equipped with a partial order $\geq$.

The definition of risk aversion below only really makes sense where it is natural to suppose preferences $\succ$ are monotone in $\geq$. For monetary lotteries this holds – more money is usually better. But if the consequences are litres of oil spilled into the ocean, say, then we would (hopefully) want to flip the definitions below for them to make sense.

Definition 17 (Absolute risk attitudes)**.** Define the *expected monetary value* of a lottery $\mu \in \Delta(\mathbb{R})$ by $\int_{\mathbb{R}} x\mu(\mathrm{d}x)$,[22] and define $\delta_\mu := \delta_{\int_{\mathbb{R}} x\mu(\mathrm{d}x)}$, the Dirac measure placing probability mass at the expected monetary value of μ.

[22] Or equivalently, $\int_0^1 f_\mu(\omega)\,\mathrm{d}\omega$ or $\int_{\mathbb{R}} x\,\mathrm{d}F_\mu(x)$.

(a) *Risk aversion.* We say preference relation $\succ$ is *risk averse* if $\delta_\mu \succsim \mu$ for all $\mu \in \Delta(\mathbb{R})$. Moreover, we say $\succ$ is *strictly risk averse* if $\delta_\mu \succ \mu$ for all $\mu \in \Delta(\mathbb{R})$.

(b) *Risk neutrality.* We say preference relation $\succ$ is *risk neutral* if $\delta_\mu \sim \mu$ for all $\mu \in \Delta(\mathbb{R})$.

(c) *Risk loving.* We say preference relation $\succ$ is *risk loving* if $\mu \succsim \delta_\mu$ for all $\mu \in \Delta(\mathbb{R})$. Moreover, we say $\succ$ is *strictly risk loving* if $\mu \succ \delta_\mu$ for all $\mu \in \Delta(\mathbb{R})$.

These are global notions – it may be that you are risk averse in some cases and risk loving in others (a classic example would be people who play the lottery and buy house insurance).

In the following proofs, we continue to use $x \oplus_\alpha y$ to denote the convex combination operation $\alpha x + (1-\alpha)y$.

Proposition 14. *Suppose $\succ$ is a weak order on $\Delta(\mathbb{R})$ and admits expected utility representation $U(\mu) = \int_{\mathbb{R}} u(x)\mu(\mathrm{d}x)$. Then*

(i) *$\succ$ is risk averse iff u is concave, and strictly risk averse iff u is strictly concave;*

(ii) *$\succ$ is risk neutral iff u is affine;*

(iii) *$\succ$ is risk loving iff u is convex, and strictly risk loving iff u is strictly convex.*

Proof. Suppose u is concave. By Jensen's inequality and the definition of δ_μ, we have

$$\int_{\mathbb{R}} u(x)\delta_\mu(\mathrm{d}x) = u\left(\int_{\mathbb{R}} x\mu(\mathrm{d}x)\right) \geq \int u(x)\mu(\mathrm{d}x),$$

and thus $\delta_\mu \succsim \mu$.

Conversely, suppose $\delta_\mu \succsim \mu$ for all $\mu \in \Delta(\mathbb{R})$. Consider the lottery μ that results in x_1 with probability α and x_2 with probability $1-\alpha$, for arbitrary $x_1, x_2 \in \mathbb{R}$ and $\alpha \in [0,1]$. Since $\delta_\mu \succsim \mu$, we have $u(x_1 \oplus_\alpha x_2) \geq u(x_1) \oplus_\alpha u(x_2)$. Thus u is concave.

The proofs for the other statements are similar. □

There are other familiar ways of thinking about absolute risk attitudes. One is in terms of *certainty equivalents.*

Definition 18 (Certainty equivalent)**.** Consider a weak order $\succ$ on $\Delta(\mathbb{R})$, and any monetary lottery $\mu \in \Delta(\mathbb{R})$.

(a) *Certainty equivalent.* A *certainty equivalent* $e_\succ(\mu)$ of lottery μ is a value $e_\succ(\mu) \in \mathbb{R}$ satisfying $\delta_{e_\succ(\mu)} \sim \mu$.

(b) *Risk premium.* If $e_\succ(\mu)$ is unique, then the *risk premium* of lottery μ is $\int_{\mathbb{R}} x\mu(\mathrm{d}x) - e_\succ(\mu)$.

In general, there could be more than one certainty equivalent for a lottery μ, or none. However, if $\succ$ has a strictly increasing Bernoulli utility representation that is continuous, then the certainty equivalent will be unique. In this case, the certainty equivalent of lottery μ can be interpreted as the decisionmaker's valuation of μ – i.e. the amount they would be willing to pay to buy the lottery, or the minimal amount they would need to be paid to sell the lottery if they already owned it. The risk premium of the lottery is simply the premium demanded by the decisionmaker compared to the expected monetary value of the lottery. In experimental settings, it is common to elicit participants' preferences over lotteries by eliciting their certainty equivalents.[23] Section 2.5.3 discusses some issues with this.

Clearly $\succ$ is risk neutral iff the certainty equivalent of any lottery is equal to its expected monetary value, or equivalently if the risk premium is 0 for all lotteries. Likewise, $\succ$ is risk averse iff $c_{\succ}(\mu) \leq \int_{\mathbb{R}} x\mu(\mathrm{d}x)$ for all lotteries μ, or iff the risk premium is always nonnegative. Finally, $\succ$ is risk loving iff $c_{\succ}(\mu) \geq \int_{\mathbb{R}} x\mu(\mathrm{d}x)$, or iff the risk premium is always nonpositive.

We can also think of *comparative* risk attitudes, generating a "more risk averse than" partial order on the set of preferences that admit expected utility representations.

Definition 19 (Comparative risk attitudes)**.** Let $\succ_1$ and $\succ_2$ be weak orders on $\Delta(\mathbb{R})$ that both admit expected utility representations. We say $\succ_1$ is *more risk averse than* $\succ_2$ if, for all $x \in \mathbb{R}$ and $\mu \in \Delta(\mathbb{R})$, we have

(i) $\delta_x \succ_2 \mu$ implies $\delta_x \succ_1 \mu$, and

(ii) $\delta_x \sim_2 \mu$ implies $\delta_x \succsim_1 \mu$.

That is, $\succ_1$ is more risk averse than $\succ_2$ if whenever $\succ_1$ rejects a lottery for a certain outcome, so too does $\succ_2$.

Proposition 15. *Let $\succ_1$ and $\succ_2$ be weak orders on $\Delta(\mathbb{R})$ admitting expected utility representations, with corresponding Bernoulli utility functions u_1 and u_2. Assume u_1 and u_2 are strictly increasing and continuous. Then $\succ_1$ is more risk averse than $\succ_2$ iff there is a concave function $\varphi : u_2(\mathbb{R}) \to \mathbb{R}$ such that $u_1 = \varphi \circ u_2$.*

Proof. Suppose that we can write $u_1 = \varphi \circ u_2$ for some concave function φ. Since both u_1 and u_2 are strictly increasing, φ must be monotone. Take any lottery μ. If $u_2(x) \geq \int_{\mathbb{R}} u_2(x)\mu(\mathrm{d}x)$ then $u_1(x) = \varphi(u_2(x)) \geq \varphi\left(\int_{\mathbb{R}} u_2(x)\mu(\mathrm{d}x)\right) \geq \int_{\mathbb{R}} \varphi(u_2(x))\mu(\mathrm{d}x) = \int_{\mathbb{R}} u_1(x)\mu(\mathrm{d}x)$, with the first inequality strict if $u_2(x) > \int_{\mathbb{R}} u_2(x)\mu(\mathrm{d}x)$. Hence $\succ_1$ is more risk averse than $\succ_2$.

[23] This is typically done by asking them directly or via the Becker-DeGroot-Marschak procedure, which is incentive compatible *under the assumption that participants' preferences have an expected utility representation*. In the typical version of this procedure, the participant submits a price for the lottery. A random price is then drawn from some distribution (usually uniform). If the random price exceeds the participant's stated price, the participant is paid their stated price. If the random price is below the participant's stated price, the participant receives the outcome of the lottery. See Karni & Safra (1987) for details of why this procedure is not necessarily incentive compatible when we drop the assumption that participants have preferences with an expected utility representation.

Conversely, suppose $\succ_1$ is more risk averse than $\succ_2$. Since u_1 and u_2 are strictly increasing, they are both one-to-one, and since they are continuous, $u_2(\mathbb{R})$ is convex. Let $\varphi : u_2(\mathbb{R}) \to \mathbb{R}$ map each $u_2(x)$ to $u_1(x)$. Towards contradiction, suppose

$$\begin{aligned}\varphi(u_2(x_1) \oplus_\alpha u_2(x_2)) &< \varphi(u_2(x_1)) \oplus_\alpha \varphi(u_2(x_2)) \\ &= u_1(x_1) \oplus_\alpha u_1(x_2)\end{aligned}$$

for some $x_1, x_2 \in \mathbb{R}$ and $\alpha \in (0, 1)$. Let μ be the lottery that results in x_1 with probability α and x_2 with probability $1 - \alpha$. There is some $x \sim_2 \mu$. We have $u_1(x) = \varphi(u_2(x)) = \varphi(u_2(x_1) \oplus_\alpha u_2(x_2)) < u_1(x_1) \oplus_\alpha u_1(x_2)$, so $\mu \succ_1 \delta_x$, contradicting that $\succ_1$ is more risk averse than $\succ_2$. Thus $\varphi(u_2(x_1) \oplus_\alpha u_2(x_2)) \geq \varphi(u_2(x_1)) \oplus_\alpha \varphi(u_2(x_2))$ for all $\alpha \in [0, 1]$ and $x_1, x_2 \in \mathbb{R}$, so φ is concave. □

Proposition 15 indicates that if a Bernoulli utility function has greater curvature, then the corresponding preferences are more risk averse. We can thus "measure" risk aversion if we can measure the curvature of Bernoulli utility functions. A popular pair of measures are the Arrow-Pratt measures, due to Pratt (1964) and Arrow (1965).

Definition 20 (Arrow-Pratt)**.** Let $\succ$ be a weak order on $\Delta(\mathbb{R})$ that has a C^2 (i.e. twice continuously differentiable) Bernoulli utility representation $u : \mathbb{R} \to \mathbb{R}$.

(a) *Coefficient of absolute risk aversion.* The *coefficient of absolute risk aversion* for u at a point $x \in \mathbb{R}$ is $-\frac{u''(x)}{u'(x)}$.

The reciprocal, $-\frac{u'(x)}{u''(x)}$, of the coefficient of absolute risk aversion is known as the *risk tolerance.*

(b) *Coefficient of relative risk aversion.* The *coefficient of relative risk aversion* for u at a point $x \in \mathbb{R}$ is $-\frac{xu''(x)}{u'(x)}$.

As we would hope of a measure of risk aversion, these measures do not change if we apply a positive affine transformation to the Bernoulli utility u: if $v = au + b$, then $-\frac{v''(x)}{v'(x)} = -\frac{au''(x)}{au'(x)} = -\frac{u''(x)}{u'(x)}$.

In general, how "risk averse" a decisionmaker will be will depend on their wealth level. It is common to work with preferences where the degree of risk aversion has a "direction" in wealth.

Definition 21. Consider preference relation $\succ$ on $\Delta(\mathbb{R})$ with Bernoulli utility representation $u : X \to \mathbb{R}$.

(a) *Direction of absolute risk aversion.* We say $\succ$ is *decreasingly risk averse* if for all $\mu \in \Delta(\mathbb{R})$ and all $x_1, x_2, z \in \mathbb{R}$ such that $x_1 < x_2$, we have that

$$\int_{\mathbb{R}} u(x_1 + y)\mu(\mathrm{d}y) > u(x_1 + z) \quad \text{implies} \quad \int_{\mathbb{R}} u(x_2 + y)\mu(\mathrm{d}y) > u(x_2 + z).$$

We say $\succ$ is *increasingly risk averse* if the implication holds with the inequalities reversed, and has constant absolute risk aversion if both implications hold.

(b) *Direction of relative risk aversion.* We say $\succ$ is *decreasingly relatively risk averse* if for all $\mu \in \Delta(\mathbb{R})$ and all $x_1, x_2, z \in \mathbb{R}$ such that $x_1 < x_2$, we have that

$$\int_{\mathbb{R}} u(x_1 y)\mu(\mathrm{d}y) > u(x_1 z) \quad \text{implies} \quad \int_{\mathbb{R}} u(x_2 y)\mu(\mathrm{d}y) > u(x_2 z).$$

We say $\succ$ is *increasingly relatively risk averse* if the implication holds with the inequalities reversed, and has constant relative risk aversion if both implications hold.

The definitions can be easily modified to preferences defined over subsets of $\mathbb{R}$. For C^2 Bernoulli utility representations, the Arrow-Pratt measures map neatly to these two definitions:

(i) $\succ$ is decreasingly [increasingly] risk averse iff $x \mapsto \left(-\frac{u''(x)}{u'(x)}\right)$ is nonincreasing [nondecreasing].

(ii) $\succ$ is decreasingly [increasingly] relatively risk averse iff $x \mapsto \left(-\frac{xu''(x)}{u'(x)}\right)$ is nonincreasing [nondecreasing].

Likewise, constant absolute or relative risk aversion correspond to constant coefficients of absolute or relative risk aversion.

A common assumption in many applications is that decisionmakers' preferences exhibit *hyperbolic absolute risk aversion.* This gives us a utility representation that proves tractable in many contexts, particularly in finance.

Definition 22 (Hyperbolic absolute risk aversion)**.** Let $X \subseteq \mathbb{R}_+$ be an interval. We say a preference relation $\succ$ on $\Delta(X)$ admitting a C^2 Bernoulli utility representation $u : X \to \mathbb{R}$ exhibits *hyperbolic absolute risk aversion* if

$$x \mapsto -\frac{u'(x)}{u''(x)}$$

is affine, that is, if $\succ$ exhibits affine risk tolerance.

The following utility representations are probably familiar.

Proposition 16. *Let $X \subseteq \mathbb{R}_+$ be an interval, and suppose $\succ$ is a preference relation on $\Delta(X)$. that exhibits hyperbolic absolute risk aversion. Then $\succ$ exhibits hyperbolic absolute risk aversion iff $\succ$ has a Bernoulli utility representation $u : X \to \mathbb{R}$ given by*

(i) *either*

$$u(x) = \frac{1-\gamma}{\gamma}\left(\frac{a}{1-\gamma}x + b\right)^{\gamma}$$

for some constants $\gamma \in \mathbb{R} - \{0, 1\}$, $a > 0$ and $b > \sup_{x \in X}\left\{-\frac{ax}{1-\gamma}\right\}$;

(ii) *or*

$$u(x) = \frac{1 - e^{-ax}}{a}$$

for some constant $a > 0$*;*

(iii) *or*

$$u(x) = \ln(ax + b)$$

for some constants $a > 0$ *and* $b > \sup_{x \in X}\{-ax\}$*.*

Proof. First suppose $\succ$ has a Bernoulli representation u as specified. In the first case, we have

$$-\frac{u'(x)}{u''(x)} = \frac{a\left(\frac{a}{1-\gamma}x + b\right)^{\gamma-1}}{a^2\left(\frac{a}{1-\gamma}x + b\right)^{\gamma-2}} = \frac{1}{1-\gamma}x + b.$$

In the second case, we have

$$-\frac{u'(x)}{u''(x)} = \frac{e^{-ax}}{ae^{-ax}} = \frac{1}{a}.$$

Finally, in the log-utility case, we have

$$-\frac{u'(x)}{u''(x)} = \frac{\frac{a}{ax+b}}{\frac{a^2}{(ax+b)^2}} = x + \frac{b}{a}.$$

In all cases, we see that $x \mapsto -\frac{u'(x)}{u''(x)}$ is affine, and thus $\succ$ exhibits hyperbolic absolute risk aversion.

Conversely, suppose $\succ$ exhibits hyperbolic absolute risk aversion. Then it admits a C^2 Bernoulli utility representation $u : X \to \mathbb{R}$. Then $\tau(x) = -\frac{u'(x)}{u''(x)}$ is affine, i.e. $\tau(x) = \alpha x + \beta$ for some constants α, β. If $\tau = 0$ then $u'(x) = 0$ everywhere, and so u is constant, and trivially admits a Bernoulli representation of form (iii) with $a = 0$. Hence suppose τ is not everywhere zero, i.e. $\alpha \neq 0$ or $\beta \neq 0$. We separate into the two cases:

(i) $\alpha \notin \{0, 1\}$. Then we have a second order ordinary differential equation,

$$(\alpha x + \beta)u''(x) - u'(x) = 0.$$

Taking $v = u'$, we can write this as a first order ordinary differential equation,

$$v'(x) = \frac{1}{\alpha x + \beta}v(x),$$

which has a solution

$$v(x) = ce^{-A(x)}$$

for arbitrary constant c, where $A(x) = \frac{1}{\alpha}\ln(\alpha x + \beta)$, which has $A'(x) = \frac{1}{\alpha x + \beta}$. This gives $u'(x) = ce^{-\frac{1}{\alpha}\ln(\alpha x + \beta)} = c\,(\alpha x + \beta)^{-\frac{1}{\alpha}}$. Integrating, we have

$$u(x) = \frac{c}{\alpha - 1}\,(\alpha x + \beta)^{\frac{\alpha-1}{\alpha}}.$$

We have, without loss, omitted constants of integration throughout.[24] Now taking $\gamma = \frac{\alpha-1}{\alpha}$, $a = c^{1/\gamma}$ and $b = \beta c^{1/\gamma}$ gives

$$u(x) = \frac{1-\gamma}{\gamma}\left(\frac{a}{1-\gamma}x + b\right)^{\gamma}.$$

(ii) $\alpha = 0$. In this case, repeating the above steps gives us $u'(x) = ce^{-x/\beta}$ for arbitrary constant c. Integrating, we have $u(x) = \frac{1}{\beta} - \beta ce^{-x/\beta}$, where we chose $\frac{1}{\beta}$ as the constant of integration. Taking $a = \frac{1}{\beta}$ and setting $c = \frac{1}{\beta^2}$ gives the representation.

(iii) $\alpha = 1$. In this case, repeating the above steps gives us $u'(x) = \frac{c}{x+\beta}$. Integrating, we have

$$u(x) = c\ln(x+\beta) + b = \ln(e^c(x + \beta e^{-c})).$$

Taking $a = e^c$ and $b = \beta e^{-c}$ gives

$$u(x) = \ln(ax + b).$$

This exhausts all possible cases. □

Hyperbolic absolute risk aversion nests a number of other special cases – constant absolute risk aversion and constant relative risk aversion, in particular. Most of the commonly assumed Bernoulli utility functions used in macroeconomics, for example, represent HARA preferences. Indeed, case (iii) is the familiar log utility, case (ii) is exponential utility, and case (i) nests isoelastic utility.

2.1.5 Stochastic dominance

It is often useful to consider partial orders over the set of monetary lotteries. More generally, these partial orders are really over sets random variables, but as noted in the previous section, we can take each monetary lottery to be the law of some random variable. Such partial orders are known as *stochastic orderings*. The most commonly used stochastic orderings in economic theory are first order and second order stochastic dominance.

Definition 23 (First order stochastic dominance)**.** Let $(\Omega, \Sigma, \mathbb{P})$ be a probability space. Let $\mathcal{Y}$ be a set of random variables $Y : \Omega \to \mathbb{R}$. For each random variable Y, let F_Y denote the cdf of Y.

For each $Y, Y' \in \mathcal{Y}$, define $Y \geq Y'$ if $F_Y(x) < F_{Y'}(x)$ for all $x \in \mathbb{R}$. Then $>$ is the *first order stochastic dominance* order on $\mathcal{Y}$.

[24] If we included arbitrary constants of integration, there would be an extra step in applying an affine transformation to get our desired Bernoulli representation.

As usual, a strict partial order can be defined from $\geq$ by $Y > Y'$ iff $Y \geq Y'$ and $Y' \not\geq Y$. Given our identification of lotteries μ with the laws of random variables, we can also equip the set of monetary lotteries $\Delta(X)$ for $X \subseteq \mathbb{R}$ with a first order or second order stochastic dominance relation $\geq$ directly.

If a lottery μ first order stochastically dominates a lottery μ', then the lottery μ pays off a higher amount for sure than lottery μ'. We would expect that any rational decisionmaker's preferences are monotone with respect to first order stochastic dominance, and this is indeed what experimental evidence usually finds, provided lotteries are not presented to participants in complicated ways.

Proposition 17. *Let $(\mathcal{Y}, \geq)$ denote a set of integrable random variables on probability space $(\Omega, \Sigma, \mathbb{P})$ partially ordered by the first order stochastic dominance order $\geq$. For each $Y \in \mathcal{Y}$, let f_Y denote the quantile function of Y. For any two random variables $Y, Y' \in \mathcal{Y}$, the following are equivalent:*

(i) *$Y \geq Y'$;*

(ii) *$f_Y(t) \geq f_{Y'}(t)$ for all $t \in (0, 1)$;*

(iii) *for all increasing functions $\varphi : \mathbb{R} \to \mathbb{R}$,*

$$\mathbb{E}[\varphi(Y)] \geq \mathbb{E}[\varphi(Y')].$$

Proof. Suppose (i). For each $t \in (0, 1)$ and arbitrary random variable Z, define $E_Z(t) = \{x \in \mathbb{R} \mid t \leq F_Z(x)\}$. Recall the quantile function f_Z of Z is defined by $f_Z(t) = \inf E_Z(t)$. Since $F_Y(x) \leq F_{Y'}(x)$, $t \leq F_Y(x)$ implies $t \leq F_{Y'}(x)$, and hence $E_Y \subseteq E_{Y'}$, giving $f_Y(t) \geq f_{Y'}(t)$. Thus (i) implis (ii).

Suppose (ii). For any increasing function φ, $f_Y \geq f_{Y'}$ implies

$$\mathbb{E}[\varphi(Y')] = \int_0^1 \varphi(f_{Y'}(t)) \, \mathrm{d}t \leq \int_0^1 \varphi(f_Y(t)) \, \mathrm{d}t = \mathbb{E}[\varphi(Y)]$$

by monotonicity of the Lebesgue integral. Hence (ii) implies (iii).

Suppose (iii). Fix arbitrary $x \in \mathbb{R}$ and define $\varphi(z) = 1_{z \geq x}$, an increasing function. Then $\mathbb{E}[\varphi(Y)] = 1 - F_Y(x) \geq 1 - F_{Y'}(x) = \mathbb{E}[\varphi(Y')]$. Thus $F_Y(x) \leq F_{Y'}(x)$. Since x is arbitrary, it follows that $F_Y(x) \leq F_{Y'}(x)$ for all $x \in \mathbb{R}$, establishing (i). $\square$

Definition 24 (Second order stochastic dominance)**.** Let $(\Omega, \Sigma, \mathbb{P})$ be a probability space. Let $\mathcal{Y}$ be a set of random variables $Y : \Omega \to \mathbb{R}$. For each random variable Y, let F_Y denote the cdf of Y.

For each $Y, Y' \in \mathcal{Y}$, define $Y \geq Y'$ if

$$\int_{-\infty}^{x} F_Y(t) \, \mathrm{d}t \leq \int_{-\infty}^{x} F_{Y'}(t) \, \mathrm{d}t \qquad \text{for all } x \in \mathbb{R}.$$

Then $\geq$ is the *second order stochastic dominance* order on $\mathcal{Y}$.

If a lottery μ second order stochastically dominates a lottery μ', then μ is "less risky" than μ'. Equivalently, if random variable Y second order stochastically dominates Y', then Y is "more precise" than Y'. This is not obvious from Definition 24 but (ii) in the following proposition clarifies what we mean by "more precise" or "less risky".

Proposition 18. *Let $(\mathcal{Y}, \geq)$ denote a set of integrable random variables on probability space $(\Omega, \Sigma, \mathbb{P})$ partially ordered by the second order stochastic dominance order $\geq$. For any two random variables $Y, Y' \in \mathcal{Y}$, the following are equivalent:*

(i) $Y \geq Y'$;

(ii) *there exists a random variable $Z : \Omega \to \mathbb{R}$ such that $Y' = Y + Z$ where $\mathbb{E}[Z \mid Y] = c$ for some constant $c \leq 0$;*

(iii) *for all increasing concave functions $\varphi : \mathbb{R} \to \mathbb{R}$,*

$$\mathbb{E}[\varphi(Y)] \geq \mathbb{E}[\varphi(Y')].$$

Proof. Suppose (i). $F_Y(x) = P(Y \leq x)$. $F_{Y'}(x) = P(Y + Z \leq x)$.

Suppose (ii). We have

$$\begin{aligned}\mathbb{E}[\varphi(Y')] &= \mathbb{E}[\mathbb{E}[\varphi(Y+Z) \mid Y]] \\ &\leq \mathbb{E}[\varphi(\mathbb{E}[Y+Z \mid Y])] = \mathbb{E}[\varphi(Y + \mathbb{E}[Z \mid Y])] \leq \mathbb{E}[\varphi(Y)],\end{aligned}$$

where the first inequality holds from conditional Jensen's inequality and the second inequality follows since φ is increasing and $\mathbb{E}[Z \mid Y] \leq 0$. Hence (ii) implies (iii).

Now suppose (iii). □

2.2 Acts, states and subjective expected utility

When it comes to subjective approaches to uncertainty, it is natural to imagine that people (should) think carefully about the aspects of the world relevant to the decision problem they face. In the words of Savage (1972), the *world* is "the object about which the person is concerned", a *state of the world* is "a description of the world, leaving no relevant aspect undescribed", and the true state of the world is "the true description of the world" (the state that actually obtains).

To put this more concretely, there are many things you might be uncertain about. You are (probably) uncertain about what the complete decimal expansion of π is, or what the complete past, present and future history of the universe is. You might also be uncertain about the quality of a piece of fruit you are considering buying at a market, or whether it will be raining at precisely 8.30am next Tuesday during your morning run.

When it comes to making a decision in some context, how you should conceive of a "state of the world" depends on what is relevant. To take a classic example from Savage (1972), suppose you are making a 6 egg omelette. You have plenty of eggs. You are uncertain whether any particular egg (or any combination of them) might be rotten. After cracking at least one good egg into the bowl, you could crack subsequent eggs directly into a bowl or crack each onto a separate plate first to check whether it is fresh before adding it to the bowl (this means more washing up). If, with five fresh eggs in the bowl, you crack a sixth directly into the bowl, then you have spoiled the five good eggs. On the other hand, if it is fresh, you can make your omelette and have saved a little washing up. In this scenario, the full decimal expansion of π is not relevant. Whether the egg is good or bad is relevant. The complete history of the universe in theory includes this information, but also an incredible amount of information that is not relevant.

2.2.1 Functions from acts to consequences

Von Neumann-Morgenstern expected utility theory does not involve explicit consideration of states of the world. The lotteries $\hat{\mu} \in \Delta(X)$ in the von Neumann-Morgenstern setting are probability measures over the set of consequences X, and these lotteries (or perhaps some subset of them) are the set of alternatives a decisionmaker faces.[25]

The language around the von Neumann-Morgenstern theory points to its original motivation – decisionmaking about *objective* gambles such as roulette, wagers on the outcome of coin flips, lottery tickets, and so on. In general, decisionmakers face uncertainty in a far broader range of contexts, and this uncertainty cannot reliably be quantified (in some objective sense) by the decisionmaker.

It would not change any of the mechanics of the von Neumann-Morgenstern theory to imagine that lotteries $\hat{\mu} \in \Delta(X)$ are potential *subjective* probability assessments. Such a model works as follows. The decisionmaker faces the problem of choosing an *act* $f \in F$, and there is uncertainty in how acts translate into consequences. The decisionmaker attaches a probability measure $\hat{\mu}_f \in \Delta(X)$ to each act $f \in F$. We interpret this probability measure as reflecting the beliefs of the decisionmaker about how likely various consequences are on selecting act f. Given the set of acts and a space of lotteries $L \supseteq \{\hat{\mu}_f \mid f \in F\}$, if the decisionmaker's preferences over L have Bernoulli utility representation $u : X \to \mathbb{R}$, these preferences over lotteries Limply a preference relation $\succ$ over acts F whereby for all $f, g \in F$,

$$f \succ g \quad \text{iff} \int_X u \, \mathrm{d}\hat{\mu}_f > \int_X u \, \mathrm{d}\hat{\mu}_f.$$

Here, we could imagine the decisionmaker thinks about how acts translate into distributions over consequences. Instead, we might imagine that acts map directly to consequences, and the decisionmaker's beliefs are reflected in a distribution over these mappings. This motivates *a* state approach (but not the one we will use!). Call a function $s : F \to X$ a *state*. A state s assigns to the act f the consequence $s(f) \in X$. We assume that at the point at which the decisionmaker makes the decision, there is some "true" state $\bar{s}$ and the act f causes the consequence $\bar{s}(f)$. In this case, we say that $\bar{s}$ *obtains*, and we write $T \subseteq S$ to mean that some $s \in T$ obtains.

[25]The hat on $\hat{\mu}$ is because I will later distinguish between measures $\hat{\mu}$ on X, measures $\tilde{\mu}$ on a set of acts A, and measures μ on a set of states S.

The space $S = X^F$ (the set of all functions $s : F \to X$) is then a *state space*. Equipping S with a σ-algebra $\tilde{\Sigma}$, we can represent a decisionmaker's beliefs over states with a probability distribution $\tilde{\mu}$ on S.

If the decisionmaker has a belief over states represented by a probability distribution $\tilde{\mu}$ on $(S, \tilde{\Sigma})$, then for each act $f \in F$, we can represent the distribution of consequences the decisionmaker associates to the act by defining the probability measure

$$\hat{\mu}'_f(E) := \tilde{\mu}(\{s \in S \mid s(f) \in E\}),$$

for all $E \in \hat{\Sigma}'_f := \{E' \subseteq X \mid \{s \in S \mid s(f) \in E'\} \in \tilde{\Sigma}\}$ (each $\hat{\Sigma}'_f$ is a σ-algebra).[26]

In view of the description of *world* and *state* above, the formulation of states we just gave is not unreasonable. However, this conception of states is not the one used by Savage (1954) or most other approaches to subjected expected utility. This said, it is a useful formulation, in particular because it gives us clarity about what we should include in state spaces when we have a natural set of actions in a decision problem. Gilboa (2009) discusses this in the context of some famous puzzles, one of which we will spell out here.

Example 16 (Newcomb's problem)**.** Newcomb's problem (Nozick, 1969) is the following decision problem. A clairvoyant presents you with two boxes, one opaque and one transparent. In the transparent box is a stack of dollar bills that amount to \$1000. In the opaque box, there may or may not be \$1 million. You have the option to take one of the boxes or both. The clairvoyant almost always correctly predicts whether you will take one or both boxes. Moreover, she is a moralistic character with a disdain for greed, so if she predicted you will take both boxes, then she will not put the \$1 million in the opaque box, whereas if she predicted you will take one box only, she puts in the \$1 million. The relevant actions are whether to take both boxes (be *greedy*, G) or just the opaque box (be *modest*, M).[27]

In typical analyses of Newcomb's problem, we end up with the following payoff matrix:

	$\$1,000,000$	$\$0$
G	$1,001,000$	1000
M	$1,000,000$	0

The rows correspond to the acts G and M and the columns correspond to the dollar amount in the opaque box. This payoff matrix suggests a strict dominance relation: you are better off being greedy and taking both boxes, because it yields you a strictly higher payoff irrespective of the content of the opaque box.

On the other hand, since the probability that the clairvoyant has made a correct prediction is 1, the expected payoff from being greedy and taking both boxes is merely

[26] You might worry this gives us a different σ-algebra in X for each act. We can always take a σ-algebra $\hat{\Sigma}'$ that includes each $\hat{\Sigma}'_f$, and extend each μ_f to $\hat{\Sigma}'$, resolving this worry. I am at pains to work with σ-algebras here. The more classic approach works with finitely additive probabilities and would define $\hat{\mu}'_f$ for all $E \subseteq X$.

[27] It is unambiguous that taking only the transparent box is worse than taking both. There are no carrying costs.

\$1000, whereas the expected payoff from being modest is \$1 million. This would suggest you are (much) better off being modest.

The trick here is that the problem is missing details that leave it ambiguous whether you have any control over the content of the opaque box. The dominance argument relies on the assertion you have no control over the content of the boxes. This would be reasonable in the real world, where magical clairvoyants (or sci-fi ones, such as predictions from perfect machine simulations of a given person's consciousness) do not exist. However, if we do face a clairvoyant, then there is reason to believe we do have control over the content of the opaque box. Thinking of states as functions from acts to consequences ensures we capture this. There are 2 acts and 4 consequences, and thus $4^2 = 16$ possible states, but because the consequences are all mediated through whether the opaque box is empty or not, we need only consider $2^2 = 4$ states at the point of making a decision. These are:

s_1: the opaque box contains \$1 million regardless of your choice G or M;

s_2: the opaque box contains \$1 million only if you choose G;

s_3: the opaque box contains \$1 million only if you choose M;

s_4: the opaque box is empty regardless of your choice G or M.

This gives us payoff matrix:

	s_1	s_2	s_3	s_4
G	$1,001,000$	$1,001,000$	1000	1000
M	$1,000,000$	0	$1,000,000$	0

We see from the new payoff matrix that G is not strictly dominant. Rather, whether G is optimal depends on the decisionmaker's belief over states. Say, before you take part, you have seen a large number of modest and greedy people face off against the clairvoyant. Gilboa (2009) reasons that if you have observed that every greedy person before you emerged with \$1000 and every modest person emerged with \$1 million, then you should place weight very close to 1 on the state being s_3 when you make the choice – certainly, if you are a Bayesian starting with a prior with full support, this would be your conclusion. Thus it would pay off to be modest.

2.2.2 Functions from states to consequences

Before, we took the set of acts F as given and derived a state space S. Savage (1954) instead takes the state space S as given, and derives the set of acts F.[28] In this case, an *act* is a measurable function $f : S \to X$, mapping each state $s \in S$ to a consequence $f(x)$, and we define $F = X^S$ to be the set of all such functions.

[28]Since we will be defining a probability measure on it, S is a measurable space with σ-algebra Σ. Or if you don't mind finite additivity of your measures, you can weaken Σ to be an algebra.

For example, in Savage's omelette example, the issue was whether the sixth egg we are about to crack into the bowl or onto the plate is good or bad. The state space $S = \{s_1, s_2\}$ then has two states, describing whether the egg is good (s_1) or bad (s_2). A juror dutifully choosing whether a defendant should be convicted must make a judgement about whether the defendant committed the crime (s_1) or did not (s_2).

We assume that in any decision problem, the state space is specified so that precisely one state obtains. We cannot have both (s_1: it rains tomorrow) and (s_2: it rains tomorrow morning) in the state space for example. This is easily remedied by removing s_2 or changing s_1 to "it rains tomorrow afternoon".

Moreover, we assume the state space is complete in the sense that it describes all possible revelant states the decisionmaker conceives of for the decision problem. Thus the decisionmaker's beliefs about the state of the world can be summarized by a probability measure μ on S. This rules out an interesting possibility that a decisionmaker can conceive that there are things they have not conceived of – we know there are "unknown unknowns", as Donald Rumsfeld once put it.

Given μ on (S, Σ) summarizing the decisionmaker's beliefs about states, we can represent the distribution the decisionmaker associates to each act $f \in F$ by defining the probability measure

$$\hat{\mu}_f^*(E) := \mu(\{s \in S \mid f(s) \in E\}),$$

for all $E \in \hat{\Sigma}_f^* := \{E' \subseteq X \mid \{s \in S \mid f(s) \in E'\} \in \Sigma\}$.[29]

We could work directly with the $\hat{\mu}_f^*$s, following the von Neumann-Morgenstern theory. However, state-of-the-world expected utility theories like Savage's instead aim to relate a decisionmaker's preference relation $\succ$ over the set of acts A to an expected utility calculation over the set of states. Namely, we search for an expected utility representation of the form

$$f \succ g \quad \text{iff} \quad \int_S u(f(s))\mu(\mathrm{d}s) > \int_S u(g(s))\mu(\mathrm{d}s) \qquad \text{for all } f, g \in F.$$

Recall we now have two different formulations of uncertainty, the first endogenizing states S taking acts F as given and the other endogenizing acts F taking states S as given. These two approaches are in fact closely related.

To make this point, let us denote the sets of states and acts in the first approach by S' and F', to distinguish them from S and F in the second formulation. Take a σ-algebra $\hat{\Sigma}$ of X. Now assume there is an isomorphism $\psi : f \mapsto f'$ so that

$$\hat{\mu}_f^*(E) = \hat{\mu}'_{\psi(f)}(E)$$

for all $E \in \hat{\Sigma}$ and all $f \in F$. Essentially, this is to assume each act $f \in F$ in the second formulation has an 'equivalent' act $f' \in F'$ in the first formulation, where 'equivalent'

[29] Again, we can extend these measures so they are defined on a σ-algebra $\hat{\Sigma}^*$ that includes all $\hat{\Sigma}_f^*$.

is in terms of inducing the same probabilities over consequences.[30] Then for any $\hat{\Sigma}$-measurable utility function $u : X \to \mathbb{R}$ and every $f \in F$, we have

$$\int_S u(f(s))\mu(\mathrm{d}s) = \int_X u \,\mathrm{d}\hat{\mu}^*_f = \int_X u \,\mathrm{d}\hat{\mu}'_{\psi(f)} = \int_{S'} u(s'(\psi(f)))\tilde{\mu}(\mathrm{d}s').$$

2.2.3 Subjective expected utility representations

As with von Neumann-Morgenstern expected utility theory, we would like to have a utility representation U of preferences $\succ$, now defined on the set of acts F, that can be written as an expectation of some Bernoulli utility function $u : X \to \mathbb{R}$ defined on consequences. The issue with *subjective* expected utility is that there are no lotteries. The probabilities are subjective, representing the decisionmaker's beliefs about the likelihood of different events in the state space. The preference relation $\succ$ thus contains information about *both* the decisionmaker's risk preferences on $\Delta(X)$ and the decisionmaker's beliefs about events in the state space. We need to be able to disentangle these to get a subjective expected utility representation. Savage (1954) famously gives us conditions under which this disentanglement is possible.

Note we state the following definition being agnostic about whether μ is a *probability measure* or a *probability charge*. A *charge* is a finitely additive probability measure, whereas whenever I describe μ as a "measure" I assume μ is countably additive. For historical reasons, finite additivity has a (misplaced) emphasis in parts of decision theory.

Definition 25 (Subjective expected utility)**.** Let X be a space of consequences, let S be a state space, let Σ be an algebra on S, and let F be a set of acts $f : S \to X$. If $\succ$ is a weak order on F, we say that $\succ$ has a *subjective expected utility representation* $U : F \to \mathbb{R}$ if there exists a function $u : X \to \mathbb{R}$ and a unique probability measure or charge $\mu : \Sigma \to [0, 1]$ such that $U(f) = \int_S u(f(s))\mu(\mathrm{d}s)$ and

$$f \succ g \quad \text{iff} \quad \int_S u(f(s))\mu(\mathrm{d}s) > \int_S u(g(s))\mu(\mathrm{d}s) \qquad \text{for all } f, g \in F.$$

We call the function $u : X \to \mathbb{R}$ in this representation a *Bernoulli utility function* representing $\succ$ on $\Delta(X)$.

Often, applied theorists call u "preferences" and μ "beliefs". As Strzalecki points out in *Decision Theory*, calling u "preferences" is a bit sloppy – the preferences $\succ$ here are over acts and contain information both on risk preferences over $\Delta(X)$ and beliefs about events. This said, if preferences $\succ$ have a subjective expected utility representation then we can draw a distinction between the decisionmaker's *tastes* – their risk preferences – and their beliefs. This distinction is enormously helpful in applied theory, because it means when a decisionmaker is exposed to new information in some setting, we do not need to worry about the decisionmaker's risk preferences on consequences changing.

[30] In comparing the two formulations, Fishburn (1970) treats an act f as though it is a common object to both formulations. This seems confused to me, but I am less smart than Fishburn.

A few brief technical comments are in order. First, in Definition 25, we specified that the probability measure/charge μ should be unique. Since the programme Savage is following is attempting to establish conditions under which we can infer a decision-maker's beliefs and tastes from their choices, if μ is non-unique then such an inference is not possible since different belief-taste combinations could justify the same complete set of choices. Really the only case uniqueness of μ rules out is one where the decision-maker is totally indifferent over everything, in which case their choices will be totally uninformative about their beliefs.

Second, in Definition 25 we allowed the set of events Σ on S to be an algebra. Modern probability theory works with countably additive probability measures, in which case Σ must be a σ-algebra. Since any σ-algebra is an algebra, our definition nests this and is a little more flexible. Savage (1954) and others worked with finitely additive probability charges, but Savage's own representation theorem requires that Σ is a σ-algebra.

2.2.4 Extraneous probabilities

There are practical reasons to prefer *subjective* conceptions of uncertainty, but there are also philosophical reasons to do so. There are two meanings for "probability". The original meaning is to do with plausibility, and the second is the notion of *objective chance*. The plausibility meaning is the one we use more often in everyday use. If you ask me if it will be sunny tomorrow morning and I respond "probably", I do not mean that I know that the physical chance that it is sunny tomorrow morning exceeds one half. Instead what I mean is that I am pretty confident it will be sunny, i.e. I *judge* it is more-likely-than-not that it will be sunny.

Because we are interested in decisionmakers, who are necessarily subjective, subjective probability is the natural approach. Some events, however, are well-described by the theory of chances. This includes randomization devices, such as roulette wheels, randomly shuffled card decks, dice, the time to decay of a particular atom of carbon-14, and so on. This is why von Neumann & Morgenstern's approach makes some sense, after all. We call the chances attached to the outcomes such devices generate *extraneous probabilities*.

Modelling extraneous probabilities is particularly useful because observing a decisionmaker's choices over lotteries associated with these devices allows us to 'calibrate' their beliefs about events that are not so easily quantified. Anscombe & Aumann (1963) is the most famous paper to do this, but the approach was also explored by Chernoff (1954), Suppes (1956) and Pratt, Raiffa & Schlaifer (1964), among others.

As Anscombe & Aumann (1963) point out, we can think of extraneous probabilities as subjective probabilities that we (i.e. the decisionmaker and the analyst) all agree on. Thus we need not depart from the subjective probability framework at all.

2.2.5 Small worlds

One of the great downsides of thinking about states of the world is that in complicated enough decision problems, thinking about *every* relevant state is not very plausible. Yet this is the approach we almost always take in applied theory, regardless of complexity. Savage calls the notion of forward-looking contingent planning the "look before you leap" principle. He is very open that there is no more general truth to the principle than to another proverb – "you can cross that bridge when you come to it." In his words (Savage, 1972, p. 16):

> Carried to its logical extreme, the "Look before you leap" principle demands that one envisage every conceivable policy for the government of his whole life (at least from now on) in its most minute details, in the light of the vast number of unknown states of the world, and decide here and now on one policy. This is utterly ridiculous, not – as some might think – because there might later be cause for regret, if things did not turn out as had been anticipated, but because the task implied in making such a decision is not even remotely resembled by human possibility. It is even utterly beyond our power to plan a picnic or to play a game of chess in accordance with the principle, even when the world of states and the set of available acts to be envisaged are artificially reduced to the narrowest reasonable limits.
>
> Though the "Look before you leap" principle is preposterous if carried to extremes, I would nonetheless argue that it is the proper subject of our further discussion, because to cross one's bridges when one comes to them means to attack relatively simple problems of decision by artificially confining attention to so small a world that the "Look before you leap" principle can be applied there. I am unable to formulate criteria for selecting these small worlds and indeed believe that their selection may be a matter of judgment and experience about which it is impossible to enunciate complete and sharply defined general principles.

There is thus a legitimate – and not often studied – question of how a decisionmaker chooses the "small world" in subjective expected utility theory. As Savage's discussion suggests, in many problems the set of states we end up considering are going to form a very partial description of the decision problem. This raises further issues. For example, it is not often obvious how to handle *residual uncertainty* – if decisionmakers know they have probably failed to isolate every relevant state, they know that they may face unforeseen contingencies, and presumably should account for that in their behaviour. In applications, this is perhaps less of a problem than you might initially think, as long as decisionmakers can form forecasts about payoffs directly. For example, suppose you are considering making a risky long-term investment. There are many factors that might influence the performance of the investment, many of which you are unlikely to conceive of ahead of time. However, while you cannot easily foresee all the possible contingencies that might affect the success of your investment, you probably do have beliefs about how likely different payoffs might be. In this case, we can just reason about the payoff space. Maskin & Tirole (1999) make this suggestion in the context of incomplete contracts.

Ironically, the world in Savage's own theory is necessarily large – his assumptions imply the state space is infinite – though we can accommodate finite state spaces in other variations of subjective expected utility theory, such as Anscombe-Aumann. Gül (1991) develops the finite-state version of Savage's theory.

2.3 Savage's subjective expected utility theory

The approach I am going to take here differs quite a bit from most expositions of Savage's subjective expected utility theory, as found in Savage's own *Foundations of Statistics* (1954, 1972), Fishburn (1970), Kreps (1988, 2013), or various sets of decision theory notes and lectures. There are two problems with the original theory.

The first problem is that Savage, like de Finetti, had views about formalizing probability theory that are pretty heterodox these days. He took probability measures to be finitely additive but not countably additive. I detail why this is a bad idea in the appendix. Villegas (1964) and Arrow (1974) reformulated subjective expected utility theory for countably additive probability, in line with standard modern treatments of formal probability theory.

The second problem is that Savage went about proving his representations in a needlessly complicated way. This is because he relied on von Neumann and Morgenstern's mixture approach. However, the domain of uncertainty in subjective expected utility, where no events are associated with a probability in the "objective" sense, is not at all like the domain of risk in von Neumann-Morgenstern expected utility theory, where all events are associated with an "objective" probability. Transforming the domain of uncertainty in subjective expected utility into the domain of risk requires richness axioms that are stronger than necessary. This just serves to complicate proofs and has made Savage's theory a "black box" to most users of it.

Abdellaoui & Wakker (2020) show that we can use weaker axioms and a far simpler proof if we do not follow Savage (or Villegas or Arrow) in using the mixture approach. We follow their approach here. This approach does allow for finite additivity, as in Savage's own version of theory, but also for countable additivity with an additional axiom.

2.3.1 Savage's axioms

Savage's theory requires a lot of axioms – seven in total, although it turns out one of them is redundant. Before stating the axioms, we should say more about the setting, and define some notation.

As a reminder, we have a given set of states S and a set of consequences X. We endow S with an algebra Σ, the elements of which are called *events*. The algebra Σ is *not* necessarily a σ-algebra, but it does no harm to require that Σ is a σ-algebra. An act is a measurable function $f : S \to X$.

Definition 26.

(a) *Simple act.* We call an act f a *simple act* if it is a measurable function taking finitely many values; otherwise, f is called a *non-simple act.*

(b) *Constant act.* A *constant act* is a constant function $f : S \to X$, i.e. $f(s) = x$ for all $s \in S$ and some $x \in X$. We use x to denote the constant act that takes value $x \in X$ everywhere on S.

(c) *Composite act.* For each event $E \in \Sigma$ and for every pair of acts $f, g \in F$, we define the act $f_E g$ by

$$f_E g(s) = \begin{cases} f(s) & \text{if } s \in E, \\ g(s) & \text{if } s \in E^c. \end{cases}$$

This is not typically given a name, but we call $f_E g$ a *composite act of f and g partitioned wrt E.*

When we write $x_E f$, for example, we have $x_E f(s) = x$ on E and $x_E f(s) = f(s)$ on E^c.

Lemma 11. *For any pair of acts $f, g \in F$ and any event $E \in \Sigma$, we have $f_E g = g_{E^c} f$.*

Proof. Fix $s \in S$. If $s \in E$ then $f_E g(s) = f(s) = g_{E^c} f(s)$. Similarly, if $s \notin E$ then $f_E g(s) = g(s) = g_{E^c} f(s)$. ☐

For the purposes of stating the axioms, we will assume that $F = X^S$. Savage's theory requires that F contains all constant acts and is closed under the composite act operation for all $E \in \Sigma$, which taking $F = X^S$ obviously ensures. The assumption that F includes all constant acts is sometimes criticized for requiring that the entire space of consequences is relevant under every state.

We have a strict preference relation $\succ$ over the set of acts F.

Definition 27.

(a) *Null events.* We call an event $E \in \Sigma$ *null* if for all acts $f, g, h \in \Sigma$, we have $f_E h \sim g_E h$.

(b) *Conditional preference.* Given an event $E \in \Sigma$ and acts $f, g \in F$, we say *$f \succ g$ given E*, denoted $f \succ_E g$ if for any $h \in F$, the restrictions

$$f'(s) = \begin{cases} f(s) & \text{if } s \in A, \\ h(s) & \text{if } s \notin A \end{cases} \quad \text{and} \quad g'(s) = \begin{cases} g(s) & \text{if } s \in A, \\ h(s) & \text{if } s \notin A \end{cases}$$

satisfy $f' \succ g'$.

These are the axioms that Savage (1954) used:

Axioms. Let S be a set of states, let X be a set of consequences equipped with algebra Σ, and let $F = X^S$ be the set of all acts $f : S \to X$. Let $\succ$ be a strict preference relation on F.

(P1) *Weak order.* The relation $\succ$ is a weak order on F.

(P2) *Sure-thing principle.* For all acts $f, g, h.h' \in F$ and all events $E \in \Sigma$, $f_E h \succ g_E h$ iff $f_E h' \succ g_E h'$.[31]

(P3) *Event monotonicity.* For every nonnull event $E \in \Sigma$, every act $f \in F$ and all constant acts $x, y \in F$, we have

$$x_E f \succsim y_E f \quad \text{iff} \quad x \succsim y.$$

(P4) *Independence of beliefs from tastes.* For all constant acts $x, x', y, y' \in F$ that satisfy $x \succ y$ and $x' \succ y'$ and for all events $E_1, E_2 \in \Sigma$, we have

$$x_{E_1} y \succ x_{E_2} y \quad \text{iff} \quad x'_{E_1} y' \succ x'_{E_2} y'.$$

(P5) *Nontriviality.* There is a pair of acts $f, g \in F$ such that $f \succ g$.

(P6) *Event continuity.* For all acts $f, g \in F$ with $f \succ g$ and for all constant acts $x \in F$, there is a partition $\mathcal{P} = \{E_1, \ldots, E_n\} \subset \Sigma$ of S such that $x_{E_i} f \succ g$ and $f \succ x_{E_i} g$ for all $E_i \in \mathcal{P}$.

(P7) *Conditional monotonicity.* For all acts $f, g \in F$ and all events $E \in \Sigma$:

(i) if $f \succsim_E g(s)$ for all $s \in E$, then $f \succsim_E g$, and

(ii) if $f(s) \succsim_E g$ for all $s \in E$, then $f \succsim_E g$.[32]

These all deserve a bit more explanation. Pay attention to the discussion on (**P6**) because we break (**P6**) into two axioms, one of which we'll actually use in the Abdellaoui-Wakker approach, in place of (**P6**).

(i) (**P1**) does not need further explanation. We're used to weak orders, and they are not immune from criticism.

(ii) The *sure-thing principle* (**P2**) says that if two acts f and g differ on an event E (i.e. $f \neq g$ on E) but are the same elsewhere ($f = g$ on E^c), then your preference over f and g never depends on the event E^c. This is essentially exactly the same as separability (**S3**) in section 1.4.2, the only difference being the slight difference in setting.[33]

As a historical note, while (**P2**) is often called the sure-thing principle, Savage's definition of the sure-thing principle is not captured by any one axiom. Savage (1954) describes the principle as follows. Say a real estate investor considers buying some property, and the outcome of the upcoming presidential election is relevant to the decision. It is certain that the winner will be either the Democratic candidate or the Republican. If she would buy knowing for sure that the Democratic candidate

[31] By Lemma 11, we also have $h_E f \succ h_E g$ iff $h'_E f \succ h'_E g$.

[32] Here, $f(s)$ and $g(s)$ are constant acts.

[33] The algebra Σ of events does not need to be the power set of X (though it could be), whereas in (**S3**), we had the condition holding for all $E \in 2^I$.

will win and she would buy knowing for sure that the Republican candidate will win, then she should buy not knowing who will win (or equivalently, she should buy regardless of the probabilities she attaches to which candidate wins). This is actually captured by a combination of axioms (**P2**), (**P3**) and (**P7**).

(iii) Event monotonicity (**P3**) is one of the two state-independence axioms. It implies that if we take an act and we modify it over a nonnull event to yield a consequence that a decisionmaker prefers over the previous consequences, then the decisionmaker prefers the new act to the original.

It took over 60 years for anyone to realize this axiom is actually redundant! We will end up needing the axiom at first because we won't be using (**P7**), but with (**P7**), Hartmann (2020) proves the other axioms imply (**P3**). I include the proof because it is quite short.

Proposition 19 (Hartmann, 2020)**.** *If $\succ$ satisfies* (**P1**)*,* (**P2**)*,* (**P4**) *and* (**P7**)*, then $\succ$ satisfies* (**P3**)*.*

Proof. Suppose $\succ$ satisfies axioms (**P1**), (**P2**) and (**P7**). First, we claim that if we have $x \succsim y$ for constant acts $x, y \in F$ then we have $x \succsim_E y$ for all events $E \in \Sigma$. To see this, fix any $E \in \Sigma$ and any constant acts x, y with $x \succsim y$. Since $x \succsim y(s)$ for every state $s \in S$, we have $x \succsim (y_E x)(s)$ for every $s \in S$. Noting $\succsim = \succsim_S$ and applying (**P7**)(i) thus gives $x \succsim y$.

Next, we claim that for every nonnull event $E \in \Sigma$, there are constant acts x, y with $x \succ_E y$ (in fact, $x \succ y$). To see this, take a nonnull event E. By definition, there are acts f, g with $f \succ_E g$. Given $\succsim$ is connected (complete), (**P7**)(ii) implies there is a state $s \in S$ such that $f \succ_E g(s)$ (if there was no such state, (**P7**)(ii) would imply $g \succsim_E f$, a contradiction). Likewise, (**P7**)(i) implies there is a state $s' \in S$ such that $f(s') \succ_E g(s)$. Taking $x = f(s')$ and $y = g(s)$ gives $x \succ_E y$. Now, if $x \not\succ y$, then connectedness of $\succsim$ implies $y \succsim x$, but from the first paragraph, this would imply $y \succsim_E x$, yielding a contradiction. Thus $x \succ y$.

Now suppose $\succ$ also satisfies (**P4**). Take any act $f \in F$, constant acts $x, y \in F$ and a nonnull event $E \in \Sigma$. Since $\succsim$ is connected, we need only show $x \succsim y$ implies $x_E f \succsim_E y_E f$ and $x \succ y$ implies $x_E f \succ_E y_E f$. If $x \succsim y$ then we have $x \succsim_E y$, and so by (**P2**), $x_E f \succsim y_E f$. Now suppose $x \succ y$. Given E is nonnull, there are constant acts x', y' with $x' \succ_E y'$, which by (**P2**) gives $x'_E y' \succ y'$. Moreover, $x' \succ y'$, and applying (**P4**) with $E_1 = E$ and $E_2 = \varnothing$ yields $x_E y \succ y$. By (**P2**), it follows that $x_E f \succ y_E f$. $\square$

(iv) Independence of beliefs from tastes (**P4**) is another state-independence axiom. Intuitively, this one implies that changing the stakes does not change which way you bet – if you prefer to put \$100 on Secretariat beating Tentam over putting \$100 on Tentam beating Secretariat, then you should also prefer to put \$1000 on Secretariat beating Tentam over putting \$1000 on Tentam beating Secretariat.

Given the decisionmaker's preferences over acts, this axiom lets us distinguish between the decisionmaker's probability ranking of events and their preference ranking over consequences.

(v) Nontriviality (**P5**) simply rules out that the decisionmaker is everywhere indifferent between any two acts. Total indifference is not a particularly interesting case anyway. The issue is that if the decisionmaker is totally indifferent, then we are not going to be able to pin down the decisionmaker's beliefs about the probability of various events from their preferences over acts.

(vi) Event continuity (**P6**) says that for any two acts, we can partition the state space sufficiently finely that changing the values of these acts on any single event in this partition does not alter the decisionmaker's preferences over the two acts.

This axiom requires the state space S to be infinite – if S is finite, then (**P6**) will imply that every state $s \in S$ is null. Given Savage (1972) contains a nuanced discussion about small worlds, it is a little odd that his theory requires the decisionmaker to keep track of an infinite set of possibilities! The version of the theory we will present allows a finite set of states in some circumstances.

(**P6**) is not itself a necessary condition for subjective expected utility representations – rather it is a technical "richness" condition that prevents the already-complex proof becoming much more complex. However, it does contain contain within it a necessary condition, captured by (**P6.1**) below.

Abdellaoui & Wakker (2020) weaken (**P6**) slightly. Just as they do, it is convenient to split the axiom up into the two axioms, the first giving a necessary condition and one giving the unnecessary-but-sufficient/convenient richness condition:

Axioms.

(**P6.1**) *Archimedeanity.* There is no infinite sequence of disjoint equally likely nonnull events.

(**P6.2**) *Event solvability.* For all acts $f, g \in F$ and all constant acts $x \succ y$, and all events $B \in \Sigma$, we have that $x_B f \succ g \succ y_B f$ implies $g \sim x_E y_{B-E} f$ for some partition $\{E, B - E\}$ of B.

Archimedeanity (**P6.1**) is an obvious necessary condition – if (**P6.1**) fails to hold, then, letting μ be the probability measure/charge representing the decisionmaker's beliefs over S, we have an infinite collection of disjoint events $\{E_i\}_{i=1}^{\infty} \subseteq \Sigma$, each with $\mu(E_i) = \mu(E_j) > 0$ for $i, j \in \mathbb{N}$. But this would imply $\mu(S) \geq \sum_{i=1}^{\infty} \mu(E_i) = +\infty$.

Event solvability (**P6.2**) is not necessary, though it is weaker than (**P6**). It implies that if a first act is not preferred to a second and we can improve the consequences of the first act so much on an event B that it the decisionmaker prefers it to the second, then there must be some subevent E of B such that improving the event on E makes the decisionmaker indifferent between the acts.

(vii) The final technical axiom (**P7**), which I am labelling "conditional monotonicity", implies that if, conditional on an event, f is weakly preferred to any possible consequence of g on that event, then f should be preferred to g conditional on the event. This axiom is unnecessary if we restrict the theory to simple acts, and only really arises because Savage wanted to work with finitely additive charges.

Savage (1954) also included another axiom that implied the Bernoulli utility function u is bounded for all acts. As Fishburn (1970, p. 194) recounts, Savage and Fishburn later realized this axiom was unnecessary, since (**P1**)-(**P7**) already imply u is bounded. The second edition of *The Foundation of Statistics* (1972) remedies this.

Finally, because finitely additive charges are incredibly inconvenient versus countable additive measures, we will be making use of another axiom due to Villegas (1964) and Arrow (1970):

Axiom.

(**P8**) *Set continuity.* For all constant acts $x, y \in F$ with $x \succ y$, for all acts $f \in F$ with $f \succ y$, and for all sequences of events $\{E_n\}_{n=1}^{\infty} \subseteq \Sigma$ with $E_n \downarrow \varnothing$, there is an N such that
$$f \succ x_{E_n} y \qquad \text{for all } n \geq N.$$

Here $E_n \downarrow \varnothing$ indicates that $\{E_n\}$ *vanishes*, i.e. $E_{n+1} \subseteq E_n$ for all n and $\bigcap_{n=1}^{\infty} E_n = \varnothing$.

Axiom (**P8**) is another technical axiom, similar in spirit to (**P6**) – it says that as events become "small enough", changing the value of two acts on these events does not alter the decisionmaker's preferences over the two acts.

Moreover, the following weaker condition, implied by (**P6**), will be helpful for proofs. Arrow (1974) used it in place of (**P6**).

Axiom.

(**P9**) *Nonatomicity.* If $E \in \Sigma$ is a nonnull event, then there are nonnull events $E_1, E_2 \in \Sigma$ such that $E_1 \cup E_2 = E$ and $E_1 \cap E_2 = \varnothing$.

If an event $E \in \Sigma$ is nonnull and *cannot* be partitioned into two disjoint nonnull events as in (**P9**), then we say E is an *atom*. Note any finite state space necessarily has atoms unless every event is null (in which case (**P5**) cannot hold).

2.3.2 Subjective expected utility for simple acts

While Savage's theory holds for all acts, it is simplest to prove for simple acts, i.e. acts that have a finite range. The proof is very long, even though we follow the "simplified" version of Abdellaoui & Wakker (2020).

Theorem 14 (Abdellaoui & Wakker, 2020)**.** *Let X be a space of consequences, let S be a state space equipped with an algebra Σ, and let F be the space of all simple acts $f : S \to X$. Let $\succ$ be a strict preference relation on F. Moreover, assume event solvability* (**P6.2**) *holds. Then:*

(i) $\succ$ *has a subjective expected utility representation with a (finitely additive) probability charge iff* (**P1**)-(**P5**) *and* (**P6.1**) *hold.*[34]

(ii) $\succ$ *has a subjective expected utility representation with a (countably additive) probability measure iff* (**P1**)-(**P5**), (**P6.1**) *and* (**P8**) *hold.*

Moreover, if $U(f) = \int_S u(f)\,\mathrm{d}\mu_1$ *and* $V(f) = \int_S v(f)\,\mathrm{d}\mu_2$ *are subjective expected utility representations of* $\succ$, *then* $\mu_1 = \mu_2$ *and* v *is a positive affine transformation of* u.

Proof. First we argue (**P1**)-(**P5**) necessary (we already argued (**P6.1**) is necessary when we introduced it). Suppose strict preference relation $\succ$ has a subjective expected utility representation $U(f) = \int_S u(f(s))\mu(\mathrm{d}s)$.

(i) To show (**P1**) holds, we need only show $\succ$ is negatively transitive. Take $f, g, h \in F$ with $f \succ h$. Then $U(f) > U(h)$. Since $>$ is a weak order on $\mathbb{R}$, either $U(g) > U(h)$ or $U(f) > U(g)$. Hence either $f \succ g$ or $g \succ h$, and so $\succ$ is negatively transitive, so (**P1**) holds.

(ii) Fix acts $f, g, h, h' \in F$ and event $E \in \Sigma$. Suppose $f_E h \succ g_E h$. Then

$$U(f_E h) = \int_E u(f)\,\mathrm{d}\mu + \int_{E^c} u(h)\,\mathrm{d}\mu > \int_E u(g)\,\mathrm{d}\mu + \int_{E^c} u(h)\,\mathrm{d}\mu = U(g_E h).$$

Thus $\int_E u(f)\,\mathrm{d}\mu > \int_E u(g)\,\mathrm{d}\mu$, and so

$$U(f_E h') = \int_E u(f)\,\mathrm{d}\mu + \int_{E^c} u(h')\,\mathrm{d}\mu > \int_E u(g)\,\mathrm{d}\mu + \int_{E^c} u(h')\,\mathrm{d}\mu = U(g_E h'),$$

giving $f_E h' \succ g_E h'$. Hence (**P2**) holds.

(iii) Fix a nonnull event $E \in \Sigma$, an act $f \in F$ and constant acts $x, y \in F$. Suppose $x_E f \succsim y_E f$. Then

$$U(x_E f) = \int_E u(x)\,\mathrm{d}\mu + \int_{E^c} u(f)\,\mathrm{d}\mu \geq \int_E u(y)\,\mathrm{d}\mu + \int_{E^c} u(f)\,\mathrm{d}\mu,$$

so $\int_E u(x)\,\mathrm{d}\mu \geq \int_E u(y)\,\mathrm{d}\mu$. Since $u(x)$ and $u(y)$ are constants, it follows that $u(x) \geq u(y)$, and thus $x \succsim y$. Thus (**P3**) holds.

(iv) Fix constant acts $x, x', y, y' \in F$ with $x \succ y$ and $x' \succ y$. Then $u(x) > u(y)$ and $u(x') > u(y')$. Fix events $E_1, E_2 \in \Sigma$ and suppose $x_{E_1} y \succ x_{E_2} y$. Then

$$\begin{aligned} \mu(E_1)u(x) + (1 - \mu(E_1))u(y) &= \int_{E_1} u(x)\,\mathrm{d}\mu + \int_{E_1^c} u(y)\,\mathrm{d}\mu = U(x_{E_1} y) \\ &> U(x_{E_2} y) = \int_{E_2} u(x)\,\mathrm{d}\mu + \int_{E_2^c} u(y)\,\mathrm{d}\mu \\ &= \mu(E_2)u(x) + (1 - \mu(E_2))u(y). \end{aligned}$$

[34] Note Proposition 19 does not apply: because (**P7**) is missing, we need (**P3**).

This implies $(\mu(E_1) - \mu(E_2))u(x) > (\mu(E_1) - \mu(E_2))u(y)$, and since $u(x) > u(y)$ we thus have $\mu(E_1) > \mu(E_2)$. Since $u(x') > u(y')$, it follows that

$$\begin{aligned} U(x'_{E_1}y') &= \mu(E_1)u(x') + (1 - \mu(E_1))u(y') \\ &> \mu(E_2)u(x') + (1 - \mu(E_2))u(y') = U(x'_{E_2}y'), \end{aligned}$$

so $x'_{E_1}y' \succ x'_{E_2}y'$. Hence (**P4**) holds.

(v) (**P5**) holds because we require μ to be unique in Definition 25. If $f \sim g$ for all acts $f, g \in F$, then μ is arbitrary.

This was the easy part. Now we turn to sufficiency. It helps to introduce a *qualitative probability* relation $\succ$ on the set of events Σ. If we write $E_1 \succ E_2$, we mean that E_1 is strictly more likely than E_2. We define this as $E_1 \succ E_2$ iff $x_{E_1}y \succ x_{E_2}y$ for any pair of constant acts $x, y \in F$ with $x \succ y$. Clearly this is a weak order under (**P1**), inheriting this from the preference relation on acts.

The equally-likely relation $\sim$ and the weak more-likely-than relation $\succsim$ is induced from $\succ$ in the same way that the indifference and weak preference relations are induced from strict preference relations.

Lemma 12. *Suppose* (**P1**)-(**P5**) *hold. Then:*

(i) *Given events $A, B, C \in \Sigma$ with $A \cap C = \varnothing$ and $B \cap C = \varnothing$, we have $A \succ B$ iff $A \cup C \succ B \cup C$, and $A \sim B$ iff $A \cup C \sim B \cup C$.*

(ii) *Let $A_1, A_2, B_1, B_2 \in \Sigma$. If $A_1 \succ B_1$, $A_2 \succ B_2$ and $A_1 \cap A_2 = B_1 \cap B_2 = \varnothing$, then $A_1 \cup A_2 \succ B_1 \cup B_2$.*

(iii) *Let $A_1, A_2, B_1, B_2 \in \Sigma$. If $A_1 \sim B_1$, $A_2 \sim B_2$ and $A_1 \cap A_2 = B_1 \cap B_2 = \varnothing$, then $A_1 \cup A_2 \sim B_1 \cup B_2$. Thus if $\{A_1, \ldots, A_n\}$ and $\{B_1, \ldots, B_n\}$ are disjoint collections of events in Σ with $A_i \sim B_i$ for all i, then $\bigcup_{i=1}^n A_i \sim \bigcup_{i=1}^n B_i$.*

(iv) *Let $A, B \in \Sigma$. If $A \succ B$ then $B^c \succ A^c$, and if $A \sim B$ then $A^c \sim B^c$.*

Proof. Abdellaoui & Wakker (2020) have visual proofs for some of these.

(i) For any constant acts $x \succ y$, (**P2**) (see footnote 31) implies $x_{A\cup C}y = x_C(x_Ay) \succ x_C(x_By) = x_{B\cup C}$ iff $x_Ay = y_C(x_Ay) \succ y_C(x_By) = x_By$. Hence $A \succ B$ iff $A \cup C \succ B \cup C$. Since $A \sim B$ iff $A \not\succ B$ and $B \not\succ A$, we have $A \sim B$ iff $A \cup C \not\succ B \cup C$ and $B \cup C \not\succ A \cup C$, so $A \sim B$ iff $A \cup C \sim B \cup C$.

(ii) Define $A_0 = (A_1 \cup A_2)^c$ and $B_0 = (B_1 \cup B_2)^c$. Now we can represent the state space S as the union of the 9 disjoint events $C_{ij} := A_i \cap B_j$ in the following matrix:

$$\begin{array}{cccc} & B_0 & B_1 & B_2 \\ A_0 & A_0 \cap B_0 & A_0 \cap B_1 & A_0 \cap B_2 \\ A_1 & A_1 \cap B_0 & A_1 \cap B_1 & A_1 \cap B_2 \\ A_2 & A_2 \cap B_0 & A_2 \cap B_1 & A_2 \cap B_2 \end{array} \quad = \quad \begin{array}{cccc} & B_0 & B_1 & B_2 \\ A_0 & C_{00} & C_{01} & C_{02} \\ A_1 & C_{10} & C_{11} & C_{12} \\ A_2 & C_{20} & C_{21} & C_{22} \end{array}$$

Moreover, each A_i is the union of the events in row i, i.e. $A_i = C_{i0} \cup C_{i1} \cup C_{i2}$, and each B_j is the union of the events in column j, i.e. $B_j = C_{0j} \cup C_{1j} \cup C_{2j}$.

Since $C_{11} \subset A_1, B_1$, we have that (i) implies

$$\begin{aligned} A_1 \succ B_1 \quad &\text{iff} \quad C_{10} \cup C_{12} \succ C_{21} \cup C_{01} \\ &\text{iff} \quad C_{20} \cup C_{10} \cup C_{12} \succ C_{20} \cup C_{21} \cup C_{01} \end{aligned}$$

Similarly,

$$\begin{aligned} A_2 \succ B_2 \quad &\text{iff} \quad C_{20} \cup C_{21} \succ C_{02} \cup C_{12} \\ &\text{iff} \quad C_{20} \cup C_{21} \cup C_{01} \succ C_{02} \cup C_{12} \cup C_{01}. \end{aligned}$$

Suppose $A_1 \succ B_1$ and $A_2 \succ B_2$. Since $\succ$ is a weak order, it is transitive, and thus

$$\begin{aligned} C_{20} \cup C_{10} \cup C_{12} &\succ C_{20} \cup C_{21} \cup C_{01} \\ &\succ C_{02} \cup C_{12} \cup C_{01}. \end{aligned}$$

By (i), adding $C_{11} \cup C_{21} \cup C_{22}$ to both sides does not change the order of likelihood, so

$$\begin{aligned} A_1 \cup A_2 &= C_{20} \cup C_{10} \cup C_{12} \cup C_{11} \cup C_{21} \cup C_{22} \\ &\succ C_{02} \cup C_{12} \cup C_{01} \cup C_{11} \cup C_{21} \cup C_{22} = B_1 \cup B_2. \end{aligned}$$

(iii) We can replace each instance of $\succ$ in the proof of (ii) with $\sim$ without altering the validity of the argument, giving $A_1 \cup A_2 \sim B_1 \cup B_2$. The second claim follows by iteration.

(iv) As with the proof of (ii), we can represent S as the union of the 4 disjoint events in the following matrix:

$$\begin{array}{ccc} & B & B^c \\ A & A \cap B & A \cap B^c \\ A^c & A^c \cap B & A^c \cap B^c \end{array}$$

By two applications of (i), we have

$$\begin{aligned} A \succ B \quad &\text{iff} \quad (A \cap B) \cup (A \cap B^c) \succ (A \cap B) \cup (A^c \cap B) \\ &\text{iff} \quad A \cap B^c \succ A^c \cap B \\ &\text{iff} \quad (A \cap B^c) \cup (A^c \cap B^c) \succ (A^c \cap B) \cup (A^c \cap B^c) \end{aligned}$$

Since $B^c = (A \cap B^c) \cup (A^c \cap B^c)$ and $A^c = (A^c \cap B) \cup (A^c \cap B^c)$, we thus have $A \succ B$ iff $B^c \succ A^c$. Replacing each instance of $\succ$ in the proof with $\sim$ gives the result for equal likelihood.

□

The next step is to use some "small" nonnull event $E \in \Sigma$ to calibrate the subjective probabilities. We will partition the state space S into equally-likely events $E_1, \dots, E_n$ that are as likely as E, along with some residual event $R \prec E$.

Call a partition $\mathcal{P} = \{E_1, \dots, E_n, R\}$ of S a *close to uniform partition* if $E_1 \sim E_i$ for all $i = 1, \dots, n$ and $E_1 \succ R$. Note the residual event can be the empty set.

Lemma 13. *Suppose* (**P1**)-(**P5**) *and* (**P6.2**) *hold. For each nonnull event $E \in \Sigma$, there is a close to uniform partition $\mathcal{P} = \{E_1, \ldots, E_n, R\}$ of S such that $E \sim E_i$ for all i and $E \succ R$.*

Proof. If $E \succ E^c$ then taking $E_1 = E$ and $R = E^c$ gives us a requisite close to uniform partition $\{E_1, R\}$. Hence suppose $E^c \succsim E$. Event solvability (**P6.2**) implies that for all events $A \subseteq E^c$ with $A \succsim E$, there is an event $E_1 \subseteq A$ with $E_1 \sim E$. Suppose we have defined $E_1, \ldots, E_k$. If $E \succ \left(\bigcup_{i=1}^k E_i\right)^c$, then setting $R = \bigcup_{i=1}^k E_i$ gives us a requisite close to uniform partition $\{E_1, \ldots, E_k, R\}$. If $\left(\bigcup_{i=1}^k E_i\right)^c \succsim E$ then applying event solvability (**P6.2**) again, we can find an $E_{k+1} \subseteq \left(\bigcup_{i=1}^k E_i\right)^c$ with $E_{k+1} \sim E$. By archimedeanity (**P6.1**), there cannot be an infinite sequence of equally likely events, so in finitely many, say n, steps we arrive at $R = \left(\bigcup_{i=1}^n E_i\right)^c \prec E$. This gives us a requisite close to uniform partition $\{E_1, \ldots, E_n, R\}$. $\square$

Lemma 14. *Suppose* (**P1**)-(**P5**) *and* (**P6.2**) *hold. If $\mathcal{P} = \{E_1, \ldots, E_n, R\}$ and $\mathcal{P}' = \{E_1', \ldots, E_m', R'\}$ are close to uniform partitions with $E_i \sim E$ for all $1 \leq i \leq n$ and $E_j' \sim E$ for all $1 \leq j \leq m$, then $n = m$ and $R' \sim R$.*

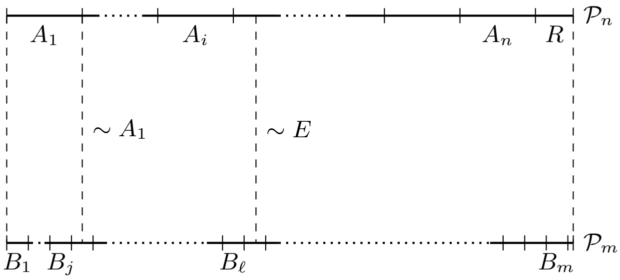
Proof. Suppose $m > n$. We have $E_i \sim E_j'$ for all i, j. By Lemma 12(iii), $\bigcup_{i=1}^n E_i \sim \bigcup_{i=1}^n E_i'$, and so by Lemma 12(iv), we have $R \sim R' \cup \bigcup_{i=n+1}^m E_i$. But $R' \cup \bigcup_{i=n+1}^m E_i \succ E$ and $E \succ R$, yielding a contradiction. Taking $n > m$ gives a similar contradiction. Hence $n = m$. Since $R = \left(\bigcup_{i=1}^n E_i\right)^c$ and $R' = \left(\bigcup_{j=1}^m E_j'\right)^c$, Lemma 12(iv) implies $R \sim R'$. $\square$

At this point, we assume Σ contains no atoms, i.e. we assume (**P9**) holds. The case with atoms will be considered after.

Lemma 15. *Suppose* (**P1**)-(**P5**), (**P6.2**) *and* (**P9**) *hold. Then for every $n \in \mathbb{N}$, there is a close to uniform partition of S with at least n elements.*

Proof. Fix a nonnull event E. By Lemma 13, we can define a close to uniform partition $\mathcal{P} = \{E_1, \ldots, E_m, R\}$ such that $E_i \sim E$ for all i. By (**P9**), there are no atoms, and so there are non-null events $E \subset E'$. If $E' - E \succsim E'$, put $A = E'$, and otherwise put $A = E' - E$. By event solvability (**P6.2**), for each E_i there is an $A_i \subseteq E_i$ with $A_i \sim A$. Now Lemma 12(ii) implies $E_i - A_i \sim E - A \succsim A \sim A_i$. Applying solvability again, there is a $B_i \subseteq E_i - A_i$ with $B_i \sim A$. Thus for each i, we can find a pair of disjoint events A_i, B_i contained in E_i such that $A_i \sim B_i \sim A$. This gives us $2m$ events. It follows that we can construct a close to uniform partition $\{E_1, \ldots, E_n, R'\}$ with $E_i \sim A$ that has $n \geq 2m$. Following this procedure repeatedly, we can make n arbitrarily large. $\square$

Now we can construct the probability measure/charge μ implied by $\succ$. For each close to uniform partition $\mathcal{P}$ and each event E, we measure how many of the equally-likely events in $\mathcal{P}$ we can fit into a union that is less likely than E, as a fraction of the total number of equally-likely events in $\mathcal{P}$. As the number of equally-likely events in these partitions grows large, this fraction stabilizes to a well defined limit, which gives us $\mu(E)$.



A representation of a pair of partitions as in the proof of Lemma 16. Here, the state space S is represented as a line and $i = k(E, \mathcal{P}_n)$, $j = k(A_1, \mathcal{P}_m)$, and $\ell = k(E, \mathcal{P}_m)$.

Lemma 16. *Suppose* (**P1**)-(**P5**), (**P6.2**) *and* (**P9**) *hold. Consider any event $E \in \Sigma$. For each close to uniform partition $\mathcal{P}_n = \{A_1, \dots, A_n, R\}$ of S, define*

$$k(E, \mathcal{P}_n) = \max\left\{ j \mid E \succsim \bigcup_{i=1}^{j} A_i \right\},$$

and put

$$\underline{\mu}(E, \mathcal{P}_n) := \frac{k(E, \mathcal{P}_n) - 1}{n+1} \quad \text{and} \quad \bar{\mu}(E, \mathcal{P}_n) := \frac{k(E, \mathcal{P}_n) + 2}{n}.$$

Then for any sequence of size n close to uniform partitions $\{\mathcal{P}_n\}_{n=1}^{\infty}$, the limit

$$\mu(E) := \lim_{n\to\infty} \underline{\mu}(E, \mathcal{P}_n) = \lim_{n\to\infty} \bar{\mu}(E, \mathcal{P}_n)$$

exists and its value does not depend on the choice of sequence of partitions.

Proof. If E is a null event, then $k(E, \mathcal{P}_n) = 0$ for all $\mathcal{P}_n$s so $\underline{\mu}(E, \mathcal{P}_n) \geq -\frac{1}{n+1}$ and $\bar{\mu}(E, \mathcal{P}_n) = \frac{2}{n}$. Thus we have $\lim_{n\to\infty} \underline{\mu}(E, \mathcal{P}_n) = \lim_{n\to\infty} \bar{\mu}(E, \mathcal{P}_n) = 0$.

Now suppose E is a nonnull event. Take any $\mathcal{P}_n = \{A_1, \dots, A_n, R\}$. We claim that for sufficiently large $m > n$, every partition $\mathcal{P}_m = \{B_1, \dots, B_m, R'\}$ has

$$\underline{\mu}(E, \mathcal{P}_n) < \underline{\mu}(E, \mathcal{P}_m) < \bar{\mu}(E, \mathcal{P}_m) < \bar{\mu}(E, \mathcal{P}_n).$$

By Lemma 12(ii)-(iii), we have $k(E, \mathcal{P}_m) \geq k(E, \mathcal{P}_n) k(A_1, \mathcal{P}_m)$. If $m \geq (k(A_1, \mathcal{P}_m) + 1)(n+1)$, then we would have at least $n+1$ disjoint events $A_i \sim A_1$ in the partition $\mathcal{P}_n$, yielding a contradiction. Thus $m < (k(A_1, \mathcal{P}_m) + 1)(n+1)$. It follows that

$$\underline{\mu}(E, \mathcal{P}_m) = \frac{k(E, \mathcal{P}_m) - 1}{m+1} \geq \frac{k(E, \mathcal{P}_n) k(A_1, \mathcal{P}_m) - 1}{(k(A_1, \mathcal{P}_m) + 1)(n+1) + 1} = \frac{k(E, \mathcal{P}_n) - \frac{1}{k(A_1, \mathcal{P}_m)}}{n + 1 + \frac{n+2}{k(A_1, \mathcal{P}_m)}}.$$

As $k(A_1, \mathcal{P}_m) \to \infty$, the rhs converges to $\frac{k(E,\mathcal{P}_n)}{n+1} > \frac{k(E,\mathcal{P}_n)-1}{n+1} = \underline{\mu}(E, \mathcal{P}_n)$.

Similarly, Lemma 12(ii)-(iii) imply that $m = k(S, \mathcal{P}_n) \geq k(S, \mathcal{P}_n) k(A_1, \mathcal{P}_m) = n k(A_1, \mathcal{P}_m)$. If $k(E, \mathcal{P}_m) \geq (k(A_1, \mathcal{P}_m) + 1)(k(E, \mathcal{P}_n) + 1)$, then $\bigcup_{i=1}^{k(E,\mathcal{P}_n)+1} A_i \precsim E$, contradicting the definition of k. Thus $k(E, \mathcal{P}_m) < (k(A_1, \mathcal{P}_m) + 1)(k(E, \mathcal{P}_n) + 1)$. It follows that

$$\begin{aligned}\bar{\mu}(E, \mathcal{P}_m) = \frac{k(E, \mathcal{P}_m) + 2}{m} &\leq \frac{(k(A_1, \mathcal{P}_m) + 1)(k(E, \mathcal{P}_n) + 1) + 2}{nk(A_1, \mathcal{P}_m)} \\ &= \frac{k(E, \mathcal{P}_n) + 1 + \frac{k(E, \mathcal{P}_n)+3}{k(A_1, \mathcal{P}_m)}}{n}.\end{aligned}$$

As $k(A_1, \mathcal{P}_m) \to \infty$, the rhs converges to $\frac{k(E,\mathcal{P}_n)+1}{n} < \frac{k(E,\mathcal{P}_n)+2}{n} = \bar{\mu}(E, \mathcal{P}_n)$. Hence we get

$$\underline{\mu}(E, \mathcal{P}_n) < \underline{\mu}(E, \mathcal{P}_m) < \bar{\mu}(E, \mathcal{P}_m) < \bar{\mu}(E, \mathcal{P}_n)$$

for all sufficiently large $m > n$. We therefore conclude that for any sequence $\{\mathcal{P}_n\}$ of such close to uniform partitions, we have

$$\underline{\mu}(E, \mathcal{P}_n) < \liminf_{m\to\infty} \underline{\mu}(E, \mathcal{P}_m) \leq \limsup_{m\to\infty} \underline{\mu}(E, \mathcal{P}_m) < \bar{\mu}(E, \mathcal{P}_n)$$

for all n. Since $\lim_{n\to\infty} \Big(\bar{\mu}(E, \mathcal{P}_n) - \underline{\mu}(E, \mathcal{P}_n)\Big) = 0$, it follows that there is a well-defined limit $\mu(E) = \lim_{n\to\infty} \underline{\mu}(E, \mathcal{P}_n) = \lim_{n\to\infty} \bar{\mu}(E, \mathcal{P}_n)$. This limit is invariant to the choice of partition sequence. For suppose that $\mu(E)$ is constructed using sequences $\{\mathcal{P}_n\}$ and that $\mu'(E)$ is constructed using sequence $\{\mathcal{P}'_n\}$. Let $\epsilon_n = \bar{\mu}(E, \mathcal{P}_n) - \underline{\mu}(E, \mathcal{P}_n)$. We have showed that for each n and all sufficiently large $m > n$, $\bar{\mu}(E, \mathcal{P}_m)$ and $\underline{\mu}(E, \mathcal{P}_m)$ must lie in an ϵ_n-ball N_{ϵ_n} about $\mu(E)$, and thus $\mu'(E)$ must lie in N_{ϵ_n}. Since $\epsilon_n \to 0$, it follows that $\mu(E) = \mu'(E)$. □

Lemma 16 demonstrates that μ is uniquely pinned down by $\succ$. Now we need to verify μ is a probability charge for statement (i) and a probability measure for statement (ii) of the theorem. For any event $E \in \Sigma$ and any close to uniform partition $\mathcal{P}$, both $\underline{\mu}(E, \mathcal{P}) \in [0, 1]$ and $\bar{\mu}(E, \mathcal{P}) \in [0, 1]$. Hence $\mu(E) \in [0, 1]$. Clearly, $\mu(\varnothing) = 0$ and $\mu(S) = 1$. All that remains to deal with is additivity:

Lemma 17. *Suppose* (**P1**)-(**P5**), (**P6.2**) *and* (**P9**) *hold. Then μ is finitely additive. Moreover, if* (**P8**) *also holds, then μ is countably additive.*

Proof. We handle finite additivity (charges) first. Suppose $E_1, E_2 \in \Sigma$ are disjoint events. For any partition $\mathcal{P}_n = \{A_1, \dots, A_n, R\}$, we have from the proof of Lemma 16 that

$$\frac{k(E_1, \mathcal{P}_n) - 1}{n + 1} \leq \mu(E_1) \leq \frac{k(E_1, \mathcal{P}_n) + 2}{n}$$

and

$$\frac{k(E_2, \mathcal{P}_n) - 1}{n + 1} \leq \mu(E_2) \leq \frac{k(E_2, \mathcal{P}_n) + 2}{n}.$$

Since E_1, E_2 are disjoint,[35]

$$k(E_1, \mathcal{P}_n) + k(E_2, \mathcal{P}_n) \leq k(E_1 \cup E_2, \mathcal{P}_n) \leq k(E_1, \mathcal{P}_n) + k(E_2, \mathcal{P}_n) + 1.$$

Hence

$$\frac{k(E_1, \mathcal{P}_n) + k(E_2, \mathcal{P}_n) - 2}{n+1} \leq \mu(E_1 \cup E_2) \leq \frac{k(E_1, \mathcal{P}_n) + k(E_2, \mathcal{P}_n) + 3}{n}.$$

Thus $\mu(E_1 \cup E_2)$ is bounded by the sum of the bounds for $\mu(E_1)$ and $\mu(E_2)$, and taking limits as in the proof of Lemma 16 gives $\mu(E_1 \cup E_2) = \mu(E_1) + \mu(E_2)$.

Now we turn to countable additivity (measures). Suppose set continuity (**P8**) also holds. Fix a sequence of disjoint events $\{E_i\}_{i=1}^{\infty}$ and put $E = \bigcup_{i=1}^{\infty} E_i$. Defining $B_n = \bigcup_{j=n}^{\infty} E_j$, we have $B_n \downarrow \varnothing$. Fix any nonnull event $A \in \Sigma$ and constant acts $x \succ y$. Since A is nonnull, $x_A y \succ y$. By set continuity (**P8**), there is an integer N such that $x_A y \succ x_{B_n} y$ for all $n \geq N$. Hence for any nonnull event $A \in \Sigma$, there is an N such that $A \succ B_n$ for all $n \geq N$. Since $E_n \subseteq B_n$, it follows that there is an N such that $A \succ E_n$ for all $n \geq N$.

By finite additivity, $\mu(E) = \mu\left(B_N \cup \bigcup_{n=1}^{N-1} E_n\right) = \mu(B_N) + \sum_{n=1}^{N-1} \mu(E_n)$ for all finite N. For any close to uniform partition $\mathcal{P}_k = \{A_1, \ldots, A_k, R\}$, we have $B_N \succ A_1$ for sufficiently large N, and thus $\mu(B_N) \downarrow 0$. Hence

$$\mu(E) = \lim_{N\to\infty} \left(\mu(B_N) + \sum_{n=1}^{N-1} \mu(E_n)\right) = \lim_{N\to\infty} \sum_{n=1}^{N-1} \mu(E_n) = \sum_{n=1}^{\infty} \mu(E_n).$$

It follows that μ is countably additive if (**P8**) holds. □

Thus μ is a probability measure or charge. We call μ a *subjective probability representation* of $\succ$ on Σ if $E_1 \succ E_2$ iff $\mu(E_1) > \mu(E_2)$.

Lemma 18. *Suppose* (**P1**)-(**P5**)*,* (**P6.2**) *and* (**P9**) *hold. Then μ is a unique subjective probability representation of $\succ$ on the set of events Σ.*

Proof. Let $E_1, E_2 \in \Sigma$ be events with $E_1 \succsim E_2$. Then for any close to uniform partition $\mathcal{P}$, we have $\underline{\mu}(E_1, \mathcal{P}) \geq \underline{\mu}(E_2, \mathcal{P})$ (and $\bar{\mu}(E_1, \mathcal{P}) \geq \bar{\mu}(E_2, \mathcal{P})$) so $\mu(E_1) \geq \mu(E_2)$. Suppose $E_1 \succ E_2$. Then by event solvability (**P6.2**), there is an event $E_2' \subset E_1$ with $E_2' \sim E_1$, and thus $\mu(E_2') = \mu(E_2)$. The relative complement $E_1 - E_2'$ is nonnull. By finite additivity, $\mu(E_1) = \mu(E_2') + \mu(E_1 - E_2') = \mu(E_2) + \mu(E_1 - E_2')$, so $\mu(E_1) > \mu(E_2)$ iff $\mu(E_1 - E_2') > 0$.

Take any nonnull event A. By archimedeanity (**P6.1**), the close to uniform partition with equally-likely elements $A_i \sim A$ has finitely many elements, so let $\mathcal{P} = \{A_1, \ldots, A_n, R\}$ be this close to uniform partition with $A_i \sim A$. Now $\mu(A_1) = \cdots = \mu(A_n) \geq \mu(R)$ and by finite additivity, $\mu(R) + \sum_{i=1}^{n} \mu(A_i) = \mu(S) = 1$. Hence $\mu(A_i) \geq$

[35] The remainder R_1 in E_1 after removing $k(E_1, \mathcal{P}_n) \sim A_1$ events from E_1 must have $A_1 \succ R_1$, and the same goes for the remainder R_2 of E_2. Removing the same events from $E_1 \cup E_2$ gives $R_1 \cup R_2$, which can contain at most one event $\sim A_1$.

$\frac{1}{n+1} > 0$, so $\mu(A) > 0$. Thus every nonnull event has strictly positive measure, so it follows that $\mu(E_1 - E_2') > 0$, establishing $\mu(E_1) > \mu(E_2)$ whenever $E_1 \succ E_2$. Hence μ represents $\succ$. $\square$

Having derived the probability measure/charge μ from preferences, we now need to derive the subjective expected utility representation $U(f) = \int_S u(f(s))\mu(\mathrm{d}s)$.

To do this, we make use of the notion of *standard gambles*:

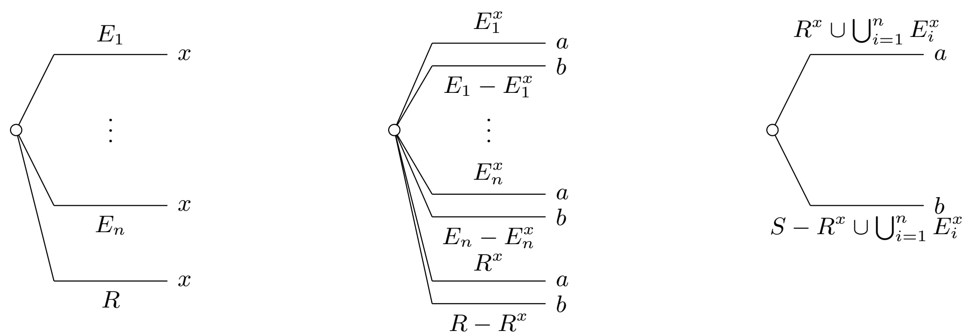
Definition 28 (Standard gambles)**.**

(a) *Standard gamble.* Given constant acts $a, b \in F$ with $a \succ b$ and a constant act $x \in F$ with $a \succsim x \succsim b$, let S^x be an event such that $a_{S^x}b \sim x$. Then we call the act $a_{S^x}b$ a *standard gamble* of x.

(b) *Conditional standard gamble.* Given constant acts $a \succ b$, a constant act x with $a \succsim x \succsim b$, a nonnull event $E \in \Sigma$, and an act $f \in F$, let $E^x \subseteq E$ be an event such that $x_E f \sim (a_{E^x}b)_E f$, or equivalently $x \sim_E a_{E^x}b$. Then we call the act $(a_{E^x}b)_E f$ a *conditional standard gamble* of x, conditional on E.

Fix constant acts $a \succ b$. By event solvability (**P6.2**), a standard gamble of x exists for any $a \succsim x \succsim b$, i.e. we can find an event S^x with $a_{S^x}b \sim x$. Whenever we write S^x or E^x, we are denoting some event that corresponds to a standard gamble of x.

We will use standard gambles to generate a Bernoulli utility for each such consequence x.[36] For each $x \in X$ with $a \succsim x \succsim b$, put $u(x) = \mu(S^x)$. Clearly, $\mu(S^x) = \mu(\tilde{S}^x)$ for any standard gambles $a_{S^x}b$ and $a_{\tilde{S}^x}b$, so the choice of S^x is immaterial.



The gambles in the proof of Lemma 19.

Lemma 19. *Suppose* (**P1**)-(**P5**), (**P6.2**) *and* (**P9**) *hold. Fix constant acts $a \succ b$ and x with $a \succsim x \succsim b$. Given an event E and an event E^x corresponding to a conditional standard gamble of x, we have*

$$u(x) = \mu(S^x) = \frac{\mu(E^x)}{\mu(E)}.$$

[36] Recall we are using x to denote both a consequence in X and the corresponding constant act in F.

Proof. First, if $x \sim b$ then $\mu(E^x) = 0$, so $u(x) = \frac{\mu(E^x)}{\mu(E)}$ is immediate. Suppose then that $x \succ b$. Note that $A \succ B$ implies $A^x \succ B^x$, i.e. A^x is monotone in A with respect to the qualitative probability relation.

First, take a close to uniform partition $\mathcal{P} = \{E_1, \ldots, E_n, R\}$ with $E_i \sim E$. Now,

$$\begin{aligned} x &= x_{E_1} x_{E_2} \cdots x_{E_n} x \\ &\sim a_{E_1^x} b_{E_1 - E_1^x} a_{E_2^x} b_{E_2 - E_2^x} \cdots a_{E_n^x} b_{E_n - E_n^x} a_{R^x} b \\ &= a_{R^x \cup E_1^x \cup \cdots \cup E_n^x} b. \end{aligned}$$

These are depicted immediately above the lemma. We have $\mu(R^x) + \sum_{i=1}^n \mu(E_i^x) = \mu(E^x) = u(x)$ by definition. Since each E_i is equally likely, $\frac{\mu(E_i^x)}{\mu(E_i)} = \frac{\mu(E^x)}{\mu(E)}$ for all i, we have

$$u(x) = \frac{\mu(R^x) + \sum_{i=1}^n \mu(E_i^x)}{\mu(S)} = \frac{\mu(R^x) + \sum_{i=1}^n \mu(E_i^x)}{\mu(R) + \sum_{i=1}^n \mu(E_i)} = \frac{\mu(R^x) + n\mu(E^x)}{\mu(R) + n\mu(E)}.$$

Since $0 \leq \mu(R^x) \leq \mu(R) < \mu(E) \leq \frac{1}{n}$, as we consider events $E \sim E_i$ for a sequence of increasingly large partitions $\{\mathcal{P}_n\}$, taking the limit as $n \to \infty$ gives $\frac{\mu(E^x)}{\mu(E)} \to u(x)$. Moreover, for each j, we have

$$\frac{\mu\left(\left(\bigcup_{i=1}^j E_i\right)^x\right)}{\mu\left(\bigcup_{i=1}^j E_i\right)} = \frac{\mu\left(\bigcup_{i=1}^j E_i^x\right)}{\mu\left(\bigcup_{i=1}^j E_i\right)} = \frac{j\mu(E^x)}{j\mu(E)} = \frac{\mu(E^x)}{\mu(E)},$$

and so $\frac{\mu(E^x)}{\mu(E)} = u(x)$ holds in the limit for unions of the elements of close to uniform partitions $\mathcal{P}_n$.

Now suppose E is an arbitrary event. Define a sequence of close to uniform partitions $\mathcal{P}_n = \{A_1, \ldots, A_n, R\}$ and for each, choose $j(\mathcal{P}_n)$ so that $\bigcup_{i=1}^{j(\mathcal{P}_n)} A_i \succsim E \succsim \bigcup_{i=1}^{j(\mathcal{P}_n)-1} A_i$. Then $\mu\left(\bigcup_{i=1}^j A_i\right) \geq \mu(A) \geq \mu\left(\bigcup_{i=1}^{j-1} A_i\right)$ and by monotonicity, $\mu\left(\left(\bigcup_{i=1}^j A_i\right)^x\right) \geq \mu(A^x) \geq \mu\left(\left(\bigcup_{i=1}^{j-1} A_i\right)^x\right)$. These upper and lower bounds converge to $\mu(A)$ and $\mu(A^x)$ as $n \to \infty$, and so the limits of $\frac{\mu\left(\left(\bigcup_{i=1}^{j-1} A_i\right)^x\right)}{\left(\bigcup_{i=1}^{j-1} A_i\right)}$ and $\frac{\mu\left(\left(\bigcup_{i=1}^{j} A_i\right)^x\right)}{\mu\left(\bigcup_{i=1}^{j} A_i\right)}$ are $\frac{\mu(E^x)}{\mu(E)}$, giving the result. $\square$

Now, by the sure-thing principle (**P2**), $x_E f \sim (a_{E^x} b)_E f$ iff $x_E f' \sim (a_{E^x} b)_E f'$ for all acts f, f', so the choice of act f does not affect preference. Since we are working with simple acts, any act f takes only finitely many values in X, say $\{x_1, \ldots, x_n\}$. Define $E_i = \{s \in S \mid f(s) = x_i\}$. We have $f \sim a_{E_1^x} b_{E_1 - E_1^x} f$ by definition of conditional standard gambles. Repeatedly making such substitutions, we get

$$\begin{aligned} f &\sim a_{E_1^x} b_{E_1 - E_1^x} a_{E_2^x} b_{E_2 - E_2^x} \cdots a_{E_n^x} b \\ &= a_{E_1^x \cup \cdots \cup E_n^x} b, \end{aligned}$$

and

$$\mu\left(\bigcup_{i=1}^{n} E_i^x\right) = \sum_{i=1}^{n} \mu(E_i^x) = \sum_{i=1}^{n} u(x_i)\mu(E_i) =: U(f),$$

which is precisely the subjective expected utility of the act. Thus any simple act f is indifferent to a standard gamble $a_{U(f)}b$. Since $a \succ b$, $f \sim a_{U(f)}b \succ a_{U(g)}b \sim g$ iff $U(f) > U(g)$. Thus U represents $\succ$ for all simple acts f with ranges in $\{x \mid a \succsim x \succsim b\}$. Note that having fixed $u(a)$ and $u(b) < u(a)$, u is uniquely determined by $u(x) = \mu(S^x)u(a) + (1 - \mu(S^x))u(b)$. For any other choice of $v(a)$ and $v(b) < v(a)$, defining $\alpha = \frac{v(b)-v(a)}{u(b)-u(a)} > 0$ and $\beta = v(a) - \alpha u(a)$ gives $v = \alpha u + \beta$, so v is a positive affine transformation of u.

Finally we extend u to all of F. To do this, consider a sequence of pairs of consequences $\{(a_n, b_n)\}$ with $a_{n+1} \succ a_n \succ b_n \succ b_{n+1}$ inducing a sequence of sets of acts $X_n := \{x \in X \mid a_n \succsim x \succsim b_n\}$ such that $X_n \uparrow X$, and let $F_n := \{f \in F \mid \mathrm{range}(f) \subseteq X_n\}$ for each n. Repeating the previous steps, we obtain a Bernoulli utility function $u_1 : F_1 \to [0, 1]$ representing $\succ$ on F_1. Given Bernoulli utility function u_{n-1}, repeating the previous steps we can obtain a Bernoulli utility function $u_n' : F_n \to [0, 1]$ representing $\succ$ on F_n. Define $\alpha_n = \frac{1}{u_n'(a_{n-1})-u_n'(b_{n-1})}$ and $\beta_n = u_{n-1}(a_{n-1}) - \alpha_n u_n'(a_{n-1})$. Define $u_n = \alpha_n u_n' + \beta_n$. Then $u_n = u_{n-1}$ on F_{n-1}, and u_n is a Bernoulli utility function representing $\succ$ on F_n. Letting $u = \lim u_n$, we have that $U(f) = \int_S u(f)\,\mathrm{d}\mu$ is a subjective expected utility representation of $\succ$ on F.

This has proved the theorem under (**P9**). We now relax the assumption that there is no atoms. Suppose there is an atom $E \in \Sigma$. If there is a nonnull event E' with $E \succ E'$, then event solvability (**P6.2**) would imply there is an event $E'' \subset E$ with $E'' \sim E'$ and thus $E \succ E'' \succ \varnothing$, contradicting that E is an atom. Hence there can be no nonnull events less likely than E. For any event $A \succ E$, we can by (**P6.1**)-(**P6.2**) write $A = \bigcup_{i=1}^{k} A_i$ for finitely many disjoint atoms $A_i \sim E$. Since $S \succ E$, we have $S = \bigcup_{i=1}^{n} E_i$ for finitely many disjoint atoms $E_i \sim E$. This gives us a close to uniform partition $\mathcal{P} = \{E_1, \ldots, E_n, R\}$ of S with $E_i \sim E$ and $R \sim \varnothing$. Now (**P6.2**) implies that for any event A and each E_i, either $E_i - A \sim \varnothing$ or $E_i \cap A \sim \varnothing$; that is, up to a null event, either E_i is contained in A or disjoint from A. Thus we can write A as a union of $j(A)$ disjoint atoms and a null event. Put $\mu(A) = \frac{j(A)}{n}$ for all events A (so $\mu(E) = \frac{1}{n}$ for all atoms E).

Now, suppose there are constant acts $x, y, z \in F$ with $x \succ y \succ z$. For any atom E, we then have $x_E z \succ y_E z \succ z_E z = z$. But then (**P6.2**) implies there is an event $B \subseteq E$ with $y_E z \sim x_B z_{e-B} z = x_B z$, which contradicts the fact that A is an atom. Hence there cannot be three constant acts $x \succ y \succ z$. By nontriviality (**P5**), we cannot have $x \sim y$ for all constant acts x, y, and so we must have $x \succ y$ for precisely two constant acts, with all other constant acts z having $z \sim x$ or $z \sim y$. Put $u(a) = 1$ for all $a \in [x]$ and $u(b) = 0$ for all $b \in [y]$.

For any simple acts $f, g \in F$, we have $f \succ g$ iff $\{s \in S \mid f(s) \sim x\} \succ \{s \in S \mid g(s) \sim x\}$, which holds iff $\int_S u(f) = \mu(\{s \in S \mid f(s) \sim x\}) > \mu(\{s \in S \mid g(s) \sim x\}) = \int_S u(g)\,\mathrm{d}\mu$. Thus $\succ$ has a subjective expected utility representation. The affine transformation remark is straightforward to see here. $\square$

Theorem 14 implies Savage's version:

Corollary 4 (Savage, 1954)**.** *Let X be a space of consequences, let S be a state space equipped with a σ-algebra Σ, and let F_S be the space of all simple acts $f : S \to X$. Let $\succ$ be a strict preference relation on F. Moreover, assume event continuity* (**P6**) *holds. Then $\succ$ has a subjective expected utility representation iff* (**P1**)-(**P5**) *hold.*

Proof. See Proposition 4 in Abdellaoui & Wakker (2020). $\square$

2.3.3 Subjective expected utility for general acts

Savage's theory is of course not restricted to simple acts. Fortunately, extending Theorem 14 to general acts is less work than the previous proof:

Theorem 15. *Let X be a space of consequences, let S be a state space equipped with an algebra Σ, and let F be the space of all acts $f : S \to X$. Let $\succ$ be a strict preference relation on F. Moreover, assume event solvability* (**P6.2**) *holds. Then:*

(i) *$\succ$ has a subjective expected utility representation with a probability charge iff* (**P1**)-(**P5**), (**P6.1**) *and* (**P7**) *hold.*

(ii) *$\succ$ has a subjective expected utility representation with a probability measure iff* (**P1**)-(**P5**), (**P6.1**), (**P7**) *and* (**P8**) *hold.*

Moreover, if $U(f) = \int_S u(f)\,\mathrm{d}\mu_1$ and $V(f) = \int_S v(f)\,\mathrm{d}\mu_2$ are subjective expected utility representations of $\succ$, then $\mu_1 = \mu_2$ and v is a positive affine transformation of u.

2.4 Anscombe-Aumann

Savage's theory is beautiful but difficult to extend once we start worrying about phenomena that cannot be captured by expected utility, such as ambiguity aversion. Anscombe & Aumann (1963) propose a nice extension to Savage's setting that is significantly easier to work with. This setting involves calibrating the "difficult" subjective probabilities that are attached to events that can't easily be quantified via "easier" extraneous probabilities that we all can agree on, like the probability that a fair coin turns up heads.

We adopt the Fishburn (1970) simplification of Anscombe & Aumann's setting, which is predominant in the literature. The main difference in setting compared to Savage (1954) is the definition of an act:

Definition 29 (Anscombe-Aumann act)**.** Given a set of consequences X and a state space (S, Σ), an *Anscombe-Aumann act* is a (measurable) function $f : S \to \Delta_S(X)$, where $\Delta_S(X)$ is the set of simple probability measures on X.

Anscombe-Aumann acts are also known as *horse lotteries*, in which case the lotteries in $\Delta_S(X)$ are known as *roulette lotteries.* The motivation here is that roulette is one of those games of chance with well-defined chances over outcomes, as with von Neumann & Morgenstern's lotteries. The outcomes of a horse race, on the other hand, cannot be quantified and reasonable people can disagree about the odds that a given horse will win the race.

The terminology developed for Savage acts carries over to Anscombe-Aumann acts – for example, a constant act in this setting is now an Anscombe-Aumann act f with $f(s) = p$ for some $p \in \Delta(X)$, and we use p to denote both the constant act and the element of $\Delta(X)$.

Note that $\Delta_S(X)$, the set of simple probability measures on X, is a mixture set when endowed with the mixture operation $p \oplus_\alpha q = \alpha p + (1-\alpha)q$. We will use p, q etc. to denote a roulette lotteries in $\Delta_S(X)$ and μ to denote the probability distribution in $\Delta(S)$ corresponding to the decisionmaker's beliefs.

Often the set of states S is assumed to be finite, but we will not make that restriction here. Let F be the set of all Anscombe-Aumann acts on S. We endow F with the mixture operation $f \oplus_\alpha g = \alpha f + (1-\alpha)g$ for all $f, g \in F$ and all $\alpha \in [0,1]$.

Axioms. Let S be a set of states, let X be a set of consequences and let F be a mixture set of all Anscombe-Aumann acts, equipped with mixture operation $\oplus$.

(**E1**) *Independence.* For all $f, g \in F$, $f \succ g$ iff $f \oplus_\alpha h \succ g \oplus_\alpha h$ for all $\alpha \in (0,1)$ and all $h \in F$.

(**E2**) *Archimedean continuity.* If $f, g, h \in F$ with $f \succ g$ and $g \succ h$, then $f \oplus_\alpha h \succ g$ and $g \succ f \oplus_\beta h$ for some $\alpha, \beta \in (0,1)$.

(**E3**) *Monotonicity.* For all $f, g \in F$, $f(s) \succsim g(s)$ for all $s \in S$ implies $f \succsim g$.

(**E4**) *Nontriviality.* $f \succ g$ for some $f, g \in F$.

Independence (**E1**) and Archimedean continuity (**E2**) are restatements of (**V1**) and (**V2**), with the set of Anscombe-Aumann acts being the relevant mixture set. Monotonicity (**E3**) is similar to Savage's (**P3**), again ensuring state-independence – note in reading (**E3**) that $f(s)$ and $g(s)$ are constant acts. Nontriviality (**E4**) is the direct counterpart of Savage's (**P5**).

Theorem 16 (Anscombe & Aumann, 1963)**.** *Let X be a set of consequences, let S be a state space equipped with algebra Σ, let F be the set of all simple Anscombe-Aumann acts on S, and let $\succ$ be a weak order on F. Then $\succ$ satisfies axioms* (**E1**)-(**E4**) *iff there is a function $u : X \to \mathbb{R}$ and a unique probability charge $\mu : 2^S \to [0,1]$ such that*

$$f \succ g \quad \textit{iff} \quad \int_S \left(\sum_{x \in \operatorname{supp} f(s)} u(x) f(s)(x) \right) \mu(\mathrm{d}s) > \int_S \left(\sum_{x \in \operatorname{supp} g(s)} u(x) g(s)(x) \right) \mu(\mathrm{d}s)$$

for all $f, g \in F$. Moreover, if $u : X \to \mathbb{R}$ and $v : X \to \mathbb{R}$ are functions representing $\succ$ in the sense above, then $v(x)$ is a positive affine transformation of $u(x)$.

Remark. The rhs of the "iff" is a bit messy. More compactly, it can be written as $\mathbb{E}_\mu[\mathbb{E}_{f(s)}u(x)] > \mathbb{E}_\mu[\mathbb{E}_{g(s)}u(x)]$, where $\mathbb{E}$ is the expectation operator. The following proof follows a note by Ozaki (2014).

Proof. First, suppose weak order $\succ$ has a representation

$$U(f) = \int_S \left(\sum_{x \in \operatorname{supp} f(s)} u(x) f(s)(x) \right) \mu(\mathrm{d}s) =: \int_S \hat{U}(f(s)) \mu(\mathrm{d}s)$$

with μ a unique probability charge. We show the axioms hold.

(i) Restricting to constant acts p, note $U(p) = \hat{U}(p)$, and $\hat{U}$ is a von Neumann-Morgenstern expected utility representation of $\succ$ on $\Delta(X)$. Thus (**E1**) holds by Theorem 12.

(ii) By the same reasoning, (**E2**) holds by Theorem 12.

(iii) Suppose $f(s) \succsim g(s)$ for all $s \in S$. Since $\hat{U}$ represents $\succ$ on $\Delta(X)$, $f(s) \succsim g(s)$ implies $\hat{U}(f(s)) \geq \hat{U}(g(s))$ for all $s \in S$, and so by monotonicity of the Lebesgue integral, $U(f) \geq U(g)$. Thus $f \succsim g$. Thus (**E3**) holds.

(iv) If $U(f) = U(g)$ for all $f, g \in F$, then we must have $f = g$ on all non-null events under μ, and so any other charge μ' that is absolutely continuous with respect to μ yields $U(f) = \int_S \hat{U}(f(s))\mu'(\mathrm{d}s)$, implying μ is not unique, a contradiction. Thus $U(f) > U(g)$ for some $f, g \in F$, and so (**E4**) holds.

Next, suppose the axioms (**E1**)-(**E4**) hold. By the mixture set theorem (Theorem 11), (**E1**) and (**E2**) imply there is a mixture preserving function $U : F \to \mathbb{R}$ representing $\succ$ on F. For each $p \in \Delta(X)$, define $\hat{U}(p) = U(p)$, where p in the right hand side denotes the constant act corresponding to $p \in \Delta(X)$ on the left hand side. Then $\hat{U} : \Delta(X) \to \mathbb{R}$ is a mixture preserving function representing $\succ$ on $\Delta(X)$. By the proof of the von Neumann-Morgenstern representation theorem (Theorem 12), it follows there is a Bernoulli utility function $u : X \to \mathbb{R}$ such that $\hat{U}(p) = \sum_{x \in \operatorname{supp} p} u(x)p(x)$ for all $p \in \Delta(X)$.

Note that there is a pair of constant acts $p_1, p_2 \in F$ such that $p_1 \succ p_2$, since otherwise (**E3**) implies $f \sim g$ for all acts $f, g \in F$, contrary to (**E4**). Thus $\hat{U}(p_1) > \hat{U}(p_2)$ for some constant acts p_1, p_2. Since $\hat{U}$ is affine, it is without loss of generality to assume $\hat{U}(p_1) = 1$ and $\hat{U}(p_2) = -1$. For now, we fix these constant acts $p_1 \succ p_2$.

Denote the range of $\hat{U}$ by $\mathcal{U} = \hat{U}(\Delta_S(X))$. Since $\Delta_S(X)$ is convex and $\hat{U}$ is mixture-preserving, $\mathcal{U}$ is convex. Define $\bar{F} = \{f \in F \mid \hat{U}(p_1) \geq U(f) \geq \hat{U}(p_2)\}$. Define $\bar{\mathcal{U}} = \{a \in \mathcal{U} \mid a = U(f) \text{for some } p_1 \succ f \succ p_2,\ f \in F\}$. We define a *utility act* to be a (measurable) function $\xi : S \to \mathcal{U}$, and we denote the set of all simple utility acts by B, and the set of all simple utility acts into $\bar{\mathcal{U}}$ by $B(\bar{\mathcal{U}})$.

Define $\pi : \bar{F} \to \mathbb{R}$ by $\pi(f)(s) = \hat{U}(f(s))$ for each $f \in \bar{F}$ and $s \in S$. This is a surjective function from $\bar{F}$ into $B(\bar{\mathcal{U}})$. We have $\pi(f) = \pi(g)$ iff $f \sim g$, since $\hat{U}$ represents $\succ$ on $\Delta_S(X)$ and (**E3**) holds. Thus π is bijective, so invertible. Since $\hat{U}$ is mixture preserving, π is also mixture preserving. Define a functional $I : B(\bar{\mathcal{U}}) \to \mathbb{R}$ by $I(\xi) = \hat{U}(\pi^{-1}(\xi))$. Now $I(\pi(f)) = U(f)$ for all $f \in \bar{F}$. Let 0 denote the constant utility act $0 = 1_\varnothing$ (that is, $0(s) = 0$ for all $s \in S$).[37] Define the constant act $p_0 = p_1 \oplus_{1/2} p_2$. We have $\pi(p_0)(s) = \hat{U}(p_1(s) \oplus_{1/2} p_2(s)) = \frac{1}{2}\hat{U}(p_1) + \frac{1}{2}\hat{U}(p_2) = 0$ for all $s \in S$, and thus $I(0) = U(p_0) = \hat{U}(p_0) = 0$.

Now, take any $\xi_1, \xi_2 \in B(\bar{\mathcal{U}})$ and any $\alpha \in [0,1]$. Take $f_1, f_2 \in \bar{F}$ such that $\xi_1 = \pi(f_1)$ and $\xi_2 = \pi(f_2)$. We have

$$\begin{aligned} I(\xi_1 \oplus_\alpha \xi_2) &= U(\pi^{-1}(\xi_1 \oplus_\alpha \xi_2)) \\ &= U(\pi^{-1}(\pi(f_1) \oplus_\alpha \pi(f_2))) \\ &= U(\pi^{-1}(\pi(f_1 \oplus_\alpha (f_2)))) \\ &= U(f_1 \oplus_\alpha f_2) \\ &= I(\pi(f_1) \oplus_\alpha \pi(f_2)). \end{aligned}$$

Hence I is mixture preserving. We now establish further properties of I that will allow us to apply a version of the Riesz representation theorem. Call a functional $G : B(\mathcal{U}) \to \mathbb{R}$ *norm-continuous* if for any sequence $\{\xi_n\}_{n=1}^\infty$ in $\bar{B}$ and any $\xi \in B(\mathcal{U})$, we have $\|\xi_n - \xi\|_\infty \to 0$ implies $\left|G(\xi_n) - G(\xi)\right| \to 0$, where $\|g\|_\infty = \sup_{s\in S}\left|g(s)\right|$ denotes the supremum norm on $B(\mathcal{U})$. That is, G is norm-continuous if $\xi_n \to \xi$ in the supremum norm topology implies $G(\xi_n) \to G(\xi)$ in the Euclidean metric topology in $\mathbb{R}$. The version of the Riesz representation theorem we will use is:

Lemma 20. *If $G : B(\mathcal{U}) \to \mathbb{R}$ is a norm-continuous linear functional with $G(1_S) = 1$, then*

$$G(\xi) = \int_S \xi(s)\mu(\mathrm{d}s)$$

for a unique probability charge μ on (S, Σ) defined by $\mu(E) = G(1_E)$ for all $E \in \Sigma$.

This is pretty involved result that we do not prove here.

Lemma 21. *Suppose $G : B \to \mathbb{R}$ is additive and monotone, and has $G(1_S) = 1$. Then G is norm-continuous.*

Proof. First we claim additivity (i.e. $G(\xi_1 + \xi_2) = G(\xi_1) + G(\xi_2)$) implies $G : \bar{B} \to \mathbb{R}$ is homogeneous for all rational numbers, i.e. $G(r\xi) = rI(\xi)$ for all $r \in \mathbb{Q}$ and $\xi \in \bar{B}$. Fix $\xi \in \bar{B}$. The case of $\xi = 0$ is immediate, and so is the case of natural numbers n, since $G(n\xi) = \sum_{i=1}^n G(\xi) = nG(\xi)$. Suppose $r = m/n$ for $n, m \in \mathbb{N}$. By additivity, $nG((m/n)\xi) = G(n(m/n)\xi) = G(m\xi) = mG(\xi)$. Next, $G(\xi) = -G(-\xi)$ since $0 = G(0) = G(\xi - \xi) = G(\xi) + G(-\xi)$ by additivity. Thus if r is a negative rational number, $G(r\xi) = G(-|r|\,\xi) = -|r|\,I(\xi) = rI(\xi)$.

Now take $\epsilon > 0$ and suppose $\{\xi_n\}_{n=1}^\infty$ converges to ξ in the supremum norm topology. Take $\delta \in \mathbb{Q}$ such that $0 < \delta < \epsilon$, and $N \in \mathbb{N}$ such that $\|\xi_n - \xi\|_\infty < \delta$ for all $n > N$.

[37] I use 1_E to denote the constant utility act that yields the indicator function of a set E in each state.

Then for all $n > N$, we have

$$\begin{aligned} G(\xi) - G(\xi_n) &= G(\xi) + G(-\xi_n) = G(\xi - \xi_n) \\ &\leq G(\|\xi - \xi_n\|_\infty 1_S) \leq G(\delta 1_S) \leq \delta G(1_S) = \delta < \epsilon. \end{aligned}$$

The first two inequalities are by monotonicity, and the third by homogeneity for rationals. An identical proof shows $G(\xi_n) - G(\xi) < \epsilon$. Hence G is norm-continuous. □

Lemma 22. *If a linear functional $G : B(\bar{\mathcal{U}}) \to \mathbb{R}$ is additive and monotone and has $G(1_S) = 1$, then there is a unique extension $G^* : B \to \mathbb{R}$ that is a linear norm-continuous functional.*

Proof. By the proof of Lemma 21, G is homogeneous for the rationals and norm-continuous in $B(\bar{\mathcal{U}})$. We claim G is homogeneous, i.e. $G(a\xi) = aG(\xi)$ for all $a \in \mathbb{R}$. Fix $\xi \in B$ and $a \in \mathbb{R}$, and let $\{r_n\}$ be a sequence of rational numbers with $r_n \to a$. Then $r_n\xi \to a\xi$ in the supremum norm topology, and since G is norm-continuous, $G(a\xi) = \lim_{n\to\infty} G(r_n\xi) = \lim_{n\to\infty} r_n G(\xi) = aG(\xi)$.

Now, define $G^*(\xi) = G(\xi)$ for $\xi \in B(\bar{\mathcal{U}})$. Any $\xi \in B - B(\bar{\mathcal{U}})$ can be written as $\xi = a\bar{\xi}$ for some $\bar{\xi} \in B(\bar{\mathcal{U}})$ and $a \in \mathbb{R}$. For such ξ, define $G^*(\xi) = aG(\bar{\xi})$. By homogeneity, this extension is a linear norm-continuous functional. □

We now claim that $I : B(\bar{\mathcal{U}}) \to \mathbb{R}$ has $I(1_S) = 1$ and is norm-continuous.

Note for the constant act p_1, we have $\pi(p_1) = (\hat{U}(p_1))_{s\in S} = 1_S$, and thus $I(1_S) = U(\pi^{-1}(1_S)) = U(p_1) = 1$, so $I(1_S) = 1$.

Since I is mixture preserving, $I(\alpha\xi) = I(\xi \oplus_\alpha 0) = \alpha I(\xi)$ for all $\xi \in B(\bar{\mathcal{U}})$. Thus for any $\xi_1, \xi_2 \in B(\bar{\mathcal{U}})$, we have $I(\xi_1 + \xi_2) = I((2\xi_1) \oplus_{1/2} (2\xi_2)) = \frac{1}{2}I(2\xi_1) + \frac{1}{2}I(2\xi_2) = I(\xi_1) + I(\xi_2)$, so I is additive.

Next, take $\xi_1, \xi_2 \in B(\bar{\mathcal{U}})$ and constant acts f_1, f_2 such that $\pi(f_1) = \xi_1$ and $\pi(f_2) = \xi_2$. Now $\xi_1 \geq \xi_2$ iff $\pi(f_1) \geq \pi(f_2)$, which implies $\hat{U}(f_1(s)) \geq \hat{U}(f_2(s))$ for all states $s \in S$. Since $\hat{U}$ represents $\succ$ on $\Delta(X)$, it follows that $f_1(s) \succsim f_2(s)$ for all $s \in S$. By (**E3**), we thus have $f_1 \succsim f_2$. This implies $U(f) \geq U(g)$, given U represents $\succ$ on F. Thus $I(\xi_1) \geq I(\xi_2)$. Hence I is monotone. By Lemma 21, I is norm-continuous on $B(\bar{\mathcal{U}})$, and by Lemma 22, there is a unique extension $I^* : B \to \mathbb{R}$ of I that is norm-continuous. By Lemma 20, we thus have

$$I(\xi) = \int_S \xi(s)\mu(\mathrm{d}s)$$

for probability charge μ defined by $\mu(E) = I(1_E)$. For all $f, g \in F$, we thus have $f \succ g$ iff $U(f) > U(g)$, iff $I(\pi(f)) > I(\pi(g))$, iff $\int_S \pi(f)(s)\mu(\mathrm{d}s) > \int_S \pi(g)(s)\mu(\mathrm{d}s)$, iff

$$U(f) = \int_S \hat{U}(f(s))\mu(\mathrm{d}s) > \int_S \hat{U}(g(s))\mu(\mathrm{d}s) = U(g),$$

which is the statement of the theorem.

The uniqueness statement is easy to prove as usual. □

The utility representation we have in Theorem 16 is a "double" expected utility representation. On the one hand,

$$\begin{aligned} U(f) &= \int_S \left(\sum_{x \in \operatorname{supp} f(s)} u(x) f(s)(x) \right) \mu(\mathrm{d}s) \\ &=: \int_S \hat{U}(f(s)) \mu(\mathrm{d}s) \end{aligned}$$

is a subjective expected utility representation of $\succ$ where we interpret

$$\hat{U}(p) = \sum_{x \in \operatorname{supp} p} u(x) p(x)$$

as a Bernoulli utility function for which the elements p of $\Delta(X)$ are consequences. But $\hat{U}$ is also itself a von Neumann-Morgenstern expected utility representation of the preferences on $\Delta(X)$ induced by $\succ$, with corresponding Bernoulli utility function $u : X \to \mathbb{R}$.

The first thing to note here is that the proof is much shorter than Savage's. This is because we could obtain the subjective expected utility representation via duality. That perhaps makes the Anscombe-Aumann framework a little less natural and intuitive than Savage's – Savage's proof may be long, but the general idea is easy to follow – we use small events as "measuring rods" to calibrate probabilities for the rest. You cannot really tell from the Anscombe-Aumann proof how calibration is going on. However, it turns out calibration is quite easy, because we can use the chance events to deduce subjective probabilities: for any event E, we need only find the number $\alpha \in [0, 1]$ such that the agent is indifferent between the lottery $\alpha\delta_x + (1 - \alpha)\delta_y$ (a constant act) and the subjective act $(\delta_x)_E(\delta_y)$. Then α is the probability the agent attaches to event E occurring.

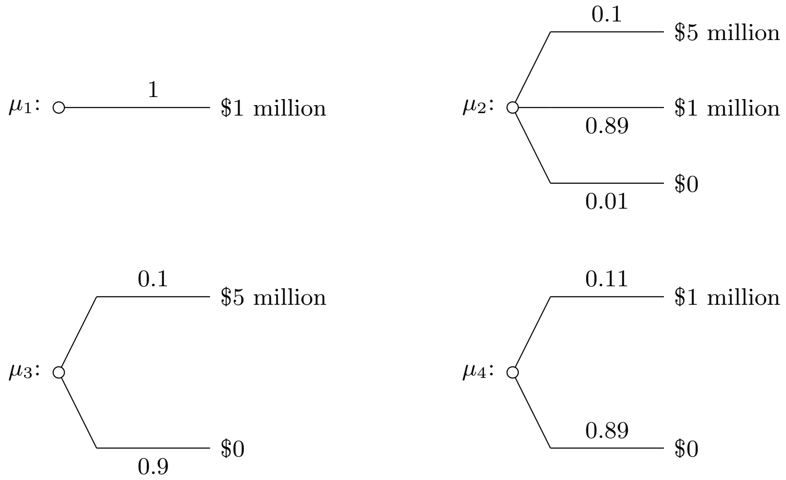
2.5 Behavioural problems II: violations of expected utility theory

The preceding models of decision under uncertainty are clear, beautiful theories, and represent some of the greatest achievements in decision theory. However, real world decisionmakers do not behave in the ways consistent with these theories – we can easily find violations of the basic axioms.

2.5.1 Violations of independence

The most famous violation of the basic axioms is the *Allais paradox* (Allais, 1953).

Example 17 (Allais, 1953)**.** Consider the following pairs of gambles:



Allais (1953), in an experiment that has been widely replicated, asked participants to indicate which of the gambles μ_1 and μ_2 they preferred and which of the gambles μ_3 and μ_4. The modal response by participants is $\mu_1 \succ \mu_2$ and $\mu_3 \succ \mu_4$. Savage himself would make the same choices (Savage, 1954, pp. 101-3), though he argued that after deliberation, his original preferences were mistaken.[38] However, $\mu_1 \succ \mu_2$ and $\mu_3 \succ \mu_4$ are together inconsistent with independence (**V1**) and the sure-thing principle (**P2**).

Let $X = \{0, 1, 5\}$ be the set of consequences, expressed in millions of dollars. We can show the violation of the sure-thing principle directly. Fix a state space $S = [0, 1]$. Let $f(s) = 1$ for all $s \in S$, let

$$g(s) = \begin{cases} 5 & \text{if } 0 \leq s \leq 0.1, \\ 0 & \text{otherwise.} \end{cases},$$

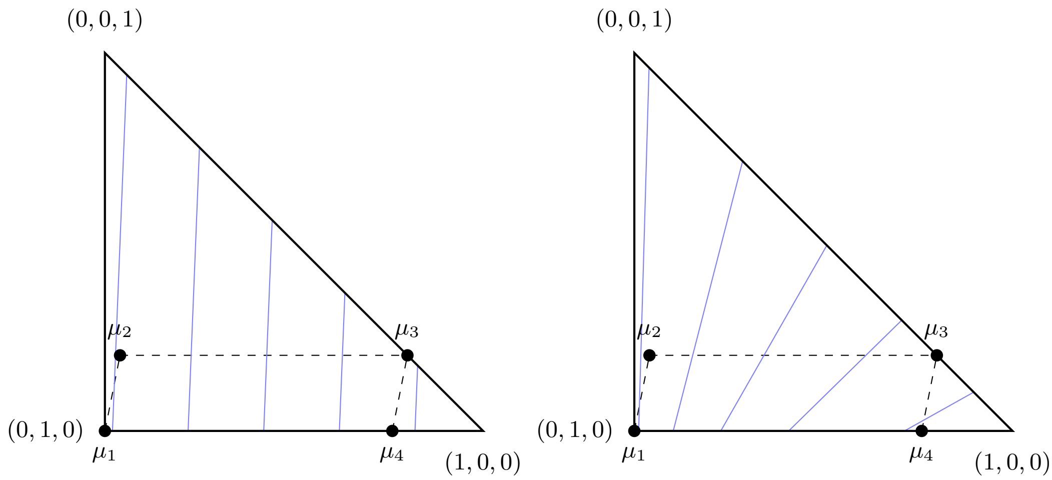
let $h(s) = 1$ for all $s \in S$, and let $h'(s) = 0$ for all $s \in S$. Finally, let $E = [0.11, 1]$. If decisionmakers place uniform probability on S,[39] then we can equate μ_1 with $f = f_E h$, μ_2 with $g_E h$, μ_3 with $g_E h'$ and μ_4 with $f_E h'$. Participants' modal preference is then $f_E h \succ g_E h$ and $g_E h' \succ f_E h'$, but (**P2**) states that $f_E h \succ g_E h$ iff $f_E h' \succ g_E h'$.

Since there is no $\alpha \in (0, 1)$ and $\nu \in \Delta(X)$ such that $\mu_3 = \mu_2 \oplus_\alpha \nu$ and $\mu_4 = \mu_1 \oplus_\alpha \nu$, the violation of (**V1**) is more subtle. However it is obvious by a visual argument. Since there are 3 consequences, we can represent the lotteries in a 2-dimensional simplex. Independence (**V1**) ensures the indifference curves for the decisionmaker are linear and parallel. However, parallel indifference lines are violated in Allais' experiment, as the

[38] One argument for theories like Savage's, even if they are easily falsified, is they provide a normative guide that provide a framework for thinking about our own preferences and ironing out inconsistencies. Savage clearly took his own theory as a normative guide, but there is not much evidence that other people find expected utility theory a useful tool for introspecting preferences. A number of classic experiments (MacCrimmon, 1968; Moskowitz, 1974; Slovic & Tversky, 1974) exposed participants to similar decision problems to Allais' and explained the arguments for and against expected utility to the participants, asking them whether they would like to revise their preferences. Few participants opted to make revisions.

[39] Given they are told the probabilities and payoffs, there is no reason for their beliefs to differ.

following figure shows.[40]



Allais' (1953) experiment represented in a simplex. The coordinates are probability vectors over $X = \{0, 1, 5\}$. In each simplex, utility increases as we move left and up. The left simplex gives the case where independence holds, and the right gives the case of "fanning" indifference curves.

The lotteries $(\mu_1, \mu_2, \mu_3, \mu_4)$ form a parallelogram in the simplex. $\mu_1 \succ \mu_2$ implies that your indifference lines must be very steep, as in the left simplex, so that there are indifference lines lying above μ_2 and below μ_1. But then there are indifference lines that lie above μ_3 and below μ_4, so $\mu_4 \succ \mu_3$. Indeed, $\mu_3 \succ \mu_4$ implies that your indifference curves are shallow in the vicinity of these lotteries, giving rise to a fanning out effect depicted in the right simplex.

The Allais paradox is not a contrived example – it is a special case of a more general effect known as the *common consequence effect*, attributable to MacCrimmon (1968), Kahneman & Tversky (1979) and others.

Example 18 (Common consequence effect)**.** Given a set of monetary outcomes X, fix an outcome $x \in X$ and let δ_x denote the degenerate lottery that yields payoff x for sure. Let μ be a lottery that places positive probability weight on some $y > x$ and on some $z < x$, and let μ' and μ'' be lotteries such that μ' first-order stochastically dominates μ''. Finally, take $\alpha \in (0, 1)$.

In experiments exhibiting the common consequence paradox, participants are asked to compare the pair of lotteries

$$\mu_1 = \alpha\delta_x + (1-\alpha)\mu' \qquad \text{and} \qquad \mu_2 = \alpha\mu + (1-\alpha)\mu',$$

and to compare the pair of lotteries

$$\mu_3 = \alpha\delta_x + (1-\alpha)\mu'' \qquad \text{and} \qquad \mu_2 = \alpha\mu + (1-\alpha)\mu''.$$

[40]The simplexes depicted in the figure are sometimes known as *Marschak-Machina triangles.*

Intuitively, we can think of each of μ_1 through μ_4 as compound lotteries, where which lottery we end up playing depends on the outcome of an α-weighted coin.

Both independence (**V1**) and the sure-thing principle (**P2**) imply either (i) $\mu_1 \succ \mu_2$ and $\mu_3 \succ \mu_4$ or (ii) $\mu_2 \succ \mu_1$ and $\mu_4 \succ \mu_3$. Yet the most common preference expressed by participants in such experiments is $\mu_1 \succ \mu_2$ and $\mu_4 \succ \mu_3$. Thus there is a "fanning out" of the indifference curves as we discussed in Example 17.

Independence and the sure-thing principle impose that your preferences over what occurs in some event should not depend on what occurs in other events. Under the compound lottery coin-flip interpretation of μ_1 through μ_4, this says your preferences over the lotteries that you play when the coin lands heads should not depend on what lottery you play when the coin lands tails.

However, the common consequence effect suggests participants act as though they are more risk averse about what happens when the coin lands heads when they are better off if the coin lands tails. Or paraphrasing Ken Arrow (Machina, 1987, footnote 10): participants act as though they are more risk averse in the event of an opportunity loss and less risk averse in the event of an opportunity gain.[41]

Finally, the other famous Allais-type effect, often described as the "easy" version of the Allais paradox, is the *common ratio effect.* Allais (1953) discusses this himself but Kahneman & Tversky (1979) seems to have become the classic reference.

Example 19 (Common ratio effect)**.** Let $0 < a < b$ be monetary gains, let $p, q \in (0,1)$ with $p > q$, and let $r \in (0,1)$. Put $X = \{0, a, b\}$.

In experiments exhibiting the common ratio paradox, participants are asked to compare the gambles

$$\mu_1 : \begin{cases} a & \text{with probability } p, \\ 0 & \text{with probability } 1-p, \end{cases} \qquad \text{and} \qquad \mu_2 : \begin{cases} b & \text{with probability } q, \\ 0 & \text{with probability } 1-q, \end{cases}$$

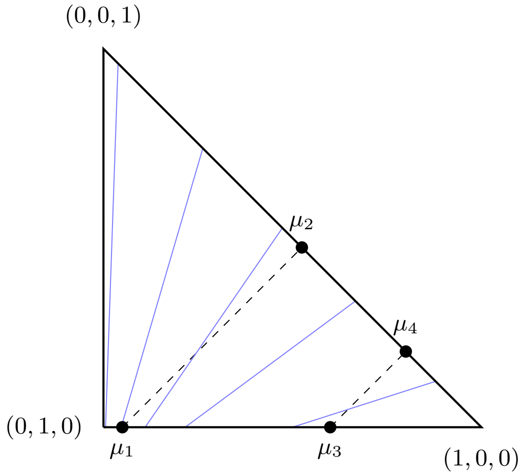
and to compare the gambles

$$\mu_3 : \begin{cases} a & \text{with probability } rp, \\ 0 & \text{with probability } 1-rp, \end{cases} \qquad \text{and} \qquad \mu_4 : \begin{cases} b & \text{with probability } rq, \\ 0 & \text{with probability } 1-rq. \end{cases}$$

Again, independence (**V1**) and the sure-thing principle (**P2**) imply either (i) $\mu_1 \succ \mu_2$ and $\mu_3 \succ \mu_4$ or (ii) $\mu_2 \succ \mu_1$ and $\mu_4 \succ \mu_3$. However, in experiments participants tend to depart from (i) and (ii), instead often opting for $\mu_1 \succ \mu_2$ and $\mu_4 \succ \mu_3$ (e.g. de Neufville,

[41] Mas-Colell, Whinston & Green (1995) describe an example (which they label Machina's paradox) that illustrates this point intuitively. Suppose we have three potential prizes: a free trip to Venice, a ticket to a critically acclaimed film about Venice, and nothing. Suppose lottery μ_1 offers a 1/10 chance of winning a trip to Venice and nothing otherwise; suppose lottery μ_2 offers a 2/5 chance of winning a ticket to the film, and nothing otherwise; suppose μ_3 offers a 1/10 chance of winning the trip, a 2/5 chance of winning the film ticket, and nothing otherwise. Suppose μ_4 offers nothing for sure. It is plausible you might have preferences $\mu_1 \succ \mu_2 \succ \mu_3 \succ \mu_4$, because you probably will not enjoy the film when you had the possibility of winning the trip.

1983, 1984). If a, b represent losses rather than gains, then it is common for participants to report $\mu_2 \succ \mu_1$ and $\mu_3 \succ \mu_4$ (e.g. Kahneman & Tversky, 1979). These findings seem to hold not just for humans but also for rats! Battalio, Kagel & MacDonald (1985), for example, offered lab rats gambles over food and record choices that exhibit the common ratio effect.



The common ratio effect in a simplex, with fanning indifference curves.

The main takeaway from these phenomena is that people do not act as though they have consistent attitudes to risk, where consistency has the sense of independence in (**V1**). Tversky & Kahneman (1992) identify a *fourfold pattern* of risk attitudes that they propose decisionmakers follow:

- Decisionmakers tend to be risk seeking over low probability gains.
- Decisionmakers tend to be risk averse over high probability gains.
- Decisionmakers tend to be risk averse over low probability losses.
- Decisionmakers tend to be risk seeking over high probability losses.

There is a body of experimental literature (including the experiment in Tversky & Kahneman, 1992) that supports this fourfold pattern. One ingredient of an explanation for this pattern is the following. It is very well-established in the experimental literature that people overweight small probabilities and underweight large probabilities. Rather than take a 0.001 probability of winning (or losing) $1000 at face value, decisionmakers often act as though they believe the probability is larger. Likewise, if offered a lottery that pays off $1000 with probability 0.99 and $0 otherwise, decisionmakers often act as though the chance of winning is lower than 0.99.[42] This lends itself to a generalization

[42] A similar effect occurs if you ask people to give their estimates of populations. For example, Kardosh, Sklar, Goldstein & Hassin (2022) ran experiments where they showed participants image matrices consisting of many faces varying in demographic, and asked the participants to estimate the proportion of the minority group in each matrix. Participants consistently overestimated the proportion of the minority group.

of expected utility theory that we discuss in section **??**.

It is worth noting that while there is evidence for the fourfold pattern in many experiments, it is not a completely robust finding. Harbaugh, Krause & Vesterlund (2010), for example, find that the fourfold pattern is not robust to the way we measure preference in experiments. In their study, participants exhibited the fourfold pattern when their preferences were elicited using a price procedure (eliciting their willingness to pay for each given lottery). However, when participants were asked to choose between risky lotteries and the corresponding expected values, participants' choices were indistinguishable from random choice. This is essentially the preference reversal phenomenon (section 2.5.3).

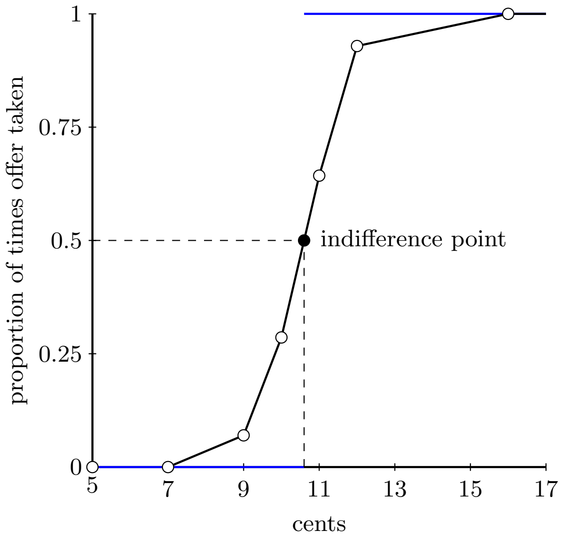
2.5.2 Choice stochasticity

In settings where we expect decisionmakers not to be indifferent over consequences, expected utility theory implies that decisionmakers should have sharp indifference points. An experimental literature, going back to Mosteller & Nogee (1951) shows this is not the case.

Example 20 (Mosteller & Nogee, 1951)**.** Mosteller and Nogee devised a poker dice game whereby participants played in tables of 5. Each player started with \$1 worth of chips in 5 cent increments (this was slightly more money than it sounds now, back in 1951). A hand in the game is a set of 5 numbers obtained by rolling 5 dice. As in poker, hands had different relative strengths – in order from lowest to highest possible hands, the possibilities were nothing, then one pair, two pair, three of a kind, full house, four of a kind and five of a kind. Straights such as 12345 had no value.

In each round, the table was shown a stimulus card, which listed a hand (e.g. 55221), and a monetary value. If a player chose to bet, they would bet 5 cents against rolling a better hand than the one listed on the stimulus card, winning the monetary offer listed on the card if successful. Each player had access to a list of all the true odds of winning against the different stimulus card hands. Fixing a hand, the different stimulus cards varied in the offers they showed. For example, for the stimulus cards with hand 55221, the offer would vary between 7 cents and 16 cents, and the odds of winning against 55221 was $2.01 : 1$.

Mosteller and Nogee tracked the betting behaviour of each player. The idea is that by looking at the proportion of times a player chose to bet given the offer on the stimulus card for each given hand, it is possible to find the point at which the player is indifferent between taking the gamble and keeping their present chips for sure. Thus Mosteller and Nogee could trace out a utility curve for each player, up to scale, by looking at the indifference points across hands. They identified an indifference point for a given hand as the point at which the player takes the offer 50% of the time.



Mosteller & Nogee's data (hollow circles) for one of their participants, playing against hand 55221. The curve in blue shows the behaviour we should expect if the participant behaved in accordance with von Neumann-Morgenstern expected utility theory, while the curve in black shows imputed actual behaviour.

The Mosteller-Nogee approach to "measuring" utility is a nice idea, but the figure shows a problem: when we look at participants behaviour, their choices are considerably more random than expected utility theory would predict. Except at the indifference point, we should expect the decisionmaker to either always take the gamble or never take it. In practice, decisionmakers tend to take the gambles some of the time, tending to the extremes as we get further from the indifference point.

2.5.3 Preference reversal and intransitivity

Under expected utility theory, the ordering of monetary lotteries by their certainty equivalents matches the ordering of the lotteries according to $\succ$. The certainty equivalent is effectively a valuation for a lottery – if the decisionmaker's certainty equivalent for a lottery μ is c, then that decisionmaker will be indifferent between maintaining her current wealth for sure and paying c to 'buy' the lottery μ.

Experimental evidence suggests this is not so. In a widely replicated experiment, Lichtenstein & Slovic (1971, 1973) presented participants with pairs of bets, each pair being a choice between a "P-bet" and a "\$-bet":

$$P\text{-bet: } \begin{cases} a_1 & \text{with probability } p, \\ a_2 & \text{with probability } 1-p, \end{cases} \quad \text{and} \quad \$\text{-bet: } \begin{cases} b_1 & \text{with probability } q, \\ b_2 & \text{with probability } 1-q, \end{cases}$$

where $p > q$, $a_1 > a_2$, $b_1 > b_2$ and $b_1 > a_1$. Thus the P-bet offers a higher chance of winning but the \$-bet offers a more valuable jackpot.

Participants were coaxed into giving their certainty equivalent for each lottery. The most elaborate (and strategyproof) way of doing this was to ask participants to state a price – the experimenter then randomly draws a price, and if it exceeds the participants' price, the participant receives the price and forgoes the lottery, and participates in the lottery otherwise; truthful reporting is the dominant action here. Other ways were to simply ask participants to state their minimum selling price if they were to own the lottery or their maximum buying price if they were to buy it.

Participants were also asked to choose which lottery they prefer in the pair. It turns out that participants tend to value the \$-bet higher than the P-bet, in terms of certainty equivalent. However, they tend to prefer the P-bet when asked to choose between them! This effect is known as the *preference reversal phenomenon*.

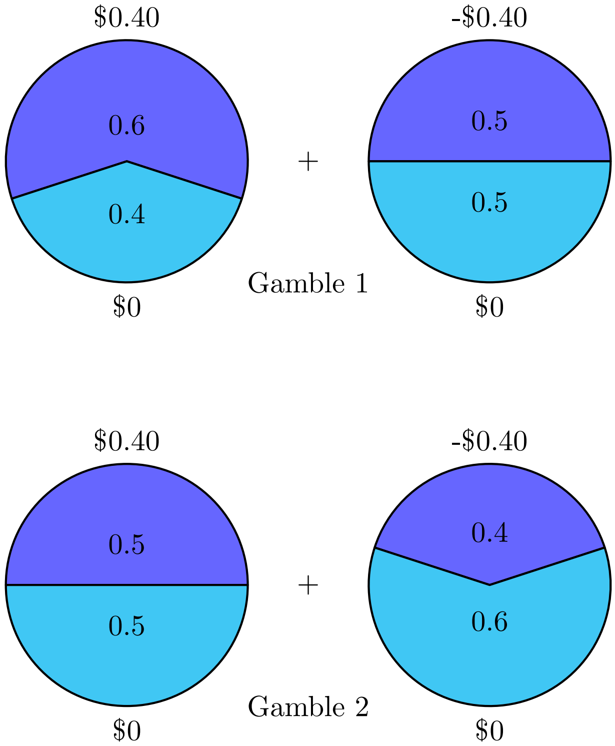
How do we explain this behaviour? A very basic assumption of the traditional economic approach is that decisionmakers have a stable preference relation over alternatives. By revealed preference, a decisionmaker exhibiting preference reversal strictly prefers the P-bet to the \$-bet. Moreover, they must be indifferent between the P-bet and the certainty equivalent on the P-bet, and they must be indifferent between the \$-bet and the certainty equivalent $c_2 > c_1$ on the \$-bet. This suggests their preferences are cyclic or intransitive, violating the weak order assumption on $\succ$ in expected utility theory.

Psychologists (like Lichtenstein and Slovic) do not buy the assumption that decisionmakers have stable global preferences in the first place. They instead focus on *response mode effects* – the idea being that how the decisionmaker arrives at a choice and a valuation comes from two separate processes that involve different ways of processing information. In choosing, people are more concerned about the probability of winning, whereas in coming up with a valuation, people focus more on the monetary amounts involved.

2.5.4 Framing

Framing effects apply to decisions under uncertainty just as they apply to decisions under certainty. A heavily replicated finding is that choice is sensitive to whether a risky gamble is framed in terms of a gain or a loss.

Example 21 (Duplex gambles; Payne & Braunstein, 1971)**.** A *duplex gamble* involves two wheels, a gain wheel and a loss wheel. Each wheel has a pointer and where the wheel lands on the pointer when spun determines the outcome of the wheel. The outcome of the gamble is the sum of the outcomes of the gain and loss wheels.



Two duplex gambles inducing identical distributions over consequences.

When presented with two duplex gambles that have the same underlying distribution over outcomes, as above, expected utility theory predicts that the decisionmaker should be indifferent between the two gambles. Yet Payne & Braunstein (1971) found participants asked to express preferences over such gambles in experiments expressed strict preferences. In particular, participants tend to strictly prefer gambles where the probability of winning on the gain wheel is greater, i.e. preferring gamble 1 to gamble 2 in the above figure.

Tversky & Kahneman (1981), allowing participants to construct a duplex gamble by choosing from a pair of gain wheels and a pair of loss wheels, find some participants end up choosing stochastically dominated gambles.

An important framing effect involves *reference points* – points against which gains or losses are judged. This is an idea originating with Markowitz (1952), and most famously associated with Tversky and Kahneman's prospect theory.

Example 22 (Tversky & Kahneman, 1981, 1986)**.** Tversky and Kahneman gave participants the following initial information:

> Imagine that the U.S. is preparing for the outbreak of an unusual Asian disease, which is expected to kill 600 people. Two alternative programs to combat the disease have been proposed. Assume that the exact scientific estimate of the consequences of the programs are as follows.

They then gave participants one of two alternative descriptions of two identical programmes:

(i) If Program A is adopted, 200 people will be saved.

If Program B is adopted, there is 1/3 probability that 600 people will be saved, and 2/3 probability that no people will be saved.

(ii) If Program C is adopted 400 people will die.

If Program D is adopted there is 1/3 probability that nobody will die, and 2/3 probability that 600 people will die.

Programmes A and C are equivalent and programmes B and D are equivalent. However, while 72% of participants given description (i) preferred A to B, only 22% of participants given description (ii) preferred C to D. Whether the options are framed in terms of a gain (relative to the expected death toll) or a loss (relative to the status quo ante) shifts participants' risk stance from risk averse to risk preferring.

2.5.5 Calibration and small stake risk aversion

Expected utility theory implies that risk averse decisionmakers are approximately risk neutral over small enough bets. However, in practice, participants in experiments often appear risk averse even when the value of the bets they are choosing over are tiny – on the order of cents rather than dollars. Rabin's (2000) famous calibration theorem shows this behaviour would imply decisionmakers should be extremely risk averse over large stake gambles if expected utility theory holds. The version of the theorem here is a streamlined version due to Balter, Chau & Schweizer (2022), which gives slightly tighter bounds.

Theorem 17 (Rabin's calibration theorem)**.** *Suppose $u : \mathbb{R} \to \mathbb{R}$ is increasing and weakly concave, and suppose there are numbers $\bar{x} > \underline{x}$ and $g > \ell > 0$ such that $u(x+g) - u(x) \leq u(x) - u(x-\ell)$ for all wealth levels $x \in [\underline{x}, \bar{x}]$.*

For any pair of integers k, m and any wealth level $x \in \mathbb{R}$ such that $[x - kg, x + mg] \subseteq [\underline{x}, \bar{x}]$, we have that

$$u(x+mg) - u(x) \leq \frac{1 - \left(\frac{\ell}{g}\right)^m}{1 - \frac{\ell}{g}}[u(x+g) - u(x)], \quad \textit{and}$$

$$u(x) - u(x-kg) \geq \frac{\left(\frac{g}{\ell}\right)^k - 1}{1 - \frac{\ell}{g}}[u(x+g) - u(x)].$$

Proof. Note $\frac{\ell}{g}(x-g) + (1-\ell/g)x = x - \ell$. Thus

$$\frac{\ell}{g}u(x-g) + (1-\ell/g)u(x) \leq u(x-\ell)$$

by concavity of u, which rearranges to $u(x) - u(x-\ell) \leq \frac{\ell}{g}[u(x) - u(x-g)]$. Iterating this, we have

$$u(x+(i+1)g) - u(x+ig) \leq \left(\frac{\ell}{g}\right)^i [u(x+g) - u(x)]$$

for $0 \leq i < m$. Now,

$$\begin{aligned} u(x+mg) - u(x) &= \sum_{i=0}^{m-1} [u(x+(i+1)g) - u(x+ig)] \\ &\leq \sum_{i=0}^{m-1} \left(\frac{\ell}{g}\right)^i [u(x+g) - u(x)] = \frac{1-\left(\frac{\ell}{g}\right)^m}{1-\frac{\ell}{g}} [u(x+g) - u(x)]. \end{aligned}$$

This establishes the upper bound on $u(x+mg) - u(x)$. Similarly, iterating $u(x) - u(x-g) \geq \frac{g}{\ell}[u(x+g) - u(x)]$ gives

$$u(x-jg) - u(x-(j+1)g) \geq \left(\frac{g}{\ell}\right)^{j+1} [u(x+g) - u(x)]$$

for $0 \leq j < k$. Thus

$$\begin{aligned} u(x) - u(x-kg) &= \sum_{j=0}^{k-1} [u(x-jg) - u(x-(j+1)g)] \\ &\geq \frac{g}{\ell} \sum_{j=0}^{k-1} \left(\frac{g}{\ell}\right)^j [u(x+g) - u(x)] = \frac{\left(\frac{g}{\ell}\right)^k - 1}{1-\frac{\ell}{g}} [u(x+g) - u(x)]. \end{aligned}$$

This establishes the lower bound on $u(x) - u(x-kg)$. $\square$

The assumption $u(x+g) - u(x) \leq u(x) - u(x-\ell)$ is equivalent to $\frac{1}{2}u(x+g) + \frac{1}{2}u(x-\ell) \leq u(x)$, i.e. for all wealth levels $x \in [\underline{x}, \bar{x}]$, the decisionmaker would reject an even gamble between a gain of g and a loss of ℓ versus maintaining current wealth x for sure.

Theorem 17 implies risk aversion over small gambles implies implausibly extreme risk aversion over large gambles. In particular, suppose you are unwilling to take a 50-50 lose \$100, gain \$110 bet for any initial wealth level below \$300,000. Then if your initial wealth is \$290,000, you would be unwilling to take a 50-50 lose \$10,000, gain \$36 billion bet! Nor would you be willing to risk a \$6000 loss for a \$1 billion gain.

2.5.6 Ambiguity

A long tradition in economics and the philosophy of probability, going back to at least Knight (1928), draws a distinction between quantifiable uncertainty (risk) and unquantifiable uncertainty (sometimes called *Knightian uncertainty*). Subjective expected utility theory implies all uncertainty a decisionmaker faces is subjectively quantifiable – a decisionmakers' beliefs can be summarized by a single probability measure over the state space. However, there is plenty of evidence that decisionmakers do not treat all forms of uncertainty alike. Ellsberg (1961) famously demonstrated this with his urn thought experiments.[43]

Example 23 (Ellsberg paradox)**.** Ellsberg (1961) considers two different thought experiments. While these are sometimes called experiments, they are thought experiments – Ellsberg did not run experiments as a modern experimental economist would. He simply asked colleagues what they would prefer.

(a) *Two urn paradox.*[44] Imagine there are two urns, Urn 1 and Urn 2. Each urn contains 100 balls, and each ball might be either red or black. For each urn i, you can choose to either bet on Red_i, or bet on Black_i. Having placed your bet, a ball will be randomly drawn from the urn. If the drawn ball is the colour you bet, then you win \$100, and if it is a different colour, you win \$0. You know that Urn 2 contains 50 red balls and 50 black balls (you were allowed to check). However, you have no information about how many balls of each colour there are in Urn 1. Consider the following questions:

   (i) Which do you prefer to bet on, Red_1 or Black_1; or are you indifferent?

   (ii) Which would you prefer to bet on, Red_2 or Black_2?

   (iii) Which would you prefer to bet on, Red_1 or Red_2?

   (iv) Which would you prefer to bet on, Black_1 or Black_2?

   In response to (i) and (ii), the typical response is indifference. However, Ellsberg found that most of those he surveyed preferred Red_2 to Red_1 and Black_2 to Black_1.

   This contradicts Savage's axioms. Suppose your preferences satisfy the axioms. If you prefer Red_2 to Red_1, then since the prizes are the same in both cases, we conclude you consider randomly drawing a Red_2 ball from Urn 2 more likely than randomly drawing a Red_1 ball from Urn 1. But since you also prefer Black_2 to Black_1, we conclude you consider randomly drawing a Black_2 ball from Urn 2 more likely than randomly drawing a Black_1 ball from Urn 1. But these two conclusions contradict each other! The violation here is of weak order (**P1**) and the sure-thing principle (**P2**).

(b) *One urn paradox.* Now there is one urn of 90 balls, and each ball might be red, black or yellow. You know there are exactly 30 red balls in the urn. However, you do not know the ratio of yellow to black balls. First suppose you can bet either on Red or on Black. A ball is randomly drawn from the urn and if it is the colour you bet, then you win \$100, and otherwise you win nothing. Second, suppose you can bet either on Red/Yellow or on Black/Yellow. Again, you receive \$100 if the randomly drawn ball matches one of the two colours.

[43] He is better known for other work.

[44] Keynes (1921) also considered this version of the paradox.

> A common response in these two cases is to prefer to bet on Red in the first case and prefer to bet on Black/Yellow in the second. This directly violates the sure-thing principle (**P2**), because we have altered both of the first two acts in the same way (changing from winning \$0 on drawing yellow to winning \$100 on drawing yellow) on the same event (drawing a yellow).

Ellsberg's article gives the responses of quite a few economists by name. Samuelson and Debreu don't violate Savage's axioms. Marschak did, proudly. Raiffa violated the axioms and felt guilty about it. What this says about them, I do not know.

Ellsberg's examples highlight the phenomenon of *ambiguity aversion* – aside from risk aversion, decisionmakers can also be averse to options where the uncertainty cannot be easily quantified. There is a large experimental literature on ambiguity aversion, confirming the conclusions of Ellsberg's non-experimental data. Recently, there have also been efforts to measure ambiguity aversion in real-world settings. For example, Brenner & Izhakian (2018) conclude ambiguity aversion is reflected in prices in equity markets.