Rational Choice

Itzhak Gilboa

The MIT Press
Cambridge, Massachusetts
London, England

Appendixes

For information about special quantity discounts, please email special_sales@mitpress.mit.edu.

This book was set in Palatino on 3B2 by Asco Typesetters, Hong Kong.
Printed and bound in the United States of America.

Library of Congress Cataloging-in-Publication Data

Gilboa, Itzhak.
Rational choice / Itzhak Gilboa.
p. cm.
Includes bibliographical references and index.
ISBN 978-0-262-01400-7 (hardcover : alk. paper)
1. Rational choice theory. 2. Social choice. 3. Decision making. 4. Game theory.
5. Microeconomics. I. Title.
HM495.G55 2010
302′.13—dc22 2009037119

10 9 8 7 6 5 4 3 2 1

Contents

A Mathematical Preliminaries

A.1 Notation

$\sum$	summation $$\sum_{i=1}^{n} a_i = a_1 + a_2 + \cdots + a_n.$$
$\prod$	product (multiplication) $$\prod_{i=1}^{n} a_i = a_1 a_2 \ldots a_n.$$
$\forall$	for every
$\exists$	there exists
$\Rightarrow$	implies
$\Leftrightarrow$	if and only if (implies and is implied by)
iff	if and only if
$\cup$	union (of sets)
$\cap$	intersection (of sets)
A^c	the complement of (the set) A
$\subset$	a subset of
$\in$	belongs to, member of
$\notin$	doesn't belong to, not a member of
$\emptyset$	the empty set
$f : A \to B$	a function from the set A to the set B
$A \times B$	the product set

$\mathbb{R}$	the set of real numbers
$\mathbb{R}^n$	the n-dimensional Euclidean space
$[a, b]$	a closed interval $\{x \in \mathbb{R} \mid a \leq x \leq b\}$
(a, b)	an open interval $\{x \in \mathbb{R} \mid a < x < b\}$
$\|\cdot\|$	norm, length of a vector
$\#, \lvert\cdot\rvert$	cardinality of a set; if the set is denoted A, then $\#A = \lvert A\rvert$ denotes its cardinality.

A.2 Sets

A set is a primitive notion. Sets are often denoted by capital letters, A, B, $C, \ldots$ and indicated by braces $\{\ \}$. Inside these braces are listed the elements of the set. For instance, $A = \{0, 1\}$ refers to the set consisting of 0 and 1. Sets can also be described without listing all elements explicitly inside the braces. For instance,

$$N = \{1, \ldots, n\}$$

denotes the set of all natural numbers from 1 to n. Similarly, we define

$$\mathbb{N} = \{1, 2, 3, \ldots\}$$

and

$$\mathbb{Z} = \{\ldots, -1, 0, 1, \ldots\}$$

to be the set of natural numbers and the set of integer numbers, respectively.

The notation $a \in A$ means that a is a *member* of the set A or that a *belongs to* A. $a \notin A$ is the negation of $a \in A$.

The symbol $\subset$ designates a relation between sets, meaning "is a subset of." Explicitly, $A \subset B$ means that A is a subset of B, that is, for all $x \in A$, it is true that $x \in B$. Thus, $x \in A$ iff $\{x\} \subset A$.

The symbol $\varnothing$ denotes the empty set, the set that has no elements.

Sets are also defined by a certain condition that elements should satisfy. For instance,

$$A = \{n \in \mathbb{N} \mid n > 3\}$$

denotes all the natural numbers greater than 3, that is, $A = \{4, 5, 6, \ldots\}$.

$\mathbb{R}$ denotes the set of real numbers. I don't define them here formally, although they can be defined using the rationals, which are, in turn, defined as the ratios of integer numbers.

When we use mathematics to model reality, we also refer to sets whose elements need not be mathematical objects. For instance,

$A = \{humans\},$

$B = \{mammals\}.$

Such sets are viewed as sets of some mathematical objects interpreted as humans or mammals, respectively. Thus, when we discuss a set of individuals, alternatives, strategies, or states of the world, we mean a set whose elements are interpreted as individuals, alternatives, and so on.

The basic set operations are as follows.

Union ($\cup$). A binary operation on sets, resulting in a set containing all the elements that are in at least one of the sets. Or, for sets A and B,

$$A \cup B = \{x \mid x \in A \quad \text{or} \quad x \in B\}.$$

Here and elsewhere, *or* is inclusive, that is, "p or q" means "p, or q, or possibly both."

Intersection ($\cap$). A binary operation resulting in elements that are in both sets. That is,

$$A \cap B = \{x \mid x \in A \quad \text{and} \quad x \in B\}$$

Two sets A and B are *disjoint* if they have an empty intersection, that is, if $A \cap B = \varnothing$.

Complement (^{c}). A unary operation containing all elements that are not in the set. To define it, we need a reference set. That is, if S is the entire universe,

$$A^c = \{x \mid x \notin A\}.$$

You may verify that

$$(A^c)^c = A,$$

$$A \cap B \subset A, B \subset A \cup B,$$

$$A \cap A^c = \varnothing,$$

$$(A \cup B)^c = A^c \cap B^c,$$

$$(A \cap B)^c = A^c \cup B^c,$$

and

$A \cap (B \cup C) = (A \cap B) \cup (A \cap C),$

$A \cup (B \cap C) = (A \cup B) \cap (A \cup C).$

Given two sets A and B, we define their *(Cartesian) product* $A \times B$ to be all the ordered pairs whose first element is from A and whose second element is from B. In formal notation,

$$A \times B = \{(x, y) \mid x \in A \quad \text{and} \quad y \in B\}.$$

Note that (x, y) is an *ordered pair* because the order matters. That is, $(x, y) \neq (y, x)$ unless $x = y$. This is distinct from the set containing x and y, in which the order does not matter. That is, $\{x, y\} = \{y, x\}$.

The notation A^2 means $A \times A$. Thus, it refers to the set of all the ordered pairs each element of which is in A. Similarly, we define

$$A^n = A \times \cdots \times A = \{(x_1, \ldots, x_n) \mid x_i \in A, i \leq n\}.$$

The *power set* of a set A is the set of all subsets of A. It is denoted

$$2^A = P(A) = \{B \mid B \subset A\}.$$

A.3 Relations and Functions

A binary relation is a subset of ordered pairs. Specifically, if R is a *binary relation* from a set A to a set B, we mean that

$$R \subset A \times B.$$

This is an extensional definition. The relation R is defined by a list of all pairs of elements in A and in B such that the former relates to the latter. For instance, consider the relation R, "located in," from the set of buildings A to the set of cities B. Then, if we have

$$R = \left\{ \begin{array}{c} (\text{Empire\_State\_Building,New\_York}), \\ (\text{Louvre,Paris}), \\ (\text{Big\_Ben,London}), \ldots \end{array} \right\}$$

we wish to say that the Empire State Building relates to New York by the relation "located in," that is, it is in New York; the building of the Louvre is in Paris; and so forth.

For a relation $R \subset A \times B$ we can define the inverse relation, $R^{-1} \subset B \times A$ by

$$R^{-1} = \{(y, x) \mid (x, y) \subset R\}.$$

Of particular interest are relations between elements of the same set. For a set A, a binary relation on A is a relation $R \subset A^2 (= A \times A)$. For instance, if A is the set of people, then "child_of" is a relation given by

$$R = \left\{ \begin{array}{c} (\text{Cain},\text{Adam}), \\ (\text{Cain},\text{Eve}), \\ (\text{Abel},\text{Adam}), \ldots \end{array} \right\},$$

and the relation "parent_of" will be

$$R^{-1} = \left\{ \begin{array}{c} (\text{Adam},\text{Cain}), \\ (\text{Eve},\text{Cain}), \\ (\text{Adam},\text{Abel}), \ldots \end{array} \right\}.$$

A *function* f from A to B, denoted

$$f : A \to B,$$

is a binary relation $R \subset A \times B$ such that for every $x \in A$, there exists precisely one $y \in B$ such that $(x, y) \in R$. We then write

$$f(x) = y$$

or

$$f : x \mapsto y.$$

The latter is also used to specify the function by a formula. For instance, we can think of the square function $f : \mathbb{R} \to \mathbb{R}$ defined by

$$f(x) = x^2$$

or write

$$f : x \mapsto x^2.$$

A function $f : A \to B$ is 1–1 (*one-to-one*) if it never attaches the same $y \in B$ to different $x_1, x_2 \in A$, that is, if

$$f(x_1) = f(x_2) \Rightarrow x_1 = x_2.$$

A function $f : A \to B$ is *onto* if every $y \in B$ has at least one $x \in A$ such that $f(x) = y$.

If $f : A \to B$ is both one-to-one and onto, we can define its inverse

$$f^{-1} : B \to A$$

by

$$f^{-1}(y) = x \Leftrightarrow y = f(x).$$

Observe that the notation f^{-1} is consistent with the notation R^{-1} for relations. Recalling that a function is a relation, one can always define f^{-1} as

$$f^{-1} = \{(y, x) \mid y = f(x)\},$$

and if f is one-to-one and onto, this relation is indeed a function, and it coincides with the inverse function of f.

We often also use the notation

$$f^{-1}(x) = \{y \in B \mid y = f(x)\}.$$

With this notation, to say that f is one-to-one is equivalent to saying that $f^{-1}(x)$ has at most one element for every x. To say that it is onto is equivalent to saying that $f^{-1}(x)$ is nonempty. And if f is both one-to-one and onto (a *bijection*), $f^{-1}(x)$ has exactly one element for each x. Then, according to the set notation, for a particular y,

$$f^{-1}(x) = \{y\},$$

and according to the inverse function notation,

$$f^{-1}(x) = y.$$

Using f^{-1} both for the element y and for the set containing only y seems problematic when one is just starting to deal with formal models, but it becomes more common as one advances. This is called an abuse of notation, and it is often acceptable as long as readers know what is meant by it.

Interesting properties of binary relations on a set $R \subset A^2$ include the following.

R is *reflexive* if $\forall x \in A$, xRx, that is, every x relates to itself. For instance, the relations $=$ and $\geq$ on $\mathbb{R}$ are reflexive, but $>$ isn't.

R is *symmetric* if $\forall x, y \in A$, xRy implies yRx, that is, if x relates y, then the converse also holds. For instance, the relation $=$ (on $\mathbb{R}$) is symmetric, but $\geq$ and $>$ aren't. Notice that $>$ does not allow any pair x, y to have both $x > y$ and $y > x$, that is,

$$> \cap >^{-1} = > \cap < = \varnothing,$$

whereas $\geq$ does because if $x = y$, it is true that $x \geq y$ and $y \geq x$. But $\geq$ is not symmetric because it is not always the case that xRy implies yRx.

R is *transitive* if $\forall x, y, z \in A$, xRy and yRz imply xRz, that is, if x relates to z through y, then x relates to z also directly. For example, $=$, $\geq$, and $>$ on $\mathbb{R}$ are all transitive, but the relation "close to" defined by

$$xRy \Leftrightarrow |x - y| < 1$$

is not transitive.

A relation that is reflexive, symmetric, and transitive is called an *equivalence relation*. Equality $=$ is such a relation. Also, "having the same square," that is,

$$xRy \Leftrightarrow x^2 = y^2,$$

is an equivalence relation.

In fact, a relation R on a set A is an equivalence relation if and only if there exist a set B and a function $f : A \to B$ such that

$$xRy \Leftrightarrow f(x) = f(y).$$

A.4 Cardinalities of Sets

The cardinality of a set A, denoted $\#A$ or $|A|$, is a measure of its size. If A is finite, the cardinality is simply the number of elements in A. If A is finite and $|A| = k$, then the number of subsets of A is

$$|P(A)| = 2^k.$$

If we have also $|B| = m$, then

$$|A \times B| = km.$$

Applied to the product of a set with itself,

$$|A^n| = k^n.$$

For infinite sets the measurement of the size, or cardinality, is more complicated. The notation ∞ denotes infinity, but it does not distinguish among infinities. And it turns out that there are meaningful ways in which infinities may differ.

How do we compare the sizes of infinite sets? The basic idea is this. Suppose we are given two sets A and B, and a one-to-one function $f : A \to B$. Then we want to say that B is at least as large as A, that is,

$|B| \geq |A|$.

If the converse also holds, that is, there also exists a one-to-one function $g : B \to A$, then we also have $|A| \geq |B|$, and together these imply that A and B have the same *cardinality*, $|A| = |B|$. (In this case it is also true that there is a one-to-one and onto function from A to B.) Otherwise, we say that the cardinality of B is larger than that of A, $|B| > |A|$.

For example, if

$A = \{1, 2, \ldots\}$,

$B = \{2, 3, \ldots\}$,

we find that the function $f : A \to B$ defined by $f(n) = n + 1$ is one-to-one and onto between A and B. Thus, the two sets are just as large. There is something counterintuitive here. A contains all of B plus one element, 1. So it feels like A should be strictly larger than B. But there is no interesting definition of the size of a set that distinguishes between A and B. The reason is that the bijection f suggests that we think of B as identical to A, with a renaming of the elements. With a bijection between two sets, it's hopeless to try to assign them different sizes.

By the same logic, the intervals

$[0, 1] = \{x \in \mathbb{R} \,|\, 0 \leq x \leq 1\}$

and

$[0, 2] = \{x \in \mathbb{R} \,|\, 0 \leq x \leq 2\}$

are of the same cardinality because the function $f(x) = 2x$ is a bijection from the first to the second. This is even more puzzling because these intervals have lengths, and the length of $[0, 2]$ is twice as large as that of $[0, 1]$. Indeed, there are other concepts of size in mathematics that would be able to capture that fact. But cardinality, attempting to count numbers, doesn't.

The cardinality of $(-1, 1)$ is identical to that of the entire real line, $\mathbb{R}$, even though the length of the former is finite and of the latter infinite. (Use the functions tag/arctag to switch between the two sets.)

Combining these arguments, we see that $\mathbb{R}$ has the same cardinality as $[0, 1]$, or $[0, 0.1]$, or $[0, \varepsilon]$ for any $\varepsilon > 0$.

Continuing with the list of counterintuitive comparisons, we find that the naturals $\mathbb{N} = \{1, 2, 3, \ldots\}$ and the integers $\mathbb{Z} = \{\ldots, -1, 0, 1, \ldots\}$ are of the same cardinality even though the integers include all the naturals, their negatives, and zero. Clearly, we can have a one-to-one function from $\mathbb{N}$ to $\mathbb{Z}$: the identity $(f(n) = n)$. But we can also map $\mathbb{Z}$ to $\mathbb{N}$ in a one-to-one way. For instance, consider the following enumeration of $\mathbb{Z}$:

$$\mathbb{Z} = \{0, 1, -1, 2, -2, 3, -3, \ldots\},$$

that is,

$$\begin{aligned} 0 &\mapsto 1 \\ 1 &\mapsto 2 \\ -1 &\mapsto 3 \\ &\vdots \\ k &\mapsto 2k \\ -k &\mapsto 2k+1 \end{aligned}$$

This function from $\mathbb{Z}$ to $\mathbb{N}$ is one-to-one (and we also made it onto).

Similarly, the set of rational numbers

$$\mathbb{Q} = \left\{ \frac{a}{b} \,\middle|\, a \in Z, b \in \mathbb{N} \right\}$$

is of the same cardinality as $\mathbb{N}$. As previously, it is easy to map $\mathbb{N}$ into $\mathbb{Q}$ in a one-to-one way because $\mathbb{N} \subset \mathbb{Q}$. But the converse is also true. We may list all the rational numbers in a sequence $q_1, q_2, \ldots$ such that any rational will appear in a certain spot in the sequence, and no two rational numbers will claim the same spot. For instance, consider the table

	0	1	−1	2	−2	...
1	q_1	q_2	q_4	q_7	...	
2	q_3	q_5	q_8	...		
3	q_6	q_9	...			
4	q_{10}	...				
...	...					

Note that different representations of the same rational number are counted several times. For instance, $q_1 = q_3 = \cdots = 0$. Hence, define the function from $\mathbb{Q}$ to $\mathbb{N}$ as follows: for $q \in \mathbb{Q}$, let $f(q)$ be the minimal

n such that $q_n = 9$, where q_n is defined by the table. Clearly, every q appears somewhere in the list $q_1, q_2, \ldots$; hence this function is well defined. It is one-to-one because each q_n can equal only one number in $\mathbb{Q}$.

It seems at this point that all infinite sets are, after all, of the same size. But this is not the case. We concluded that the sets

$$\mathbb{N}, \mathbb{Z}, \mathbb{Q}$$

are of the same cardinality, and so are

$$\mathbb{R}, [0,1], [0,\varepsilon]$$

for any $\varepsilon > 0$. But the cardinality of the first triple is lower than the cardinality of the second.

Clearly, the cardinality of $\mathbb{N}$ cannot exceed that of $\mathbb{R}$, because $\mathbb{N} \subset \mathbb{R}$, and thus the identity function maps $\mathbb{N}$ into $\mathbb{R}$ in a one-to-one manner. The question is, can we have the opposite direction, namely, can we map $\mathbb{R}$ into $\mathbb{N}$ in a one-to-one way, or equivalently, can we count the elements in $\mathbb{R}$? The answer is negative. There are at least three insightful proofs of this fact (not provided here). It suffices to know that there are sets that are not countable, and any interval with a positive length is such a set. Thus, in a well-defined sense, there are as many rational numbers as there are natural numbers, and there are as many numbers in any interval as there are in the entire real line (and, in fact, in any $\mathbb{R}^n$), but any interval (with a positive length) has more points than the natural (or the rational) numbers.

A.5 Calculus

A.5.1 Limits of Sequences

The notion of a limit is intuitive and fundamental. What is the limit of $\frac{1}{n}$ as $n \to \infty$? It is zero. We write this as

$$\lim_{n\to\infty} \frac{1}{n} = 0$$

or

$$\frac{1}{n} \to_{n\to\infty} 0.$$

Formally, we say that a sequence of real numbers $\{a_n\}$ converges to a number b, denoted

$$a_n \to_{n\to\infty} b$$

or

$$\lim_{n\to\infty} a_n = b$$

if the following holds: for every $\varepsilon > 0$ there exists N such that

$$n \geq N$$

implies

$$|a_n - b| < \varepsilon.$$

Intuitively, $\{a_n\}$ converges to b if it gets closer and closer to b. How close? As close as we wish. We decide how close to b we want the sequence to be, and we can then find a place in the sequence, N, such that all numbers in the sequence from that place on are as close to b as we requested.

If the sequence converges to ∞ (or $-\infty$), we use a similar definition, but we have to redefine the notion of "close to." Being close to ∞ doesn't mean having a difference of no more than ε, but rather, being large. Formally, $a_n \to_{n\to\infty} \infty$ if, for every M, there exists N such that

$$n \geq N \Rightarrow a_n > M,$$

and a similar definition is used for convergence to $-\infty$.

A.5.2 Limits of Functions

Again, we intuitively understand what is the limit of a function at a point. For instance, if x is a real-valued variable ($x \in \mathbb{R}$), we can agree that

$$\lim_{x\to\infty} \frac{1}{x} = 0$$

and

$$\lim_{x\to 0} \frac{1}{x^2} = \infty.$$

The formal definition of a limit is the following. The statement

$$\lim_{x\to a} f(x) = b$$

or

$f(x) \to_{x \to a} b$

means that for every $\varepsilon > 0$, there exists a $\delta > 0$ such that

$|x - a| < \delta$

implies

$|f(x) - b| < \varepsilon.$

That is, if we know that we want the value of the function to be close to (within ε of) the limit b, we just have to be close enough to (within δ of) the argument a.

The proximity of the argument is defined a little differently when we approach infinity. Being close to ∞ doesn't mean being within δ of it, but being above some value. Explicitly, the statement

$\lim_{x \to \infty} f(x) = b$

or

$f(x) \to_{x \to \infty} b$

means that for every $\varepsilon > 0$, there exists M such that

$x > M$

implies

$|f(x) - b| < \varepsilon.$

Similarly, if we wish to say that the function converges to ∞ as x converges to a, we say that for every M, there exists a $\delta > 0$ such that

$|x - a| < \delta$

implies

$f(x) > M.$

Similar definitions apply to $\lim_{x \to \infty} f(x) = \infty$ and to the case in which x or $f(x)$ is $-\infty$.

A.5.3 Continuity

A function $f : \mathbb{R} \to \mathbb{R}$ is continuous at a point a if it equals its own limit, that is, if

$$\lim_{x \to a} f(x) = f(a).$$

The same definition applies to multiple variables. If we have $f : \mathbb{R}^n \to \mathbb{R}$, we say that f is continuous at x if $f(x) \to f(a)$ whenever $x \to a$. Specifically, f is continuous at $a \in \mathbb{R}^n$ if for every $\varepsilon > 0$, there exists $\delta > 0$ such that $\|x - a\| < \delta$ implies $|f(x) - f(a)| < \varepsilon$.

A.5.4 Derivatives

The *derivative* of a real-valued function of a single variable, $f : \mathbb{R} \to \mathbb{R}$, at a point a, is defined as

$$f'(a) = \frac{df}{dx}(a) = \lim_{x \to a} \frac{f(x) - f(a)}{x - a}.$$

If we draw the graph of the function and let $x \neq a$ be close to a, $\frac{f(x)-f(a)}{x-a}$ is the slope of the string connecting the point on the graph corresponding to a, $(a, f(a))$ and the point corresponding to x, $(x, f(x))$. The derivative of f at the point a is the limit of this slope. Thus, it is the slope of the graph at the point a, or the slope of the tangent to the function.

When we say that a function has a derivative at a point a, we mean that this limit exists. It may not exist if, for instance, the function has a kink at a (for instance, $f(x) = |x - a|$), or if the function is too wild to have a limit even when x approaches a from one side.

The geometric interpretation of the derivative f' is therefore the slope of the function, or its rate of increase, that is, the ratio between the increase (positive or negative) in the value of the function relative to a small change in the variable x. If x measures time, and $f(x)$ measures the distance from a given point, $f'(x)$ is the velocity. If x measures the quantity of a good, and $u(x)$ measures the utility function, then $u'(x)$ measures the marginal utility of the good.

A function that always has a derivative is called *differentiable*. At every point a, we can approximate it by the linear function that is its tangent,

$$g(x) = f(a) + (x - a)f'(a),$$

and for values of x close to a, this approximation will be reasonable. Specifically, by definition of the derivative, the difference between the approximation, $g(x)$, and the function, $f(x)$, will converge to zero faster than x converges to a:

$$\frac{g(x) - f(x)}{x - a} = \frac{f(a) - f(x) - (x - a)f'(a)}{x - a}$$
$$= \frac{f(a) - f(x)}{x - a} - f'(a),$$

where the definition of the derivative means that the latter converges to zero as $x \to a$.

Thus, the zero-order approximation to the function f around a is the constant $f(a)$. The first-order approximation is the linear function $f(a) + (x - a)f'(a)$. Using higher-order derivatives (derivatives of derivatives of... the derivative), one can get higher-order approximations of f by higher-order polynomials in x.

A.5.5 Partial Derivatives

When we have a function of several variables,

$$f : \mathbb{R}^n \to \mathbb{R}$$

we can consider the rate of the change in the function relative to each of the variables. If we wish to see what is the impact (on f) of changing, say, only the first variable x_1, we can fix the values of the other variables $\bar{x}_2, \ldots, \bar{x}_n$ and define

$$f_{\bar{x}_2,\ldots,\bar{x}_n}(x_1) = f(x_1, \bar{x}_2, \ldots, \bar{x}_n).$$

Focusing on the impact of x_1, we can study the derivative of $f_{\bar{x}_2,\ldots,\bar{x}_n}$. Since the other variables are fixed, we call this a *partial derivative*, denoted

$$\frac{\partial f}{\partial x_1}(x_1, \bar{x}_2, \ldots, \bar{x}_n) = \frac{d f_{\bar{x}_2,\ldots,\bar{x}_n}}{dx_1}(x_1).$$

A function $f : \mathbb{R}^n \to \mathbb{R}$ is called differentiable if it can be approximated by a linear function. Specifically, at a point a, define

$$g(x) = f(a) + \sum_{i=1}^{n}(x_i - a_i)\frac{\partial f}{\partial x_1}(x)$$

and require that

$$\frac{|g(x) - f(x)|}{\|x - a\|}$$

converge to 0, as $\|x - a\|$ does.

A.6 Topology

Topology is the study of the abstract notion of convergence. We only need the standard topology here, and the definitions of convergence are given with respect to this topology, as are the definitions that follow. However, it is worthwhile to recall that there can be other topologies and, correspondingly, other notions of convergence.

A set $A \subset \mathbb{R}^n$ is *open* if for every $x \in A$, there exists $\varepsilon > 0$ such that

$$\|x - y\| < \varepsilon \Rightarrow y \in A.$$

That is, around every point in the set A we can draw a small ball, perhaps very small but with a positive radius ε (the more general concept is an open neighborhood) such that the ball will be fully contained in A.

The set

$$(0, 1) = \{x \in \mathbb{R} \,|\, 0 < x < 1\}$$

is open (the open interval). Similarly, for $n = 2$, the following sets are open:

$$\{(x, y) \in \mathbb{R}^2 \,|\, x^2 + y^2 < 1\}$$

$$\{(x, y) \in \mathbb{R}^2 \,|\, 3x + 4y < 17\}$$

$$\mathbb{R}^2$$

A set $A \subset \mathbb{R}^n$ is *closed* if for every convergent sequence of points in it, $(x_1, x_2, \ldots)$ with $x_n \in A$ and $x_n \to_{n\to\infty} x^*$, the limit point is also in the set, that is, $x^* \in A$.

The set $[0, 1] = \{x \in \mathbb{R} \,|\, 0 \le x \le 1\}$ is closed in $\mathbb{R}$. The following subsets of $\mathbb{R}^2$ are closed (in $\mathbb{R}^2$):

$$\{(x, y) \in \mathbb{R}^2 \,|\, x^2 + y^2 \le 1\}$$

$$\{(x, y) \in \mathbb{R}^2 \,|\, 3x + 4y \le 17\}$$

$$\mathbb{R}^2$$

The set

$$[0, 1) = \{x \in \mathbb{R} \,|\, 0 \le x < 1\}$$

is neither open nor closed. It is not open because $0 \in [0, 1)$, but no open neighborhood of 0 is (fully) contained in A. It is not closed because the sequence $x_n = 1 - 1/n$ is a convergent sequence of points in A, whose limit (1) is not in A.

In $\mathbb{R}^n$, the only two sets that are both open and closed are the entire space ($\mathbb{R}^n$ itself) and the empty set. This is true in any space that we call connected.

A.7 Probability

A.7.1 Basic Concepts

Intuitively, an event is a fact that may or may not happen, a proposition that may be true or false. The probability model has a set of states of the world, or possible scenarios, often denoted by Ω or by S. Each state $s \in S$ is assumed to describe all the relevant uncertainty. An *event* is then defined as a subset of states, that is, as a subset $A \subset S$. When S is infinite, we may not wish to discuss all subsets of S. But when S is finite, there is no loss of generality in assuming that every subset is an event that can be referred to.

The set-theoretic operations of complement, union, and intersection correspond to the logical operations of negation, disjunction, and conjunction. For example, if we roll a die and

$$S = \{1, \ldots, 6\},$$

we can think of the events

$$A = \text{“The die comes up on an even number”} = \{2, 4, 6\}$$

$$B = \text{“The die comes up on a number smaller than 4”} = \{1, 2, 3\}$$

and then $A^c = \{1, 3, 5\}$ designates the event “the die comes up on an odd number,” that is, the negation of the proposition that defines A, and $B^c = \{4, 5, 6\}$ is the event described by “the die comes up on a number that is not smaller than 4.” Similarly, $A \cup B = \{1, 2, 3, 4, 6\}$ stands for “the die comes up on a number that is smaller than 4, or even, or both,” and $A \cap B = \{2\}$ is defined by “the die comes up on a number that is both even and smaller than 4.”

Probability is an assignment of numbers to events, which is supposed to measure their plausibility. The formal definition is simpler when S is finite, and we can refer to all subsets of S. That is, the set of events is

$$2^S = \{A \mid A \subset S\}.$$

A probability is a function

$$P : 2^S \to \mathbb{R}$$

that satisfies three properties:

1. $P(A) \geq 0$ for every $A \subset S$;
2. Whenever $A, B \subset S$ are disjoint (i.e., $A \cap B = \varnothing$),

$$P(A \cup B) = P(A) + P(B);$$

3. $P(S) = 1$.

The logic behind these conditions is derived from two analogies. First, we can think of a probability of an event as its relative frequency. Relative frequencies are non-negative (property 1), and they are added up when we discuss two disjoint events (property 2). The relative frequency of S, the event that always occurs, is 1 (property 3).

The second analogy, which is particularly useful when an event is not repeated in the same way and relative frequencies cannot be defined, is the general notion of a measure. When we measure the mass of objects or the length of line segments or the volume of bodies, we use numerical functions on subsets (of matter, of space) that satisfy the first two properties. For example, the masses of objects are never negative, and they add up when we take together two objects that had nothing in common. The last property is a matter of normalization, or a choice of the unit of measurement so that the sure event will always have the probability 1.

It is easy to verify that a function P satisfies the *additivity* condition (property 2) if and only if it satisfies, for every $A, B \subset S$,

$$P(A \cup B) = P(A) + P(B) - P(A \cap B).$$

These three properties imply that $P(\varnothing) = 0$, so that the impossible event has probability 0.

When S is finite, say, $S = \{1, \ldots, n\}$, we say that $p = (p_1, \ldots, p_n)$ is a probability vector on S if

$$p_i \geq 0, \quad \forall i \leq n,$$

and

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

For every probability $P : 2^S \to [0,1]$, there exists a probability vector p such that

$$P(A) = \sum_{i \in A} p_i, \quad \forall A \subset S,$$

and vice versa, every probability vector p defines a probability P by this equation. Thus, the probabilities on all events are in a one-to-one correspondence with the probability vectors on S.

A.7.2 Random Variables

Consider a probability model with a state space S and a probability on it P, or equivalently, a probability vector p on S. In this model a random variable is defined to be a function on S. For example, if X is a random variable that assumes real numbers as values, we can write it as

$$X : S \to \mathbb{R}.$$

The point of this definition is that a state $s \in S$ contains enough information to know anything of importance. If the focus is on a variable X, each state should specify the value that X assumes. Thus, $X(s)$ is a well-defined value, about which there is no uncertainty. Any previous uncertainty is incorporated into the uncertainty about which state s obtains. But given such a state s, no uncertainty remains.

Observe that we can use a random variable X to define events. For instance, "X equals a" is the name of the event

$$\{s \in S \mid X(s) = a\},$$

and "X is no more than a" is

$$\{s \in S \mid X(s) \leq a\},$$

and so forth.

Often, we are interested only in the probability that a random variable will assume certain values, not at which states it does so. If X takes values in some set $\mathcal{X}$, we can then define the *distribution* of a random variable X, as a function $f_X : \mathcal{X} \to [0,1]$ by

$$f_X(x) = P(X = x) = P(\{s \in S \mid X(s) = x\}).$$

For real-valued random variables, there are several additional useful definitions. The cumulative distribution of X, $F_X : \mathbb{R} \to [0,1]$ is

$$F_X(x) = P(X \leq x) = P(\{s \in S \mid X(s) \leq x\}).$$

It is thus a nondecreasing function of x going from 0 (when x is below the minimal value of X) to 1 (when x is greater than or equal to the maximal value of X). This definition can also be used when the state space is infinite and X may assume infinitely many real values.

Trying to summarize the information about a random variable X, there are several central measures. The most widely used is the *expectation*, or the *mean*, which is simply a weighted average of all values of X, where the probabilities serve as weights:

$$EX = \sum_x f_X(x)x$$

and (in a finite state space with generic element i),

$$EX = \sum_{i=1}^{n} p_i X(i).$$

The most common measure of dispersion around the mean is the *variance*, defined by

$$\text{var}(X) = E[(X - EX)^2].$$

It can be verified that

$$\text{var}(X) = E[X^2] - [EX]^2.$$

Since the variance is defined as the expectation of squared deviations from the expectation, its unit of measurement is not intuitive (it is the square of the unit of measurement of X). Therefore, we can often use the *standard deviation*, defined by

$$\sigma_X = \sqrt{\text{var}(X)}.$$

Expectation behaves in a linear way. If X, Y are real-valued random variables, and $\alpha, \beta \in \mathbb{R}$, then

$$E[\alpha X + \beta Y] = \alpha EX + \beta EY.$$

For the variance of sums (or of linear functions in general), we need to take into account the relation between X and Y. The *covariance* of X and Y is defined as

$$\text{cov}(X, Y) = E[(X - EX)(Y - EY)].$$

Intuitively, the covariance tries to measure whether X and Y go up and down together, or whether they tend to go up and down in different directions. If they do go up and down together, whenever X is relatively high (above its mean, EX), Y will be relatively high (above its mean, EY), and $(X - EX)(Y - EY)$ will be positive. And whenever X is below its mean, Y will be also be below its mean, resulting in a positive product $(X - EX)(Y - EY)$. By contrast, if Y tends to be relatively high (above EY) when X is relatively low (below EX), and vice versa, there will be more negative values of $(X - EX)(Y - EY)$. The covariance is an attempt to summarize the values of this variable. If $\mathrm{cov}(X, Y) > 0$, then X and Y are *positively correlated*; if $\mathrm{cov}(X, Y) < 0$, X and Y are *negatively correlated*; and if $\mathrm{cov}(X, Y) = 0$, X and Y are *uncorrelated.*

Equipped with the covariance, we can provide a formula for the variance of a linear combination of random variables:

$$\mathrm{var}[\alpha X + \beta Y] = \alpha^2 \,\mathrm{var}(X) + 2\alpha\beta \,\mathrm{cov}(X, Y) + \beta^2 \,\mathrm{var}(Y).$$

The formulas for expectation and variance also extend to more than two random variables:

$$E\left(\sum_{i=1}^{n} \alpha_i X_i\right) = \sum_{i=1}^{n} \alpha_i E X_i$$

and

$$\mathrm{var}\left(\sum_{i=1}^{n} \alpha_i X_i\right) = \sum_{i=1}^{n} \alpha_i^2 \,\mathrm{var}(X_i) + 2\sum_{i=1}^{n}\sum_{j\neq i} \alpha_i\alpha_j \,\mathrm{cov}(X_i, X_i).$$

A.7.3 Conditional Probabilities

The unconditional probability of an event A, $P(A)$, is a measure of the plausibility of A occurring a priori, when nothing is known. The conditional probability of A given B, $P(A|B)$, is a measure of the likelihood of A occurring once we know that B has already occurred.

Bayes suggested that this conditional probability be the ratio of the probability of the intersection of the two events to the probability of the event that is known to have occurred. That is, he defined the conditional probability of A given B to be

$$P(A|B) = \frac{P(A \cap B)}{P(B)}.$$

(This definition only applies if $P(B) > 0$.)

The logic of this definition is as follows. Assume that event B has occurred. What do we think about A? For A to occur now, the two events, A and B, have to occur simultaneously. That is, we need their intersection, $A \cap B$, to occur. The probability of this happening was estimated (a priori) to be the numerator $P(A \cap B)$. However, if we just take this expression, probabilities will not sum up to 1. Indeed, the sure event will not have a probability higher than $P(B)$. We have a convention that probabilities sum up to 1. It is a convenient normalization because when we say that an event has probability of, say, .45, we don't have to ask .45 out of how much. We know that the total has been normalized to 1. To stick to this convention, we divide the measure of likelihood of A in the presence of B, $P(A \cap B)$, by the maximal value of this expression (over all A's), which is $P(B)$, and this results in Bayes' formula.

Observe that this formula makes sense in extreme cases. If A is implied by B, that is, if B is a subset of A (whenever B occurs, so does A), then $A \cap B = B$, and we have $P(A \cap B) = P(B)$ and $P(A|B) = 1$; that is, given that B has occurred, A is a certain event. At the other extreme, if A and B are logically incompatible, then their intersection is the empty set, $A \cap B = \varnothing$ and there is no scenario in which both materialize. Then $P(A \cap B) = P(\varnothing) = 0$ and $P(A|B) = 0$; that is, if A and B are incompatible, then the conditional probability of A given B is zero.

If two events are independent, the occurrence of one says nothing about the occurrence of the other. In this case the conditional probability of A given B should be the same as the unconditional probability of A. Indeed, one definition of independence is

$$P(A \cap B) = P(A)P(B),$$

which implies

$$P(A|B) = \frac{P(A \cap B)}{P(B)} = \frac{P(A)P(B)}{P(B)} = P(A).$$

Rearranging the terms in the definition of conditional probability, for any two events A and B (independent or not),

$$\begin{aligned} P(A \cap B) &= P(B)P(A|B) \\ &= P(A)P(B|A), \end{aligned}$$

that is, the probability of the intersection of two events (the probability of both occurring) can be computed by taking the unconditional probability of one of them and multiplying it by the conditional probability of the second given the first. Clearly, if the events are independent, and

$$P(A|B) = P(A),$$

$$P(B|A) = P(B),$$

the two equations boil down to

$$P(A \cap B) = P(A)P(B).$$

Note that the formula $P(A \cap B) = P(B)P(A|B)$ applies also if independence does not hold. For example, the probability that a candidate wins a presidency twice in a row is the probability that she wins the first time, multiplied by the conditional probability that she wins the second time given that she has already won the first time.

Let there be two events A and B such that $P(B), P(B^c) > 0$. We note that

$$A = (A \cap B) \cup (A \cap B^c)$$

and

$$(A \cap B) \cap (A \cap B^c) = \varnothing.$$

Hence

$$P(A) = P(A \cap B) + P(A \cap B^c)$$

and, combining the equalities,

$$P(A) = P(A|B)P(B) + P(A|B^c)P(B^c).$$

Thus, the overall probability of A can be computed as a weighted average, with weights $P(B)$ and $P(B^c) = 1 - P(B)$, of the conditional probability of A given B and the conditional probability of A given B^c.

A.7.4 Independence and i.i.d. Random Variables

Using the concept of independent events, we can also define independence of random variables. Let us start with two random variables X, Y that are defined on the same probability space. For simplicity of notation, assume that they are real-valued:

$$X, Y : S \to \mathbb{R}.$$

Then, given a probability P on S, we can define the joint distribution of X and Y to be the function $f_{X,Y} : \mathbb{R}^2 \to [0,1]$ defined by

$$\begin{aligned} f_{X,Y}(x,y) &= P(X = x, Y = y) \\ &= P(\{s \in S \mid X(s) = x, Y(s) = y\}). \end{aligned}$$

We say that X and Y are *independent random variables* if, for every x, y,

$$f_{X,Y}(x,y) = f_X(x) f_Y(y).$$

In other words, every event that is defined in terms of X has to be independent of any event that is defined in terms of Y. Intuitively, anything we know about X does not change our belief about (the conditional distribution of) Y.

If the state space is not finite, similar definitions apply to cumulative distributions. We can then define independence by the condition

$$\begin{aligned} F_{X,Y}(x,y) &= P(X \le x, Y \le y) \\ &= P(X \le x) P(Y \le y) \\ &= F_X(x) F_Y(y). \end{aligned}$$

All these definitions extend to any finite number of random variables. Thus, if $X_1, \ldots, X_n$ are random variables, their joint distribution and their joint cumulative distributions are, respectively,

$$f_{X_1,\ldots,X_n} : \mathbb{R}^n \to [0,1]$$

and

$$F_{X_1,\ldots,X_n} : \mathbb{R}^n \to [0,1]$$

defined by

$$f_{X_1,\ldots,X_n}(x_1, \ldots, x_n) = P(X_1 = x_1, \ldots, X_n = x_n)$$

and

$$F_{X_1,\ldots,X_n}(x_1, \ldots, x_n) = P(X_1 \le x_1, \ldots, X_n \le x_n).$$

Independence of n random variables is similarly defined, by the product rule

$$F_{X_1,\ldots,X_n}(x_1, \ldots, x_n) = \prod_{i=1}^{n} P(X_i \le x_i),$$

and it means that nothing that may be learned about any subset of the variables will change the conditional distribution of the remaining ones. If n random variables are independent, then so are any pair of them. The converse, however, is not true. There may be random variables that are pairwise independent but that are not independent as a set. For example, consider $n = 3$, and let X_1 and X_2 have the joint distribution

	0	1
0	0.25	0.25
1	0.25	0.25

with $X_3 = 1$ if $X_1 = X_2$, and $X_3 = 0$ if $X_1 \neq X_2$. Any pair of (X_1, X_2, X_3) are independent, but together the three random variables are not independent. In fact, any two of them fully determine the third.

Two random variables X and Y are *identically distributed* if they have the same distribution, that is, if

$$f_X(a) = f_Y(a)$$

for any value a. Two random variables X and Y are *identical* if they always assume the same value. That is, if, for every state $s \in S$,

$$X(s) = Y(s).$$

Clearly, if X and Y are identical, they are also identically distributed. This is so because, for every a,

$$f_X(a) = P(X = a) = P(Y = a) = f_Y(a),$$

where $P(X = a) = P(Y = a)$ follows from the fact that $X = a$ and $Y = a$ define precisely the same event. That is, since $X(s) = Y(s)$, any $s \in S$ belongs to the event $X = a$ if and only if it belongs to the event $Y = a$.

By contrast, two random variables that are not identical can still be identically distributed. For example, if X can assume the values $\{0, 1\}$ with equal probabilities, and $Y = 1 - X$ (that is, for every s, $Y(s) = 1 - X(s)$), then X and Y are identically distributed, but they are not identical. In fact, they never assume the same value.

The notion of identical distribution is similarly defined for more than two variables. That is, $X_1, \ldots, X_n$ are identically distributed if, for every a,

$$f_{X_1}(a) = P(X_1 = a) = \cdots = P(X_n = a) = f_{X_n}(a).$$

The variables $X_1, \ldots, X_n$ are said to be i.i.d. (identically and independently distributed) if they are identically distributed and independent.

A.7.5 Law(s) of Large Numbers

Consider a sequence of i.i.d. random variables $X_1, \ldots, X_n, \ldots$. Since they all have the same distribution, they all have the same expectation,

$$EX_i = \mu,$$

and the same variance. Assume that this variance is finite

$$\mathrm{var}(X_i) = \sigma^2.$$

When two random variables are independent, they are also uncorrelated (that is, their covariance is zero). Hence the variance of their sum is the sum of their variances.

When we consider the average of the first n random variables,

$$\bar{X}_n = \frac{1}{n}\sum_{i=1}^{n} X_i,$$

we observe that

$$E(\bar{X}_n) = \frac{1}{n}\sum_{i=1}^{n} EX_i = \mu,$$

and since any two of them are uncorrelated,

$$\mathrm{var}(\bar{X}_n) = \frac{1}{n^2}\sum_{i=1}^{n} \mathrm{var}(X_i) = \frac{\sigma^2}{n},$$

which implies that the more variables we take in the average, the lower will be the variance of the average. This, in turn, means that the average, $\bar{X}_n$, will be, with very high probability, close to its expectation, which is μ.

In fact, more can be said. We may decide how close we want $\bar{X}_n$ to be to μ, and with what probability, and then we can find a large enough N such that, for all n starting from N, $\bar{X}_n$ will be as close to μ as we wish with the probability we specify. Formally, for every $\varepsilon > 0$ and every $\delta > 0$, there exists N such that

$$P(\{s \mid |\bar{X}_n - \mu| < \delta \ \forall n \geq N\}) > 1 - \varepsilon.$$

It is also the case that the probability of the event that $\overline{X}_n$ converges to μ is 1:

$$P\left(\left\{s \,\middle|\, \exists \lim_{n\to\infty} \overline{X}_n = \mu\right\}\right) = 1.$$

LLN and Relative Frequencies Suppose a certain trial or experiment is repeated infinitely many times. In each repetition, the event A may or may not occur. The different repetitions/trials/experiments are assumed to be identical in terms of the probability of A occurring in each, and independent. Then we can associate, with experiment i, a random variable

$$X_i = \begin{cases} 1 & A \text{ occurred in experiment } i \\ 0 & A \text{ did not occur in experiment } i \end{cases}$$

The random variables $(X_i)_i$ are independently and identically distributed (i.i.d.) with $E(X_i) = p$, where p is the probability of A occurring in each of the experiments. The relative frequency of A in the first n experiments is the average of these random variables,

$$\overline{X}_n = \frac{1}{n}\sum_{i=1}^{n} X_i = \frac{\#\{i|A \text{ occurred in experiment } i\}}{n}.$$

Hence, the law of large numbers guarantees that the relative frequency of A will converge to its probability p.

B Formal Models

B.1 Utility Maximization

B.1.1 Definitions

Suppose there is a set of alternatives X. A binary relation $\succsim$ on X is simply a set of ordered pairs of elements from X, that is, $\succsim \subset X \times X$, with the interpretation that for any two alternatives $x, y \in X$,

$$(x, y) \in \succsim,$$

also denoted

$$x \succsim y,$$

means "alternative x is at least as good as alternative y in the eyes of the decision maker" or "given the choice between x and y, the decision maker may choose x."

It is useful to define two binary relations associated with $\succsim$, which are often called the symmetric and the asymmetric parts of $\succsim$. Specifically, let us introduce the following definitions. First, we define the inverse of the relation $\succsim$:

$$\precsim = \succsim^{-1} = \{(y, x) \mid (x, y) \in \succsim\},$$

that is, $y \precsim x$ if and only if $x \succsim y$ (for any x, y). The symbol $\precsim$ was selected to make $y \precsim x$ and $x \succsim y$ similar, but it is a new symbol and requires a new definition.

Do not confound the relation $\succsim$ between alternatives and the relation $\geq$ between their utility values. Later, when we have a representation of $\succsim$ by a utility function, we will be able to do precisely that—to think of

$$x \succsim y$$

as equivalent to

$$u(x) \geq u(y),$$

but this equivalence is the representation we seek, and until we prove that such a function u exists, we should be careful not to confuse $\succsim$ with $\geq$.[1]

Next define the symmetric part of $\succsim$ to be the relation $\sim \subset X \times X$ defined by

$$\sim = \succsim \cap \precsim,$$

that is, for every two alternatives $x, y \in X$,

$$x \sim y \Leftrightarrow [(x \succsim y) \quad \text{and} \quad (y \succsim x)].$$

Intuitively, $x \succsim y$ means "alternative x is at least as good as alternative y in the eyes of the decision maker," and $x \sim y$ means "the decision maker finds alternatives x and y equivalent" or "the decision maker is indifferent between alternatives x and y."

The asymmetric part of $\succsim$ is the relation $\succ \subset X \times X$ defined by

$$\succ = \succsim \setminus \precsim,$$

that is, for every two alternatives $x, y \in X$,

$$x \succ y \Leftrightarrow [(x \succsim y) \quad \text{and not} \quad (y \succsim x)].$$

Intuitively, $x \succ y$ means "the decision maker finds alternative x strictly better than alternative y."

B.1.2 Axioms

The main axioms that we impose on $\succsim$ are as follows.

Completeness For every $x, y \in X$, $x \succsim y$ or $y \succsim x$.

(Recall that *or* in mathematical language is inclusive unless otherwise stated. That is, "A or B" should be read as "A or B or possibly both").

The completeness axiom states that the decision maker can make up her mind between any two alternatives. This means that at each and every possible instance of choice between x and y something will be chosen. But it also means, implicitly, that we expect some regularity in

1. Observe that the sequence of symbols $x \geq y$ need not make sense at all because the elements of X need not be numbers or vectors or any other mathematical entities.

these choices: x is always chosen (and then we would say that $x \succ y$), or y is always chosen ($y \succ x$), or sometimes x is chosen and sometimes y. But this latter case would be modeled as equivalence ($x \sim y$), and the implicit assumption is that the choice between x and y would be completely random. If, for instance, the decision maker chooses x on even dates and y on odd dates, it would seem inappropriate to say that she is indifferent between the two options. In fact, we may find that the language is too restricted to represent the decision maker's preferences. The decision maker may seek variety and always choose the option that has not been chosen on the previous day. In this case, one would like to say that preferences are history- or context-dependent and that it is, in fact, a modeling error to consider preferences over x and y themselves (rather than, say, on sequences of x's and y's). More generally, when we accept the completeness axiom we do not assume only that at each given instance of choice one of the alternatives will end up being chosen. We also assume that it is meaningful to define preferences over the alternatives, and that these alternatives are informative enough to tell us anything that might be relevant for the decision under discussion.

Transitivity For every $x, y, z \in X$, if ($x \succsim y$ and $y \succsim z$), then $y \succsim z$.

Transitivity has a rather obvious meaning, and it almost seems like part of the definition of preferences. Yet, it is easy to imagine cyclical preferences. Moreover, such preferences may well occur in group decision making, for instance, if the group is using a majority vote. This is the famous Condorcet paradox (see section 6.2 of the main text). Assume that there are three alternatives, $X = \{x, y, z\}$ and that one-third of society prefers

$x \succ y \succ z$,

one-third

$y \succ z \succ x$,

and the last third

$z \succ x \succ y$.

It is easy to see that when every two pairs of alternatives come up for a separate majority vote, there is a two-thirds majority for $x \succ y$, a two-thirds majority for $y \succ z$, but also a two-thirds majority for $z \succ x$.

In other words, a majority vote may violate transitivity and even generate a cycle of strict preferences: $x \succ y \succ z \succ x$.

Once we realize that this can happen in a majority vote in a society, we can imagine how this can happen inside the mind of a single individual as well. Suppose Daniel has to choose among three cars, and he ranks them according to three criteria, such as comfort, speed, and price. He finds it hard to quantify and trade off these criteria, so he decides to adopt the simple rule that if one alternative is better than another according to most criteria, then it should be preferred. In this case Daniel can be thought of as if he were the aggregation of three decision makers—one who cares only about comfort, one who cares only about speed, and one who cares only about price—where his decision rule as a "society" is to follow a majority vote. Then Daniel would find that his preferences are not transitive. But if this happens, we expect him to be confused about the choice and to dislike the situation of indecision. Thus, even if the transitivity axiom does not always hold, it is generally accepted as a desirable goal.

B.1.3 Result

We are interested in a representation of a binary relation by a numerical function. Let us first define this concept more precisely.

A function $u : X \to \mathbb{R}$ is said to *represent* $\succsim$ if, for every $x, y \in X$,

$$x \succsim y \Leftrightarrow u(x) \geq u(y). \tag{B.1}$$

Proposition 1 Let X be finite. Let $\succsim$ be a binary relation on X, i.e., $\succsim \subset X \times X$. The following are equivalent: (i) $\succsim$ is complete and transitive; (ii) there exists a function $u : X \to \mathbb{R}$ that represents $\succsim$.

B.1.4 Generalization to a Continuous Space

In many situations of interest, the set of alternatives is not finite. If we consider a consumer who has preferences over the amount of wine she consumes, the amount of time she spends in the pool, or the amount of money left in her bank account, we are dealing with variables that are continuous and that therefore may assume infinitely many values. Thus, the set X, which may be a set of vectors of such variables, is infinite.

Physicists might say that the amount of wine can only take finitely many values because there are a finite number of particles in a glass of wine (and perhaps also in the world). This is certainly true of the amount of money—it is only measured up to cents. And the accuracy of measurement is also limited in the case of time, temperature, and so forth. So maybe the world is finite after all, and we don't need to deal with extension of proposition 1?

The fact is that finite models may be very awkward and inconvenient. For example, assume there are supply and demand curves that slope in the right directions[2] but fail to intersect because they are only defined for finitely many prices. (In fact, they are not really curves but only finite collections of points in $\mathbb{R}^2$.) It would be silly to conclude that the market will never be at equilibrium simply because there is no precise price at which supply and demand are equal. You might recall a similar discussion in statistics. The very first time you were introduced to continuous random variables, you might have wondered who really needs them in a finite world. But then you find out that many assumptions and conclusions are greatly simplified by the assumption of continuity.

In short, we would like to have a similar theorem, guaranteeing utility representation of a binary relation, also in the case that the set of alternatives is infinite. There are several ways to obtain such a theorem. The one presented here also guarantees that the utility function be continuous. To make this a meaningful statement, we have to have a notion of convergence in the set X, a topology. But in order to avoid complications, let us simply assume that X is a subset of $\mathbb{R}^n$ for some $n \geq 1$ and think of convergence as it is usually defined in $\mathbb{R}^n$.

It is not always the case that a complete and transitive relation on $\mathbb{R}^n$ can be represented by a numerical function. (A famous counterexample was provided by Gerard Debreu.[3]) An additional condition that we may impose is that the relation $\succsim$ be continuous. What is meant by this is that if $x \succ y$, then all the points that are very close to x are also strictly better than y, and vice versa, all the points that are very close to y are also strictly worse than x.

Continuity For every $x \in X$, the sets $\{y \in X \mid x \succ y\}$ and $\{y \in X \mid y \succ x\}$ are open in X.

(Recall that a set is open if, for every point in it, there is a whole neighborhood contained in the set.)

---

2. The supply curve, which indicates the quantity supplied as a function of price, is increasing. The demand curve, which specifies the quantity demanded as a function of price, is decreasing.

3. G. Debreu, *The Theory of Value: An Axiomatic Analysis of Economic Equilibrium* (New Haven: Yale University Press, 1959), ch. 2, prob. 6.

To see why this axiom captures the notion of continuity, we may think of a function $f : \mathbb{R} \to \mathbb{R}$ and a point $x \in \mathbb{R}$ for which $f(x) > 0$. If f is continuous, then there is a neighborhood of x for which f is positive. If we replace "positive" by "strictly better than y" for a fixed y, we see the similarity between these two notions of continuity.

Alternatively, we can think of continuity as requiring that for every $x \in X$, the sets

$$\{y \in X \,|\, x \succsim y\}$$

and

$$\{y \in X \,|\, y \succsim x\}$$

be closed in X. That is, if we consider a convergent sequence $(y_n)_{n \geq 1}$, $y_n \to y$, such that $y_n \succsim x$ for all n, then also $y \succsim x$, and if $x \succsim y_n$ for all n, then also $x \succsim y$. In other words, if we have a weak preference all along the sequence (either from below or from above), we should have the same weak preference at the limit. This condition is what we were after.

Theorem 2 (Debreu) Let $\succsim$ be a binary relation on X, that is, $\succsim \subset X \times X$. The following are equivalent: (i) $\succsim$ is complete, transitive, and continuous; (ii) there exists a continuous function $u : X \to \mathbb{R}$ that represents $\succsim$.

B.2 Convexity

As a preparation for the discussion of constrained optimization, it is useful to have some definitions of convex sets, convex and concave functions, and so on.

B.2.1 Convex Sets

A set $A \subset \mathbb{R}^n$ is *convex* if, for every $x, y \in A$ and every $\lambda \in [0, 1]$, $\lambda x + (1 - \lambda)y \in A$. That is, whenever two points are in the set, the line segment connecting them is also in the set. If we imagine the set A as a room, convexity means that any two people in the room can see each other.

B.2.2 Convex and Concave Functions

A function $f : \mathbb{R}^n \to \mathbb{R}$ is convex if its graph is never above the strings that connect points on it. As an example, we may think of

$$f(x) = x^2$$

for $n = 1$. If we draw the graph of this function and take any two points on the graph, when we connect them by a segment (the string), the graph of the function will be below it (or at least not above the segment). The same will be true for

$$f(x_1, x_2) = x_1^2 + x_2^2$$

if $n = 2$.

Formally, $f : \mathbb{R}^n \to \mathbb{R}$ is *convex* if for every $x, y \in A$ and every $\lambda \in [0, 1]$,

$$f(\lambda x + (1 - \lambda)y) \leq \lambda f(x) + (1 - \lambda)f(y),$$

and it is strictly convex if this inequality is strict whenever $x \neq y$ and $0 < \lambda < 1$.

To see the geometric interpretation of this condition, imagine that $n = 1$, and observe that $\lambda x + (1 - \lambda)y$ is a point on the interval connecting x and y. Similarly, $\lambda f(x) + (1 - \lambda)f(y)$ is a point on the interval connecting $f(x)$ and $f(y)$. Moreover, if we connect the two points

$$(x, f(x)), (y, f(y)) \in \mathbb{R}^2$$

by a segment (which is a string of the function f), we get precisely the points

$$\{(\lambda x + (1 - \lambda)y, \lambda f(x) + (1 - \lambda)f(y)) \mid \lambda \in [0, 1]\}.$$

For $\lambda = 1$ the point is $(x, f(x))$; for $\lambda = 0$ it is $(y, f(y))$; for $\lambda = 0.5$, the point has a first coordinate that is the arithmetic average of x and y, and a second coordinate that is the average of their f values. Generally, for every λ, the first coordinate is the $(\lambda, (1 - \lambda))$ average between x and y, and the second coordinate is corresponding average of their f values.

Convexity of the function demands that for every $\lambda \in [0, 1]$, the value of the function at the $(\lambda, (1 - \lambda))$ average between x and y, that is, $f(\lambda x + (1 - \lambda)y)$, will not exceed the height of the string (connecting $(x, f(x))$ and $(y, f(y))$) at the same point.

Next assume that $n = 2$ and repeat the argument to show that this geometric interpretation is valid in general.

A function $f : \mathbb{R}^n \to \mathbb{R}$ is convex if and only if the following set is convex[4]

$$\{(x, z) \in \mathbb{R}^{n+1} \mid z \geq f(x)\}.$$

If $n = 1$ and f is twice differentiable, convexity of f is equivalent to the condition that $f'' \geq 0$, that is, that the first derivative, f', is nondecreasing. When $n > 1$, there are similar conditions, expressed in terms of the matrix of second derivatives, which are equivalent to convexity.

Concave functions are defined in the same way, with the converse inequality. All that is true of convex functions is true of concave functions, with the opposite inequality. In fact, we could define f to be concave if $-f$ is convex. But we will spell it out.

A function $f : \mathbb{R}^n \to \mathbb{R}$ is *concave* if for every $x, y \in A$ and every $\lambda \in [0, 1]$,

$$f(\lambda x + (1 - \lambda)y) \geq \lambda f(x) + (1 - \lambda)f(y),$$

and it is strictly concave if this inequality is strict whenever $x \neq y$ and $0 < \lambda < 1$.

Thus, f is concave if the graph of the function is never below the strings that connect points on it. Equivalently, $f : \mathbb{R}^n \to \mathbb{R}$ is concave if and only if the following set is convex

$$\{(x, z) \in \mathbb{R}^{n+1} \mid z \leq f(x)\}.$$

(This set is still required to be convex, not concave. In fact, we didn't define the notion of a concave set, and we don't have such a useful definition. The difference between this condition for convex and concave functions is in the direction of the inequality. The resulting set in both cases is required to be a convex set as a subset of $\mathbb{R}^{n+1}$.)

If $n = 1$ and f is twice differentiable, concavity of f is equivalent to the condition that $f'' \leq 0$, that is, that the first derivative, f', is nonincreasing.

An affine function is a shifted linear function. That is, $f : \mathbb{R}^n \to \mathbb{R}$ is *affine* if

4. Observe that the vector (x, z) refers to the concatenation of x, which is a vector of n real numbers, with z, which is another real number—together a vector of $(n + 1)$ real numbers.

$$f(x) = \sum_{i=1}^{n} a_i x_i + c,$$

where $\{a_i\}$ and c are real numbers.

An affine function is both convex and concave (but not strictly so). The converse is also true: a function that is both convex and concave is affine.

If we take a convex function f, we can, at each x, look at the tangent to the graph of f. This would be a line if $n = 1$ and a hyperplane more generally. Formally, for every x there exists an affine function $l_x : \mathbb{R}^n \to \mathbb{R}$ such that

$$l_x(x) = f(x),$$

and for every $y \in \mathbb{R}^n$

$$l_x(y) \leq f(y).$$

If we take all these functions $\{l_x\}_x$, we find that their maximum is f. That is, for every $y \in \mathbb{R}^n$

$$f(y) = \max_x \; l_x(y).$$

Thus, a convex function can be described as the maximum of a collection of affine functions. Conversely, the maximum of affine functions is always convex. Hence, a function is convex if and only if it is the maximum of affine functions.

Similarly, a function is concave if and only if it is the minimum of a collection of affine functions.

B.2.3 Quasi-convex and Quasi-concave Functions

Consider the convex function

$$f(x_1, x_2) = x_1^2 + x_2^2.$$

Suppose that I cut it at a given height, z, and ask which points (x_1, x_2) do not exceed z in their f value. That is, I look at

$$\{x \in \mathbb{R}^2 \mid f(x) \leq z\}.$$

It is easy to see that this set will be convex. This gives rise to the following definition.

A function $f : \mathbb{R}^n \to \mathbb{R}$ is *quasi-convex* if, for every $z \in \mathbb{R}$,

$$\{x \in \mathbb{R}^n \mid f(x) \leq z\}$$

is a convex set.

Observe that this set is a subset of $\mathbb{R}^n$ and that we have a (potentially) different such set for every value of z, whereas in the characterization of convex functions given previously we used the convexity of a single set in $\mathbb{R}^{n+1}$.

The term *quasi* should suggest that every convex function is also quasi-convex. Indeed, if

$$y, w \in \{x \in \mathbb{R}^n \mid f(x) \leq z\},$$

then

$$f(y), f(w) \leq z,$$

and for every $\lambda \in [0,1]$,

$$\begin{aligned} f(\lambda y + (1-\lambda)w) &\leq \lambda f(y) + (1-\lambda) f(w) \\ &\leq \lambda z + (1-\lambda) z = z, \end{aligned}$$

and this means that

$$\lambda y + (1-\lambda)w \in \{x \in \mathbb{R}^n \mid f(x) \leq z\}.$$

Since this is true for every y, w in $\{x \in \mathbb{R}^n \mid f(x) \leq z\}$, this set is convex for every z, and this is the definition of quasi-convexity of the function f.

Is every quasi-convex function is convex? The answer is negative, (otherwise we wouldn't use a different term for quasi-convexity). Indeed, it suffices to consider $n = 1$ and observe that

$$f(x) = x^3$$

is quasi-convex but not convex. Indeed, when we look at the sets

$$\{x \in \mathbb{R}^n \mid f(x) \leq z\}$$

for various values of $z \in \mathbb{R}$, we simply get the convex sets $(-\infty, \alpha]$ for some α (in fact, for $\alpha = z^{1/3}$). The collection of these sets, when we range over all possible values of z, does not look any different for the original function, x^3, than it would if we looked at the function x or $x^{1/3}$.

Again, everything we can say of quasi-convex functions has a counterpart for quasi-concave ones. A function $f : \mathbb{R}^n \to \mathbb{R}$ is *quasi-concave* if, for every $z \in \mathbb{R}$,

$$\{x \in \mathbb{R}^n \mid f(x) \geq z\}$$

is a convex set.

Imagine now the parabola upside down,

$$f(x_1, x_2) = -x_1^2 - x_2^2,$$

and when we cut it at a certain height, z, and look at the dome above the cut, the projection of this dome on the x_1, x_2 plane is a circle. The fact that it is a convex set follows from the fact that f is quasi-concave.

B.3 Constrained Optimization

B.3.1 Convex Problems

Constrained optimization problems are much easier to deal with when they are convex. Roughly, we want everything to be convex, both on the feasibility and on the desirability side.

Convexity of the feasible set is simple to define. We require that the set F be convex.

What is meant by "convex preferences"? The answer is that we wish the "at least as desirable as" sets to be convex. Explicitly, for every $x \in \mathbb{R}^n$, we may consider the "at least as good as" set

$$\{y \in X \mid y \succsim x\}$$

and require that it be convex. If we have a utility function u that represents $\succsim$, we require that the function be quasi-concave. Indeed, if u is quasi-concave, then for every $\alpha \in \mathbb{R}$, the set

$$\{y \in X \mid u(y) \geq \alpha\}$$

is convex. When we range over all values of α, we obtain all sets of the form $\{y \in X \mid y \succsim x\}$, and thus a quasi-concave u defines convex preferences. Observe that quasi-concavity is the appropriate term when the utility is given only up to an arbitrary (increasing) monotone transformation (utility is only ordinal). Whereas a concave function can be replaced by a monotone transformation that results in a nonconcave function, a quasi-concave function will remain quasi-concave after any increasing transformation.

Convex problems (in which both the feasible set and preferences are convex) have several nice properties. In particular, local optima are also global optima. This means that looking at first-order conditions is often sufficient. If these conditions identify a local maximum, we can rest assured that it is also a global one. Another important feature of convex problems is that for such problems one can devise simple algorithms of small local improvements that converge to the global optimum. This is very useful if we are trying to solve the problem on a computer. But, more important, it also says that real people may behave as if they were solving such problems optimally. If a decision maker makes small improvements when these exist, we may assume that, as time goes by, he converges to the optimal solution. Thus, for large and complex problems, the assumption that people maximize utility subject to their feasible set is much more plausible in convex problems than it is in general.

B.3.2 Example: The Consumer Problem

Let us look at the consumer problem again. The decision variables are $x_1, \ldots, x_n \in \mathbb{R}_+$, where x_i ($x_i \geq 0$) is the amount consumed of good i. The consumer has an income $I \geq 0$, and she faces prices $p_1, \ldots, p_n \in \mathbb{R}_{++}$ (that is, $p_i > 0$ for all $i \leq n$). The problem is therefore

$$\max_{x_1,\ldots,x_n} u(x_1, \ldots, x_n)$$

subject to

$$p_1 x_1 + \cdots + p_n x_n \leq I$$

$$x_i \geq 0$$

B.3.3 Algebraic Approach

Let us further assume that u is strictly monotone in all its arguments, namely, that the consumer prefers more of each good to less. Moreover, we want to assume that u is quasi-concave, so that the "better than" sets are convex. Under these assumptions we may conclude that the optimal solution will be on the budget constraint, namely, will satisfy

$$p_1 x_1 + \cdots + p_n x_n = I,$$

and if a point $x = (x_1, \ldots, x_n)$ is a local maximum, it is also a global one. Hence it makes sense to seek a local maximum, namely, to ask whether a certain point on the budget constraint happens to maximize utility in a certain neighborhood (of itself on this constraint).

If the utility function is also differentiable, we may use calculus to help identify the optimal solution. Specifically, the first-order condition for this problem can be obtained by differentiating the Lagrangian

$$L(x_1, \ldots, x_n, \lambda) = u(x_1, \ldots, x_n) - \lambda[p_1x_1 + \cdots + p_nx_n - I]$$

and equating all first (partial) derivatives to zero. This yields

$$\frac{\partial L}{\partial x_i} = \frac{\partial u}{\partial x_i} - \lambda p_i = 0$$

for all $i \leq n$ and

$$\frac{\partial L}{d\lambda} = -[p_1x_1 + \cdots + p_nx_n - I] = 0.$$

The second equality is simply the budget constraint, whereas the first implies that for all i,

$$\frac{u_i}{p_i} = \text{const} = \lambda,$$

where $u_i = \frac{\partial u}{dx_i}$. Thus, for any two goods i, j, we have

$$\frac{u_i}{p_i} = \frac{u_j}{p_j} \tag{B.2}$$

or

$$\frac{u_i}{u_j} = \frac{p_i}{p_j}. \tag{B.3}$$

B.3.4 Geometric Approach

Each of these equivalent conditions has an intuitive interpretation. Let us start with the second, which can be understood geometrically. We argue that it means that the feasible set and the desirable ("better than") set are tangent to each other. To see this, assume there are only two goods, $i = 1, 2$. Consider the budget constraint

$$p_1x_1 + p_2x_2 = I,$$

and observe that its slope, at a point x, can be computed by taking differentials:

$$p_1\,dx_1 + p_2\,dx_2 = 0,$$

which means

$$\frac{dx_1}{dx_2} = -\frac{p_1}{p_2}. \tag{B.4}$$

Next consider the "better than" set, and focus on the tangent to this set at the point x. We get a line (more generally, a hyperplane) that goes through the point x, and satisfies

$$du = u_1\,dx_1 + u_2\,dx_2 = 0,$$

that is, a line with a slope

$$\frac{dx_1}{dx_2} = -\frac{u_1}{u_2}. \tag{B.5}$$

This will also be the slope of the indifference curve (the set of points that are indifferent to x) at x. Clearly, condition (B.3) means that the slope of the budget constraint, (B.4), equals the slope of the indifference curve (B.5).

Why is this a condition for optimality? We may draw several indifference curves and superimpose them on the budget constraint. In general, we can always take this geometric approach to optimization: draw the feasible set, and then compare it to the "better than" sets. If an indifference curve, which is the boundary of a "better than" set, does not intersect the feasible set, it indicates a level of utility that cannot be reached. It is in the category of wishful thinking. A rational decision maker will be expected to give up this level of utility and settle for a lower one.

If, on the other hand, the curve cuts *through* the feasible set, the corresponding level of utility is reachable, but it is not the highest such level. Since the curve is strictly in the interior of the feasible set, there are feasible points on either side of it. Assuming that preferences are monotone, that is, that the decision maker prefers more to less, one side of the curve has a higher utility level than the curve itself. Since it is feasible, the curve we started with cannot be optimal. Here a rational decision maker will be expected to strive for more and look for a higher utility level.

What is the highest utility level that is still feasible? It has to be represented by an indifference curve that is not disjoint with the feasible set yet does not cut through it. In other words, the intersection of the feasible set and the "better than" set is nonempty but has an empty interior (it has zero volume, or zero area in a two-dimensional problem). If both sets are smooth, they have to be tangent to each other. This tangency condition is precisely what equation (B.3) yields.

B.3.5 Economic Approach

The economic approach is explained in the main text. But it may be worthwhile to repeat it in a slightly more rigorous way. Consider condition (B.2). Again, assume that I already decided on spending most of my budget, and I'm looking at the last dollar, asking whether I should spend it on good i or on good j. If I spend it on i, how much of this good will I get? At price p_i, one dollar would buy $\frac{1}{p_i}$ units of the good. How much additional utility will I get from this quantity? Assuming that one dollar is relatively small and that correspondingly the amount of good i, $\frac{1}{p_i}$, is also relatively small, I can approximate the marginal utility of $\frac{1}{p_i}$ extra units by

$$\frac{1}{p_i} \cdot \frac{\partial u}{\partial x_i} = \frac{u_i}{p_i}.$$

Obviously, the same reasoning would apply to good j. Spending the dollar on j would result in an increase in utility that is approximately $\frac{u_j}{p_j}$. Now, if

$$\frac{u_i}{p_i} > \frac{u_j}{p_j},$$

one extra dollar spent on i will yield a higher marginal utility than the same dollar spent on j. Put differently, we can take one dollar of the amount spent on j and transfer it to the amount spent on i, and be better off, since the utility lost on j, $\frac{u_j}{p_j}$, is more than compensated for by the utility gained on i, $\frac{u_i}{p_i}$.

This argument assumes that we can indeed transfer one dollar from good j to good i. That is, that we are at an interior point. If we consider a boundary point, where we don't spend any money on j in any case, this inequality may be consistent with optimality.

If one dollar is not relatively small, we can repeat this argument with ε dollars, where ε is small enough for the derivatives to provide good approximations. Then we find that ε dollars are translated to quantities $\frac{\varepsilon}{p_i}$ and $\frac{\varepsilon}{p_j}$, if spent on goods i or j, respectively, and that these quantities yield marginal utilities of $\frac{\varepsilon u_i}{p_i}$ and $\frac{\varepsilon u_j}{p_j}$, respectively. Hence, any of the inequalities

$$\frac{u_i}{p_i} > \frac{u_j}{p_j} \quad \text{or} \quad \frac{u_i}{p_i} < \frac{u_j}{p_j}$$

indicates that we are not at an optimal (interior) point. Condition (B.2) is a powerful tool in identifying optimal points. It says that small changes in the budget allocation, in other words, small changes along the boundary of the budget constraint, will not yield an improvement.

B.3.6 Comments

Two important comments are in order. First, the previous arguments are not restricted to a feasible set defined by a simple budget constraint, that is, by a linear inequality. The feasible set may be defined by one or many nonlinear constraints. What is crucial is that it be convex.

Second, condition (B.2) is necessary only if the optimal solution is at a point where the sets involved—the feasible set and the "better than" set—are smooth enough to have a unique tangent (supporting hyperplane, that is, a hyperplane defined by a linear equation that goes through the point in question, and such that the entire set is on one of its sides). An optimal solution may exist at a point where one of the sets has kinks, and in this case slopes and derivatives may not be well defined.

Still, the first-order conditions, namely, the equality of slopes (or ratios of derivatives) are sufficient for optimality in convex problems, that is, problems in which both the feasible set and the "better than" sets are convex. It is therefore a useful technique for finding optimal solutions in many problems. Moreover, it provides us with very powerful insights. In particular, the marginal way of thinking about alternatives, which we saw in the economic interpretation, appears in many problems within and outside of economics.

B.4 vNM's Theorem

B.4.1 Setup

vNM's original formulation involved decision trees in which compound lotteries were explicitly modeled. We use here a more compact formulation, due to Niels-Erik Jensen and Peter Fishburn,[5] which

5. N. E. Jensen, "An Introduction to Bernoullian Utility Theory," pts. I and II, *Swedish Journal of Economics* 69 (1967): 163–183, 229–247; P. C. Fishburn, *Utility Theory for Decision Making* (New York: Wiley, 1970).

implicitly assumes that compound lotteries are simplified according to Bayes' formula. Thus, lotteries are defined by their distributions, and the notion of mixture implicitly supposes that the decision maker is quite sophisticated in terms of his probability calculations.

Let X be a set of alternatives. X need not consist of sums of money or consumption bundles, and it may include outcomes such as death.

The objects of choice are lotteries. We can think of a lottery as a function from the set of outcomes, X, to probabilities. That is, if P is a lottery and x is an outcome, $P(x)$ is the probability of getting x if we choose lottery P. It will be convenient to think of X as potentially infinite, as is the real line, for example. At the same time, we don't need to consider lotteries that may assume infinitely many values. We therefore assume that while X is potentially infinite, each particular lottery P can only assume finitely many values.

The set of all lotteries is therefore

$$L = \left\{ P : X \to [0,1] \;\middle|\; \begin{array}{c} \#\{x|P(x) > 0\} < \infty, \\ \sum_{x \in X} P(x) = 1 \end{array} \right\}.$$

Observe that the expression $\sum_{x \in X} P(x) = 1$ is well defined thanks to the finite support condition that precedes it.

A *mixing operation* is performed on L, defined for every $P, Q \in L$ and every $\alpha \in [0,1]$ as follows: $\alpha P + (1-\alpha)Q \in L$ is given by

$$(\alpha P + (1-\alpha)Q)(x) = \alpha P(x) + (1-\alpha)Q(x)$$

for every $x \in X$. The intuition behind this operation is of conditional probabilities. Assume that I offer you a compound lottery that will give you the lottery P with probability α and the lottery Q with probability $(1-\alpha)$. You can ask what is the probability of obtaining a certain outcome x, and observe that it is indeed α times the conditional probability of x if you get P plus $(1-\alpha)$ times the conditional probability of x if you get Q.

Since the objects of choice are lotteries, the observable choices are modeled by a binary relation on L, $\succsim \subset L \times L$.

B.4.2 The vNM Axioms

The vNM axioms are

Weak order $\succsim$ is complete and transitive.

Continuity For every $P, Q, R \in L$, if $P \succ Q \succ R$, there exist $\alpha, \beta \in (0,1)$ such that $\alpha P + (1-\alpha)R \succ Q \succ \beta P + (1-\beta)R$.

Independence For every $P, Q, R \in L$, and every $\alpha \in (0,1)$, $P \succsim Q$ if and only if $\alpha P + (1-\alpha)R \succsim \alpha Q + (1-\alpha)R$.

The weak order axiom is not very different from the same assumption in chapter 2 of the main text. The other two axioms are new and deserve a short discussion.

B.4.3 Continuity

Continuity may be viewed as a technical condition needed for the mathematical representation and for the proof to work. To understand its meaning, consider the following example, supposedly challenging continuity. Assume that P guarantees one dollar, Q guarantees zero dollars, and R guarantees death. You are likely to prefer one dollar to no dollars, and no dollars to death. That is, you would probably exhibit preferences $P \succ Q \succ R$. The axiom then demands that for a high enough $\alpha < 1$, you will also exhibit the preference

$$\alpha P + (1-\alpha)R \succ Q,$$

namely, that you will be willing to risk your life with probability $(1-\alpha)$ in order to gain one dollar. The point of the example is that you are supposed to say that no matter how small the probability of death $(1-\alpha)$, you will not risk your life for one dollar.

A counterargument to this example (suggested by Howard Raiffa) is that we often do indeed take such risks. For instance, suppose you are about to buy a newspaper, which costs one dollar. But you see that it is freely distributed on the other side of the street. Would you cross the street to get it at no cost? If you answer in the affirmative, you are willing to accept a certain risk, albeit very small, of losing your life (in traffic) in order to save one dollar.

This counterargument can be challenged in several ways. For instance, you may argue that even if you don't cross the street, your life is not guaranteed with probability 1. Indeed, a truck driver who falls asleep may hit you anyway. In this case, we are not comparing death with probability 0 to death with probability $(1-\alpha)$. And, the argument goes, it is possible that if you had true certainty on your side of the street, you would not have crossed the street, thereby violating the axiom.

It appears that framing also matters in this example. I may be about to cross the street in order to get the free copy of the newspaper, but if you stop me and say, "What are you doing? Are you nuts, to risk your life this way? Think of what could happen! Think of your family!" I might cave in and give up the free paper. It is not obvious which behavior is more relevant, namely, the decision making without the guilt-inducing speech or with it. Presumably, this depends on the application.

In any event, we understand the continuity axiom. Moreover, if we consider applications that do not involve extreme risks such as death, it appears to be a reasonable assumption.

B.4.4 Independence

The independence axiom is related to dynamic consistency. However, it involves several steps. Consider the following four choice situations:

1. You are asked to make a choice between P and Q.

2. Nature will first decide whether, with probability $(1-\alpha)$, you get R, and then you have no choice to make. Alternatively, with probability α, nature will let you choose between P and Q.

3. The choices are as in (2), but you have to commit to making your choice before you observe Nature's move.

4. You have to choose between two branches. In one, Nature will first decide whether, with probability $(1-\alpha)$, you get R, or, with probability α, you get P. The second branch is identical, with Q replacing P.

Clearly, (4) is the choice between $\alpha P+(1-\alpha)R$ and $\alpha Q+(1-\alpha)R$. To relate the choice in (1) to that in (4), we can use (2) and (3) as intermediary steps. Compare (1) and (2). In (2), *if* you are called upon to act, you are choosing between P and Q. At that point R will be a counterfactual world. Why would it be relevant? Hence, it is argued, you can ignore the possibility that did not happen, R, and make your decision in (2) identical to that in (1).

The distinction between (2) and (3) has to do only with the timing of your decision. Should you make different choices in these scenarios, you would not be dynamically consistent. It is as if you plan in (3) to make a given choice, but when you get the chance to make it, you do (or would like to do) something else in (2). Observe that when you make a choice in (3), you know that this choice is conditional on getting to the decision node. Hence, the additional information you have should not change this conditional choice.

Finally, the alleged equivalence between (3) and (4) relies on changing the order of your move (to which you already committed) and

Nature's move. As such, this is an axiom of reduction of compound lotteries, assuming that the order of the draws does not matter as long as the distributions on outcomes, induced by your choices, are the same.

B.4.5 The Theorem

Finally, the theorem can be stated.

Theorem 3 (vNM) Let there be given a relation $\succsim \subset L \times L$. The following are equivalent: (i) $\succsim$ satisfies weak order, continuity, and independence; (ii) there exists $u : X \to \mathbb{R}$ such that, for every $P, Q \in L$,

$$P \succsim Q \quad \text{iff} \quad \sum_{x \in X} P(x)u(x) \geq \sum_{x \in X} Q(x)u(x).$$

Moreover, in this case u is unique up to a positive linear transformation (plt). That is, $v : X \to \mathbb{R}$ satisfies, for every $P, Q \in L$,

$$P \succsim Q \quad \text{iff} \quad \sum_{x \in X} P(x)v(x) \geq \sum_{x \in X} Q(x)v(x)$$

if and only if there are $a > 0$ and $b \in \mathbb{R}$ such that $v(x) = au(x) + b$ for every $x \in X$.

Thus, we find that the theory of expected utility maximization is not just one arbitrary generalization of expected value maximization. There are quite compelling reasons to maximize expected utility (in a normative application) as well as to believe that this is what people naturally tend to do (in a descriptive application). If we put aside the more technical condition of continuity, we find that expected utility maximization is equivalent to following a weak order that is linear in probabilities; this linearity is basically what the independence axiom says.

B.5 Ignoring Base Probabilities

The disease example discussed in section 5.4.1 of the main text illustrates that people often mistake $P(A|B)$ for $P(B|A)$. In that example, if you had the disease, the test would show it with probability 90 percent; if you didn't, the test might still show a false positive with probability 5 percent. Suppose you took the test and you tested positive. What was the probability of your actually having the disease?

Let D be the event of having the disease and T be the event of testing positive. Then

$$P(T|D) = .90,$$

$$P(T|D^c) = .05.$$

What is $P(D|T)$?

The definition of conditional probability says that

$$P(D|T) = \frac{P(D \cap T)}{P(T)}.$$

Trying to get closer to the given data, we may split the event T into two disjoint events:

$$T = (D \cap T) \cup (D^c \cap T).$$

In other words, one may test positive if one is sick $(D \cap T)$ but also if one is healthy $(D^c \cap T)$, so

$$P(T) = P(D \cap T) + P(D^c \cap T)$$

and

$$P(D|T) = \frac{P(D \cap T)}{P(T)}$$

$$= \frac{P(D \cap T)}{P(D \cap T) + P(D^c \cap T)}.$$

Now we can try to relate each of the probabilities in the denominator to the conditional probabilities we are given. Specifically,

$$P(D \cap T) = P(D)P(T|D) = .90P(D)$$

and

$$P(D^c \cap T) = P(D^c)P(T|D^c) = .05[1 - P(D)]$$

(recalling that the probability of no disease, $P(D^c)$, and the probability of disease have to sum up to 1.) Putting it all together, we get

$$P(D|T) = \frac{P(D \cap T)}{P(T)}$$

$$= \frac{P(D \cap T)}{P(D \cap T) + P(D^c \cap T)}$$

$$= \frac{P(D)P(T|D)}{P(D)P(T|D) + P(D^c)P(T|D^c)}$$

$$= \frac{.90P(D)}{.90P(D) + .05[1 - P(D)]}.$$

This number can be anywhere in $[0, 1]$. Indeed, suppose that we are dealing with a disease that is known to be extinct. Thus, $P(D) = 0$. The accuracy of the test remains the same: $P(T|D) = .90$, and $P(T|D^c) = .05$, but we have other reasons to believe that the a priori probability of having the disease is zero. Hence, whatever the test shows, your posterior probability is still zero. If you test positive, you should attribute it to the inaccuracy of the test (the term $.05[1 - P(D)]$ in the denominator) rather than to having the disease (the term $.90P(D)$). By contrast, if you are in a hospital ward consisting only of previously diagnosed patients, and your prior probability of having the disease is $P(D) = 1$, your posterior probability will be 1 as well (and this will be the case even if you tested negative).

To see why Kahneman and Tversky called this phenomenon "ignoring base probabilities," observe that what relates the conditional probability of A given B to the conditional probability of B given A is the ratio of the unconditional (base) probabilities:

$$P(A|B) = \frac{P(A)}{P(B)} P(B|A),$$

and the confusion of $P(B|A)$ for $P(A|B)$ is tantamount to ignoring the term $P(A)/P(B)$.

B.6 Arrow's Impossibility Theorem

Let $N = \{1, 2, \ldots, n\}$ be the set of individuals, and let X be the set of alternatives. Assume that X is finite, with $|X| \geq 3$. Each individual is assumed to have a preference relation over X. For simplicity, assume that there are no indifferences, so that for each $i \in N$, there is a relation $\succsim_i \subset X \times X$ that is complete, transitive, and antisymmetric (namely, $x \succsim_i y$ and $y \succsim_i x$ imply $x = y$.) Alternatively, we may assume that for each individual $i \in N$ there is a "strictly prefer" relation $\succ_i \subset X \times X$ that is transitive and that satisfies

$$x \neq y \Leftrightarrow [x \succ_i y \text{ or } y \succ_i x].$$

(If $\succsim_i$ is complete, transitive, and antisymmetric, its asymmetric part $\succ_i$ satisfies this condition.)

The list of preference relations $(\succsim_1, \ldots, \succsim_n) = (\succsim_i)_i$ is called a *profile.* It indicates how everyone in society ranks the alternatives. Arrow's theorem does not apply to one particular profile but to a function that is assumed to define a social preference for *any* possible profile of individual preferences. Formally, let

$$R = \{\succsim \subset X \times X \,|\, \succsim \text{ is complete, transitive, antisymmetric}\}$$

be the set of all possible preference relations. We consider functions that take profiles, or n-tuples of elements in R into R itself. That is, the theorem will be about creatures of the type

$$f : R^n \to R.$$

Note that all profiles, that is, all n-tuples of relations (one for each individual), are considered. This can be viewed as an implicit assumption that is sometimes referred to explicitly as "full domain."

For such functions f we are interested in two axioms:

Unanimity For all $x, y \in X$, if $x \succsim_i y \; \forall i \in N$, then $xf((\succsim_i)_i)y$.

The unanimity axiom says that if everyone prefers x to y, then so should society.

Independence of Irrelevant Alternatives (IIA) For all $x, y \in X$, $(\succsim_i)_i$, $(\succsim_i')_i$, if $x \succsim_i y \Leftrightarrow x \succsim_i' y$, $\forall i \in N$, then $xf((\succsim_i)_i)y \Leftrightarrow sf((\succsim_i')_i)y$.

The IIA axiom says that the social preference between two specific alternatives, x and y, only depends on individual preferences between these two alternatives. That is, suppose we compare two different profiles, $(\succsim_i)_i$, $(\succsim_i')_i$, and find that they are vastly different in many ways, but it so happens that when we restrict attention to the pair $\{x, y\}$, the two profiles look the same: for each and every individual, x is considered to be better than y according to $\succsim_i$ if and only if it is better than y according to $\succsim_i'$. The axiom requires that when we aggregate preferences according to the function f, and consider the aggregation of $(\succsim_i)_i$, that is $f((\succsim_i)_i)$, and the aggregation of $(\succsim_i')_i$, which is denoted $f((\succsim_i')_i)$, we find that these two aggregated relations rank x and y in the same way.

The final definition we need is the following:

A function f is *dictatorial* if there exists $j \in N$ such that for every $(\succsim_i)_i$ and every $x, y \in X$,

$$xf((\succsim_i)_i)y \Leftrightarrow x \succsim_j y.$$

That is, f is dictatorial if there exists one individual, j, such that, whatever the others think, society simply adopts j's preferences. We can finally state

Theorem 4 (Arrow) f satisfies unanimity and IIA iff it is dictatorial.

Arrow's theorem can be generalized to the case in which the preference relations admit indifferences (that is, are not necessarily antisymmetric). In this case, the unanimity axiom has to be strengthened to apply both to weak and to strict preferences.[6]

B.7 Nash Equilibrium

A *game* is a triple $(N, (S_i)_{i \in N}, (h_i)_i)$, where $N = \{1, \ldots, n\}$ is a set of *players*, S_i is the (nonempty) set of *strategies* of player i, and

$$h_i : S \equiv \prod_{i \in N} S_i \to \mathbb{R}$$

is player i's vNM *utility function.*

A selection of strategies $s = (s_1, \ldots, s_n) \in S$ is a *Nash equilibrium* (in pure strategies) if for every $i \in N$,

$$h(s) \geq h(s_{-i}, t_i), \qquad \forall t_i \in S_i,$$

where $(s_{-i}, t_i) \in S$ is the n-tuple of strategies obtained by replacing s_i by t_i in s. In other words, a selection of strategies is a Nash equilibrium if, given what the others are choosing, each player is choosing a best response.

To model random choice, we extend the strategy set of each player to mixed strategies, that is, to the set of distributions over the set of pure strategies:

$$\Sigma_i = \left\{ \sigma_i : S_i \to [0, 1] \,\middle|\, \sum_{s_i \in S_i} \sigma_i(s_i) = 1 \right\}.$$

6. Other formulations of Arrow's involve a choice function, selecting a single alternative $x \in X$ for a profile $(\succsim_i)_i$. In these formulations the IIA axiom is replaced by a monotonicity axiom stating that if x is chosen for a given profile $(\succsim_i)_i$, x will also be chosen in any profile where x is only "better," in terms of pairwise comparisons with all the others. This axiom is similar in its strengths and weaknesses to the IIA in that it requires that direct pairwise comparisons, not concatenations thereof, would hold sufficient information to determine social preferences.

Given a mixed strategy $\sigma_i \in \Sigma_i$ for each $i \in N$, we define i's payoff to be the expected utility

$$H_i(\sigma_1, \ldots \sigma_n) = \sum_{s \in S} \left[\prod_{j \in N} \sigma_j(s_j) \right] h_i(s),$$

and we define a Nash equilibrium in mixed strategies to be a Nash equilibrium of the extended game in which the sets of strategies are $(\Sigma_i)_i$ and the payoff functions—$(H_i)_i$.

Mixed strategies always admit Nash equilibria.

Theorem 5 (Nash) Let $(N, (S_i)_{i \in N}, (h_i)_i)$ be a game in which S_i is finite for each i.[7] Then it has a Nash equilibrium in mixed strategies.

7. Recall that in the formulation here N was also assumed finite.

C Exercises

Chapter 1 Feasibility and Desirability

1. In the first example we saw an instance of

impossible ⇒ *undesirable*,

whereas the second was an instance of

possible ⇒ *undesirable*.

The third example is one in which

desirable ⇒ *possible*,

and this raises the question, what would be an example in which

desirable ⇒ *impossible*?

2. Symmetry requires that we also look for examples in which

possible ⇒ *desirable*,

impossible ⇒ *desirable*,

undesirable ⇒ *possible*,

undesirable ⇒ *impossible*.

Can you find such examples?

3. George says, "I wish to live in a peaceful world. Therefore, I favor policies that promote world peace."

a. Explain why this statement violates the separation of feasibility and desirability.

b. Suppose George thinks that if a peaceful world is impossible, he is not interested in living any more, and further, he doesn't care about anything else that might happen in this world, to himself or to others. Explain why, under these assumptions, George's statement is compatible with rationality.

4. In the previous exercises, the symbol $\Rightarrow$ referred to causal implication. First there is the antecedent (on the left side), and then, *as a result*, the consequent (on the right side) follows. Another notion of implication is *material implication*, often denoted by $\rightarrow$:

$$
\begin{aligned}
p \rightarrow q \quad &\text{iff} \quad \neg p \vee q \\
&\text{iff} \quad \neg(p \wedge \neg q).
\end{aligned} \tag{C.1}
$$

Then some of the eight implications in exercises 1 and 2 are redundant. Which are they?[1]

5. Convince yourself that for the material implication (C.1),
a. $p \rightarrow q$ is equivalent to $\neg q \rightarrow \neg p$,
but
b. $p \rightarrow q$ is *not* equivalent to $q \rightarrow p$.

Chapter 2 Utility Maximization

1. To what degree is the function u in proposition 1 and theorem 2 (see section B.1 in appendix B) unique? That is, how much freedom does the modeler have in choosing the utility function u for a given relation $\succsim$?

2. Assume that apart from preferences between pairs of alternatives $x \succsim y$ or $y \succsim x$, more data are available, such as (1) the probability that x is chosen out of $\{x, y\}$; or (2) the time it takes the decision maker to make up her mind between x and y; or (3) some neurological data that show the strength of preference between x and y. Consider different representations of preferences, corresponding to (1)–(3), which will also restrict the set of utilities one can ascribe to the decision maker.

3. Assume that $X = \mathbb{R}^2$ and that because of some axioms, you are convinced that your utility function should be of the form

$$u(x_1, x_2) = v_1(x_1) + v_2(x_2).$$

Discuss how this additional structure may help you to estimate your own utility function, and contrast this case with the (end of the dialogue) we started out with.

1. Notation: $\neg p$ is the negation of p, i.e., not-p. $\vee$ means *or* and $\wedge$ means *and*.

4. Prove that if $\succsim$ is transitive, then so are $\succ$ and $\sim$.

5. Assume that $\succsim$ is complete. Prove that u represents $\succsim$ if and only if, for every $x, y \in X$,

$$x \succ y \Leftrightarrow u(x) > u(y).$$

6. Assume that $X = [0,1]^2$ and that $\succsim$ is defined by

$$(x_1, x_2) \succsim (y_1, y_2)$$

if

$$[x_1 > y_1]$$

or

$$[(x_1 = y_1) \quad \text{and} \quad (x_2 \geq y_2)].$$

Prove that $\succsim$ is complete and transitive but not continuous. Prove that $\succsim$ cannot be represented by any utility u (continuous or not).

Chapter 3 Constrained Optimization

1. You are organizing an interdisciplinary conference and wish to have a good mix of psychologists, economists, and sociologists. There are many scientists of each type, but the cost of inviting them grows with distance; it is relatively inexpensive to invite those that are in your city, but it gets expensive to fly them from remote countries. State the problem as a constrained optimization problem. Is this a convex problem? What do the first-order conditions tell you?

2. Provide an example of a consumer problem in which the optimal solution does not satisfy the first-order conditions. (Hint: Use two goods and a simple budget set such as that defined by $x_1 + x_2 \leq 100$.)

3. Suppose you have to allocate a given amount of time among several friends. Unfortunately, since they live far away, you can't meet more than one friend at the same time. Let x_i be the amount of time you spend with friend i. Formulate the problem as a constrained optimization problem. Is it convex?

4. Show that in the presence of discounts for quantity (that is, the price per unit goes down as you buy large quantities) the feasible set of the consumer is not convex.

5. Show that the intersection of convex sets is convex.

6. A half-space is defined by a (weak) linear inequality. That is, for a linear function $f : \mathbb{R}^n \to \mathbb{R}$ and a number $c \in \mathbb{R}$, it is the set of points

$$H(f, c) \equiv \{x \in \mathbb{R}^n \mid f(x) \leq c\}.$$

Show that the intersection of (any number of) half-spaces is convex.

Chapter 4 Expected Utility

1. A concave utility function can explain why people buy insurance with a negative expected value. And a convex utility function can explain why they buy lottery tickets, whose expected value is also negative. But how would you explain the fact that some people do both simultaneously?

2. Assume that a decision maker's preference $\succsim$ is representable by median utility maximization. That is, for a function $u : X \to \mathbb{R}$, and a lottery $P \in L$, define

$$med_P u = \max\left\{ \alpha \,\middle|\, \sum_{u(x)<\alpha} P(x) \leq \frac{1}{2} \right\}$$

and

$$P \succsim Q \Leftrightarrow med_P u \geq med_Q u$$

for all $P, Q \in L$.

Show that $\succsim$ is a weak order but that it violates continuity and independence.

3. If $\succsim$ is representable by median u maximization as in exercise 2, how unique is u? That is, what is the class of functions v such that median v maximization also represents $\succsim$?

4. Suppose that the utility function from money, u, is twice differentiable and satisfies $u' > \delta > 0$ and $u'' < 0$. Let X be a random variable assuming only two values, with $EX > 0$.

a. Show that for every wealth level W, there exists $\varepsilon > 0$ such that

$$E[u(W + \varepsilon X)] > u(W).$$

b. Show that there exists a wealth level W such that for all $w \geq W$,

$$E[u(w + X)] > u(w).$$

5. Show that if $\succsim$ satisfies the vNM axioms, then whenever $P \succ Q$,

$$\alpha P + (1-\alpha)Q \succsim \beta P + (1-\beta)Q \quad \text{iff} \quad \alpha \geq \beta.$$

6. a. Show that if $x, y, z \in X$ satisfy $x \succ y \succ z$ (where we abuse notation and identify each $x \in X$ with the lottery $P_x \in L$ such that $P_x(x) = 1$), there exists a unique $\alpha = \alpha(x, y, z)$ such that

$$\alpha x + (1-\alpha)z \sim y.$$

b. Assume that for some $x, z \in X$, we have $x \succsim y \succsim z$ for all $y \in X$. Define

$$u(y) = 1 \quad \text{if} \quad y \sim x,$$

$$u(y) = 0 \quad \text{if} \quad y \sim z,$$

$$u(y) = \alpha(x, y, z) \quad \text{if} \quad x \succ y \succ z.$$

Explain why maximization of the expectation of u represents $\succsim$. (Sketch the proof or, even better, write the complete proof formally.)

Chapter 5 Probability and Statistics

1. Explain what is wrong with the claim, "Most good chess players are Russian; therefore a Russian is likely to be a good chess player."

2. When one sails along the shores of the Mediterranean, it seems that much more of the shoreline has hills and cliffs than one would have thought. One theory is that the Earth was created with the tourism industry in mind. Another is that this is an instance of biased sampling. Explain why.

(Hint: Assume that the Earth is unidimensional and that its surface varies in slope. To be concrete, assume that the surface is made of the segments connecting $((0,0),(90,10))$ and $((90,10),(100,100))$, where the first coordinate denotes distance and the second height. Assume that the height of the water is randomly determined according to a uniform distribution over $[0,100]$. Compare the probability of the shore's being at a point of a steep slope to the probability you get if you sample a point at random (uniformly) on the distance axis.)

3. Comment on the claim, "Some of the greatest achievements in economics are due to people who studied mathematics. Therefore, all economists had better study mathematics first."

4. Consider exercise 5, following (even if you do not solve it), and explain how many prejudices in the social domain may result from ignoring base probabilities.

5. Trying to understand why people confuse $P(A|B)$ with $P(B|A)$, it is useful to see that qualitatively, if A makes B more likely, it will also be true that B will make A more likely.
a. Show that for any two events A, B,

$$P(A|B) > P(A|B^c)$$

iff

$$P(A|B) > P(A) > P(A|B^c)$$

iff

$$P(B|A) > P(B|A^c)$$

iff

$$P(B|A) > P(B) > P(B|A^c),$$

where A^c is the complement of A. (Assume that all probabilities involved are positive, so that all the conditional probabilities are well defined.)
b. If the proportion of Russians among the good chess players is higher than their proportion overall in the population, what can be said?

6. Consider a regression line relating the height of children to that of their parents. We know that its slope should be in $(0, 1)$. Now consider the following generation, and observe that the slope should again be in $(0, 1)$. Does this mean that because of regression to the mean, all the population will converge to a single height?

Chapter 6 Aggregation of Preferences

1. In order to determine a unique utility function for each individual, to be used in the summation of utilities across individuals, it was suggested to measure an individual's vNM utility functions (for choice under risk) and to set two arbitrary outcomes to given values (shared across individuals). Discuss this proposal.

2. The Eurovision song contest uses a scoring rule according to which each country ranks the other countries' songs and gives them scores according to this ranking. It has been claimed that the given scores favor standard songs over more innovative ones. Does this claim make sense? Is it more convincing when the score scale is convex or concave?

3. It turns out that for two particular individuals, Pareto domination defines a complete relation. (That is, for every two distinct alternatives, one Pareto-dominates the other.) Assume that

$$u(X) = \{(u_1(x), u_2(x)) \mid x \in X\}$$

is convex. What can you say about the utility functions of these individuals?

4. Assume that individual i has a utility function u_i. For $\alpha = (\alpha_1, \ldots, \alpha_n)$ with $\alpha_i > 0$, let

$$u_\alpha = \sum_{i=1}^{n} \alpha_i u_i.$$

Show that if x maximizes u_α for some α, it is Pareto-efficient.

5. Is it true that every Pareto-efficient alternative maximizes u_α for some α? (Hint: For $n = 2$, consider the feasible sets

$$X_1 = \{(x_1, x_2) \mid \sqrt{x_1} + \sqrt{x_2} \leq 1; \qquad x_1, x_2 \geq 0\}$$

and

$$X_2 = \{(x_1, x_2) \mid x_1^2 + x_2^2 \leq 1; \qquad x_1, x_2 \geq 0\},$$

where $u_i = x_i$.)

6. Show that under approval voting, it makes sense for each individual to approve of her most preferred alternative(s) and not to approve of the least preferred one(s) (assuming that the voter is not indifferent among all alternatives).

Chapter 7 Games and Equilibria

1. Suppose prisoner's dilemma is played T times between two players. Show that playing D is not a dominant strategy but that only Nash equilibria still result in consecutive play of D.

2. Consider the following story. In a certain village there are n married couples. It is the case that if one married woman is unfaithful to her husband, all other men are told about it immediately but not the husband. This fact is commonly known in the village. The law of the land is that if a husband knows that his wife has been unfaithful to him, he must shoot her to death on the same night. But he is not allowed to hurt her unless he knows that for sure.

One day a visitor comes to the village, gets everyone to meet in the central square, and says, "There are unfaithful wives in this village." He then leaves.

That night and the following night, nothing happens. On the third night, shots are heard.

a. How many shots were heard on the third night?

b. What information did the visitor add that the village inhabitants did not have before his visit?

3. Consider an extensive form game, and show how a player might falsify common knowledge of rationality (by deviating from the backward induction solution). Show an example in which it may be in the player's best interest to do so.

4. Compute the mixed strategies equilibria in the following games (see section 7.3 of main text):

Game 6. Pure Coordination 1

	R	L
R	$(1,1)$	$(0,0)$
L	$(0,0)$	$(1,1)$

Game 7. Pure Coordination 2

	A	B
A	$(3,3)$	$(0,0)$
B	$(0,0)$	$(1,1)$

Game 8. Battle of the Sexes

	Ballet	Boxing
Ballet	$(2,1)$	$(0,0)$
Boxing	$(0,0)$	$(1,2)$

5. Show that a 2×2 game in which all payoffs are different cannot have precisely two Nash equilibria.

6. A computer sends a message to another computer, and it is commonly known that the message never gets lost and that it takes 60 seconds to arrive. When it arrives, it is common knowledge (between the two computers) that the message has indeed been sent and has arrived. Next, a technological improvement was introduced, and the message can now take any length of time between 0 and 60 seconds to arrive. How long after the message was sent will it be commonly known that it has been sent?

Chapter 8 Free Markets

1. Discuss the reasons that equilibria might not be efficient in the following cases:

a. A physician should prescribe tests for a patient.

b. A lawyer assesses the probability of success of a legal battle.

c. A teacher is hired to teach a child.

2. The dean has to decide whether to give a department an overall budget for its activities or split the budget among several activities such as conferences, visitors, and so forth. Discuss the pros and cons of the two options.

3. Consider the student course assignment problem described in section 8.4 of the main text. Show that for every n it is possible to have examples in which n is the minimal number of students that can find a Pareto-improving reallocation of courses.

D Solutions

Chapter 1 Feasibility and Desirability

1. In the first example we saw an instance of

impossible ⇒ *undesirable*,

whereas the second was an instance of

possible ⇒ *undesirable*.

The third example is one in which

desirable ⇒ *possible*,

and this raises the question, what would be an example in which

desirable ⇒ *impossible*?

Solution You may be reflecting such a belief if you think, for instance, that any potential spouse you may like is bound to be married. Indeed, you may rely on statistics and a reasonable theory that says when you like someone, so do others, and therefore the object of desire is less likely to be available. But if we all believed that anything desirable is automatically impossible, those desirable potential spouses would end up remaining single. By analogy, it is true that one doesn't often see $100 bills on the sidewalk, but the reason is that they have indeed been picked up. Someone who sees the bill believes that it might be real and is willing to try to pick it up. Thus, you are justified in believing that something really worthwhile may not be easy to find, but you would be wrong to assume that anything worthwhile is automatically unreachable.

2. Symmetry requires that we also look for examples in which

possible $\Rightarrow$ *desirable*,

impossible $\Rightarrow$ *desirable*,

undesirable $\Rightarrow$ *possible*,

undesirable $\Rightarrow$ *impossible*.

Can you find such examples?

Solution

possible $\Rightarrow$ *desirable*. Habits may provide an example in which you do not try to optimize and assume that something is what you want only because you know you can have it.

impossible $\Rightarrow$ *desirable*. By the same token, it would also be irrational to want things just because you don't have them. Whereas the previous example leads to too little experimentation and may make you settle for suboptimal solutions, this example might lead to too much experimentation and not let you settle on an optimal solution even if you found it.

undesirable $\Rightarrow$ *possible*. The pessimistic assumption that you might be doing something just because you hope not to is reminiscent of "If something can go wrong, it will."

undesirable $\Rightarrow$ *impossible*. This is the optimistic version of the preceding, a bit similar to "This won't happen to me" (referring to negative events such as accidents).

3. George says, "I wish to live in a peaceful world. Therefore, I favor policies that promote world peace."
a. Explain why this statement violates the separation of feasibility and desirability.
b. Suppose George thinks that if a peaceful world is impossible, he is not interested in living any more, and further, he doesn't care about anything else that might happen in this world, to himself or to others. Explain why, under these assumptions, George's statement is compatible with rationality.

Solution This exercise is supposed to point out that people often think about what they want and then reason backward to see what's needed for that. This may be incompatible with rationality if they forget to ask what is feasible. In George's case, if (as in 3b) he doesn't care about anything but his one goal, then it makes sense to ignore what is precisely feasible. If peace is not feasible, he doesn't care about anything anyway. But for many people feasibility is important. Even when people say they want peace at all costs, they do not literally mean it. What we expect of rational decision makers is not to state just what is desirable but also what is feasible.

4. In the previous exercises, the symbol ⇒ referred to causal implication. First there is the antecedent (on the left side), and then, *as a result*, the consequent (on the right side) follows. Another notion of implication is *material implication*, often denoted by →:

$$p \to q \quad \text{iff} \quad \neg p \vee q \tag{C.1}$$

$$\text{iff} \quad \neg(p \wedge \neg q).$$

Then some of the eight implications in exercises 1 and 2 are redundant. Which are they?[1]

Solution With material implication, $p \to q$ is equivalent to $\neg q \to \neg p$ (see exercise 5, following). Hence,

possible → *desirable*

is equivalent to

undesirable → *impossible*,

and

possible → *undesirable*

is equivalent to

desirable → *impossible*.

Hence, half of the implications are redundant.

5. Convince yourself that for the material implication (C.1),
a. $p \to q$ is equivalent to $\neg q \to \neg p$.
b. $p \to q$ is *not* equivalent to $q \to p$.

1. Notation: $\neg p$ is the negation of p, i.e., not-p. $\vee$ means *or* and $\wedge$ means *and*.

Solution

Part 5a First, using proof by negation, assume that $p \to q$. We want to show that $\neg q \to \neg p$. Assume that indeed $\neg q$, that is, q is false. Ask whether p can be true. If it were (contrary to what we want to show), then we could use $p \to q$ to conclude that q is true as well, in contradiction to the assumption $\neg q$. Hence, by assuming $\neg q$, we obtained $\neg p$, which is the first part of what we wanted to prove.

Another way to see part 5a is to observe that $p \to q$ is simply the statement, "We cannot observe p and not-q simultaneously." That is, of the four possible combinations of truth values of p and of q, only three combinations are possible. The possible ones are marked by $+$, and the impossible one, by $-$:

	q is false	q is true
p is false	$+$	$+$
p is true	$-$	$+$

If we denote "p is true" by A and "q is false" by B, the statement $p \to q$ means that A and B cannot happen together. To say that two events, A and B, are incompatible is like saying "If A, then not B," or "If B, then not A."

Part 5b This is the converse. That is, assume $\neg q \to \neg p$; then $p \to q$. We could go through a similar proof as for part 5a (or use the previous one), observing that $\neg\neg p \leftrightarrow p$ and $\neg\neg q \leftrightarrow q$.

To see part 5b, take a simple example such as "Because all humans are mortal, *mortal* $\to$ *human*." But because dogs are also mortal, it is false that *mortal* $\to$ *human*. This looks trivial in such simple examples, and yet people make such mistakes often in the heat of a debate or when probabilities are involved.

Chapter 2 Utility Maximization

1. To what degree is the function u in proposition 1 and theorem 2 (see section B.1 in appendix B) unique? That is, how much freedom does the modeler have in choosing the utility function u for a given relation $\succsim$?

Solution The utility function is unique up to a monotone transformation. That is, if u represents $\succsim$, then so will any other function

$$v : X \to \mathbb{R}$$

such that there exists a (strictly) monotonically increasing

$$f : \mathbb{R} \to \mathbb{R}$$

for which

$$v(x) = f(u(x)) \tag{D.1}$$

for every $x \in X$. Conversely, if both u and v represent $\succsim$, it is easy to see that for every $x, y \in X$,

$$u(x) > u(y) \Leftrightarrow v(x) > v(y),$$

and this means that there exists a (strictly) monotonically increasing $f : \mathbb{R} \to \mathbb{R}$ such that (D.1) holds.

2. Assume that apart from preferences between pairs of alternatives $x \succsim y$ or $y \succsim x$, more data are available, such as (1) the probability that x is chosen out of $\{x, y\}$; or (2) the time it takes the decision maker to make up her mind between x and y; or (3) some neurological data that show the strength of preference between x and y. Consider different representations of preferences, corresponding to (1)–(3), which will also restrict the set of utilities one can ascribe to the decision maker.

Solution Assume that in reality there exists some numerical measure of desirability, $u(x)$, which is not directly observable. Yet we may find the following observable manifestations of this measure:

• The probability of choosing x out of $\{x, y\}$ may be increasing as a function of the utility difference, $u(x) - u(y)$. The standard model implicitly assumes that this probability is

$$\Pr(x\_over\_y) = \begin{cases} 1 & u(x) > u(y) \\ .5 & u(x) = u(y) \\ 0 & u(x) < u(y) \end{cases} \tag{D.2}$$

but this discontinuous function is not very realistic. Instead, we may consider a function such as the cumulative distribution function of the normal (Gaussian) distribution with parameters $(0, \sigma)$, namely,

$$\Pr(x\_over\_y) = \int_{-\infty}^{u(x)-u(y)} e^{-t^2/\sqrt{2\pi}\sigma}\, dt$$

such that $\Pr(x\_over\_y)$ converges to (D.2) as $\sigma \to 0$. This function would make utility differences observable by the probability of choice.

• The standard model ignores response time, or the time it takes the decision maker to reach a decision. We might consider a function such as

$$R(x,t) = c + de^{-\theta[u(x)-u(y)]^2}$$

such that $c > 0$ is the minimal response time, obtained when the choice is very clear (when the absolute difference between the utility levels is approaching infinity), and the maximal response time, $c + d$, is obtained when the two alternatives are equivalent in the eyes of the decision maker.

• Finally, the standard model treats anything that goes into our brains as unobservable. But recent neurological studies identify zones of the brain that tend to be activated when the alternatives are close to equivalent but not otherwise. Thus, neurological data may be another source of information on the strength of preferences.

Overall, the model in which the utility function is "only ordinal" and we therefore cannot discuss strength of preferences is a result of our highly idealized assumption that only choice is observable and that choice is deterministic, as in (D.2). It is our choice to focus on such a model. In reality, much more information is available, and this additional information may suffice to pinpoint a cardinal utility function, one that is more or less unique, at least up to a linear transformation of the type

$$v(x) = \alpha u(x) + \beta$$

with $\alpha > 0$.

3. Assume that $X = \mathbb{R}^2$ and that because of some axioms, you are convinced that your utility function should be of the form

$$u(x_1, x_2) = v_1(x_1) + v_2(x_2).$$

Discuss how this additional structure may help you to estimate your own utility function, and contrast this case with the (end of the dialogue) we started out with.

Solution In this case, one can try to learn something about one's preferences in complex choices from one's preferences in simple ones. For example, suppose that after intensive introspection you realize that your preferences satisfy

$$(x_1, x_2) \sim (y_1, y_2)$$

and

$$(z_1, x_2) \sim (w_1, y_2).$$

The first equivalence means that

$$v_1(x_1) - v_1(y_1) = v_2(y_2) - v_2(x_2), \quad \text{(D.3)}$$

and the second, that

$$v_1(z_1) - v_1(w_1) = v_2(y_2) - v_2(x_2). \quad \text{(D.4)}$$

Next suppose also that

$$(x_1, s_2) \sim (y_1, r_2),$$

which means that

$$v_1(x_1) - v_1(y_1) = v_2(r_2) - v_2(s_2). \quad \text{(D.5)}$$

It then follows that we should also have

$$(z_1, s_2) \sim (w_1, r_2)$$

because we already know that (combining (D.3) and (D.4))

$$v_1(z_1) - v_1(w_1) = v_2(y_2) - v_2(x_2) = v_1(x_1) - v_1(y_1)$$

and because of (D.5), also

$$v_1(z_1) - v_1(w_1) = v_1(x_1) - v_1(y_1) = v_2(r_2) - v_2(s_2).$$

In other words, additional structure on the utility function will make the elicitation of utility a noncircular exercise.

4. Prove that if $\succsim$ is transitive, then so are $\succ$ and $\sim$.

Solution In this type of proof the main thing is to keep track of what is given and what is to be proved. Most mistakes in such exercises arise from getting confused about this. Also, much of the proof is a translation of the symbols using their definitions. For these reasons it is best to write things down very carefully and precisely, even though it might seem silly or boring.

Let us begin with $\sim$. Assume that (for some $x, y, z \in X$) $x \sim y$ and that $y \sim z$. We need to show that $x \sim z$. Let us first translate both premise and desired conclusion to the language of the relation $\succsim$ about which we know something (i.e., that it is transitive).

By definition of $\sim$, $x \sim y$ means that

$$x \succsim y \quad \text{and} \quad y \succsim x, \tag{D.6}$$

whereas $y \sim z$ is a shorthand for

$$y \succsim z \quad \text{and} \quad z \succsim y. \tag{D.7}$$

What we need to prove is that $y \sim z$, namely, that

$$x \succsim z \quad \text{and} \quad z \succsim x. \tag{D.8}$$

The first parts of (D.6) and (D.7) are, respectively, $x \succsim y$ and $y \succsim z$, and given the transitivity of $\succsim$, they yield $x \succsim z$. This is the first part of (D.8).

Similarly, the second parts of (D.6) and (D.7) are, respectively, $y \succsim x$ and $z \succsim y$, which given transitivity of $\succsim$ imply $z \succsim x$. This is the second part of (D.8).

Since this is true for any x, y, z with $x \sim y$ and $y \sim z$, transitivity of $\sim$ has been established.

Next turn to transitivity of $\succ$. Assume that (for some $x, y, z \in X$) $x \succ y$ and $y \succ z$. We need to show that $x \succ z$. Again, let us translate both the premises and the desired conclusion to the language of $\succsim$.

By definition of $\succ$, $x \succ y$ means that

$$x \succsim y \quad \text{and} \quad \neg(y \succsim x), \tag{D.9}$$

whereas $y \succ z$ is the statement

$$y \succsim z \quad \text{and} \quad \neg(z \succsim y). \tag{D.10}$$

We need to show that $x \succ z$, which means

$$x \succsim z \quad \text{and} \quad \neg(z \succsim x). \tag{D.11}$$

The first part of (D.11) follows from the first parts of (D.9) and of (D.10) by transitivity of $\succsim$. The second part of (D.11) will be proved by negation. Suppose that, contrary to our claim, $z \succsim x$ does hold. Combining this with $x \succsim y$ (the first part of (D.9)), we get, by transitivity of $\succsim$, that $z \succsim y$. But this would contradict the second part of (D.10). Hence $z \succsim x$ cannot be true, and the second part of (D.11) is also true. Again, this holds for every x, y, z, and this completes the proof.

5. Assume that $\succsim$ is complete. Prove that u represents $\succsim$ if and only if, for every $x, y \in X$,

$$x \succ y \Leftrightarrow u(x) > u(y).$$

Solution The simplest way to see this is to observe that for real numbers a, b,

$$\neg(a \geq b) \Leftrightarrow b > a,$$

and, because $\succsim$ is complete, a similar fact holds for preferences: for every $z, w \in X$,

$$\neg(x \succsim y) \Leftrightarrow y \succ x.$$

Once this is established, we can use the contrapositive (exercise 5a in the previous section (chapter 1)) to conclude the proof. But before we do so, a word of warning. We know that

$$x \succsim y \Leftrightarrow u(x) \geq u(y), \qquad \forall x, y \in X, \tag{D.12}$$

and we need to show that

$$x \succ y \Leftrightarrow u(x) > u(y), \qquad \forall x, y \in X. \tag{D.13}$$

This is rather simple unless we get ourselves thoroughly confused with the x's in (D.12) and in (D.13). It is therefore a great idea to replace (D.13) by

$$z \succ w \Leftrightarrow u(z) > u(w), \qquad \forall z, w \in X. \tag{D.14}$$

You can verify that (D.13) and (D.14) mean the same thing. Since we range over all x, y in (D.13) and over all z, w in (D.14), these variables have no existence outside the respective expressions. Replacing "all x" by "all z" is similar to changing the index inside a summation. That is, just as

$$\sum_{i=1}^{n} a_i = \sum_{j=1}^{n} a_j,$$

the statements (D.13) and (D.14) are identical.

If we agree on this, we can now observe that, for every x, y,

$$x \succsim y \Rightarrow u(x) \geq u(y)$$

is equivalent to

$\neg(u(x) \geq u(y)) \Rightarrow \neg(x \succsim y)$

or to

$u(y) > u(x) \Rightarrow y \succ x.$

Thus, for every $z(= y)$, and $w(= x)$,

$u(z) > u(w) \Rightarrow z \succ w.$

Similarly, for every x, y,

$u(x) \geq u(y) \Rightarrow x \succsim y$

is equivalent to

$\neg(x \succsim y) \Rightarrow \neg(u(x) \geq u(y))$

or to

$y \succ x \Rightarrow u(y) > u(x)$

and, again, for every $z(= y)$, and $w(= x)$,

$z \succ w \Rightarrow u(z) > u(w).$

6. Assume that $X = [0, 1]^2$ and that $\succsim$ is defined by

$(x_1, x_2) \succsim (y_1, y_2)$

if

$[x_1 > y_1]$

or

$[(x_1 = y_1) \quad \text{and} \quad (x_2 \geq y_2)].$

Prove that $\succsim$ is complete and transitive but not continuous. Prove that $\succsim$ cannot be represented by any utility u (continuous or not).

Solution To see that $\succsim$ is complete, consider (x_1, x_2) and (y_1, y_2). If $x_1 > y_1$, then $(x_1, x_2) \succsim (y_1, y_2)$. Similarly, $y_1 > x_1$ implies $(y_1, y_2) \succsim (x_1, x_2)$. We are left with the case $x_1 = y_1$. But then $x_2 \geq y_2$ (and $(x_1, x_2) \succsim (y_1, y_2)$) or $y_2 \geq x_2$ (and then $(y_1, y_2) \succsim (x_1, x_2)$).

Next turn to transitivity. Assume that $(x_1, x_2) \succsim (y_1, y_2)$ and $(y_1, y_2) \succsim (z_1, z_2)$. If $x_1 > y_1$, or $y_1 > z_1$, then $x_1 > z_1$ and $(x_1, x_2) \succsim (z_1, z_2)$ follows. Otherwise, $x_1 = y_1 = z_1$. Then $(x_1, x_2) \succsim (y_1, y_2)$ implies

$x_2 \geq y_2$, and $(y_1, y_2) \succsim (z_1, z_2)$ implies $y_2 \geq z_2$. Together we have $x_2 \geq z_2$, which implies (since we already know that $x_1 = z_1$) that $(x_1, x_2) \succsim (z_1, z_2)$.

To see that continuity does not hold, consider $y = (y_1, y_2)$ with

$$y_1 = 0.5,$$

$$y_2 = 1,$$

and the set

$$B(y) = \{(x_1, x_2) \mid (y_1, y_2) \succ (x_1, x_2)\}.$$

You can verify that $(0.5, 0) \in B(y)$, but for every $\varepsilon > 0$, $(0.5 + \varepsilon, 0) \notin B(y)$ (because $(0.5 + \varepsilon, 0) \succ (0.5, 1) = y$). Hence $B(y)$ is not open, which is sufficient to show that continuity of $\succsim$ does not hold.

We know from Debreu's theorem that $\succsim$ cannot be represented by a continuous utility function. This can also be verified directly in this example. Indeed, if there were a continuous u that represented $\succsim$, we would have

$$u((0.5 + \varepsilon, 0)) > u(y) > u((0.5, 0)) \tag{D.15}$$

for every $\varepsilon > 0$. But this is incompatible with continuity because

$$(0.5 + \varepsilon, 0) \to (0.5, 0)$$

as $\varepsilon \to 0$, and continuity would have implied that

$$u((0.5 + \varepsilon, 0)) \to u((0.5, 0)),$$

whereas (D.15) means that the left side of the preceding statement is bounded below by a number ($u(y)$) strictly larger than $u((0.5, 0))$.

To see that no utility function can represent $\succsim$ requires a little more knowledge of set theory. We can try intuition here. If u represented $\succsim$, then the function

$$w(z) = u((z, 0))$$

has a discontinuity from the right at $z = 0.5$. That is, as we have just seen,

$$\lim_{\varepsilon \to 0} u((z + \varepsilon, 0)) > u((z, 0)).$$

By now we don't expect u to be continuous. But the above is true not only for $z = 0.5$ but for *any* $z \in (0, 1)$. And $w(z)$ is a monotone function (the higher is z, the better is $(z, 0)$ and the higher should be $u((z, 0)) = w(z)$).

The contradiction arises from the fact that a monotone function can have jumps, but not *everywhere*. Roughly, this has to do with the fact that the set of jumps of a monotone function is countable, whereas the interval $(0, 1)$ is not.

This lexicographic example might seem like a mathematical oddity. But lexicographic relations often appear in everyday speech. For instance, one can imagine a politician's saying that we will give the public the best health care possible, but subject to this level of health care, we will save on costs. Or that we will promote minority candidates, provided that we do not compromise on quality. These are examples of lexicographic relations. These are also often examples of dishonesty. Typically, trade-offs do exist. If one needs to save money on health care, one might have to compromise on the quality of health care. If one wants to promote a social agenda and help minorities, one might have to compromise on quality. Politicians often try to disguise such compromises. This lexicographic example, showing that we can easily describe a function that cannot be represented numerically, suggests that perhaps politicians do not really mean what they say. That is, it might be more honest to describe a continuous trade-off, as in, "We have to cut on health costs, and we will try to do it without hurting the quality of the service *too much*." Or, "It's important to have affirmative action, and we are willing to pay some price for that." When you hear someone describing preferences lexicographically, ask whether they really mean what they say.

Chapter 3 Constrained Optimization

1. You are organizing an interdisciplinary conference and wish to have a good mix of psychologists, economists, and sociologists. There are many scientists of each type, but the cost of inviting them grows with distance; it is relatively inexpensive to invite those that are in your city, but it gets expensive to fly them from remote countries. State the problem as a constrained optimization problem. Is this a convex problem? What do the first-order conditions tell you?

Solution Suppose that you invite x_1 psychologists, x_2 economists, and x_3 sociologists, and make the unrealistic but convenient assumption that these are real numbers, that is, that scientists are divisible. Let

$u(x_1, x_2, x_3)$ be a measure of how good the conference is as a function of the number of scientists invited from each group. An even more unrealistic assumption here is that all psychologists are interchangeable, as are all economists (among themselves) and all sociologists. This is implicit in the formulation asking "How many should we invite?" while ignoring their identity.

The story implicitly refers to only one constraint, namely, cost. However, it is clear that cost is not linear because it grows with distance. Let $c_i(x_i)$ be the cost of inviting x_i scientists of type i (1 for psychologists, 2 for economists, and 3 for sociologists), and let B be the overall budget. The optimization problem is then

$$\max_{x_1, x_2, x_3} u(x_1, x_2, x_3)$$

subject to

$$c_1(x_1) + c_2(x_2) + c_3(x_3) \leq B$$

$$x_i \geq 0$$

This will be a convex problem provided that the cost functions are weakly convex and that the utility function is quasi-concave. Specifically, if the cost functions are (weakly) convex, then for every $i = 1, 2, 3$, every $x_i, y_i \geq 0$, and every $\alpha \in [0, 1]$,

$$\alpha c_i(x_i) + (1 - \alpha)c_i(y_i) \geq c_i(\alpha x_i + (1 - \alpha)y_i),$$

and this means that $x = (x_1, x_2, x_3)$ and $y = (y_1, y_2, y_3)$ are in the feasible set, then so is $\alpha x + (1 - \alpha)y$. Non-negativity of x_i and of y_i implies non-negativity of $\alpha x_i + (1 - \alpha)y_i$, and

$$\sum_{i=1}^{3} c_i(\alpha x_i + (1 - \alpha)y_i) \leq \sum_{i=1}^{3} [\alpha c_i(x_i) + (1 - \alpha)c_i(y_i)]$$

$$= \alpha \sum_{i=1}^{3} c_i(x_i) + (1 - \alpha) \sum_{i=1}^{3} c_i(y_i) \leq \alpha B + (1 - \alpha)B = B.$$

Quasi-concavity of u means precisely that the "better than" sets are convex, that is, the set

$$\{x \in R^3 \mid u(x_1, x_2, x_3) \geq \gamma\}$$

is convex for every γ. With a convex feasible set and convex "better than" sets, the problem is convex.

The first-order conditions can be obtained from taking the derivatives of the Lagrangian,

$$L(x_1, x_2, x_3, \lambda) = u(x_1, x_2, x_3) + \lambda[B - c_1(x_1) + c_2(x_2) + c_3(x_3)],$$

which yield

$$\frac{\partial L}{\partial x_i} = u_i(x_1, x_2, x_3) - \lambda c_i'(x_i),$$

with $u_i(x_1, x_2, x_3) = \frac{\partial u}{\partial x_i} u(x_1, x_2, x_3)$. Equating all to zero, we get

$$\frac{u_i(x_1, x_2, x_3)}{c_i'(x_i)} = \lambda,$$

that is, the ratio of the marginal utility to marginal cost should be the same across all decision variables x_i. Given that the problem is convex, if we find such a point, it is optimal. Note, however, that such a point may not exist, and the optimal problem may well be at a corner solution, for example, if sociologists turn out to be too expensive and the optimal solution is to invite none of them ($x_3 = 0$).

2. Provide an example of a consumer problem in which the optimal solution does not satisfy the first-order conditions. (Hint: Use two goods and a simple budget set such as that defined by $x_1 + x_2 \leq 100$.)

Solution Given the budget constraint $x_1 + x_2 \leq 100$, consider the utility function

$$u(x_1, x_2) = 2x_1 + x_2.$$

Clearly, the optimal solution is at $(100, 0)$. You can also generate such an example if the utility function is strictly quasi-concave. All you need to guarantee is that the slope of the indifference curves will be steep enough so that there will be no tangency point between these curves and the budget line. Specifically, if throughout the range

$$\frac{u_1}{u_2} > 1,$$

the optimal solution will be at $(100, 0)$ without the marginality condition holding.

3. Suppose you have to allocate a given amount of time among several friends. Unfortunately, since they live far away, you can't meet more than one friend at the same time. Let x_i be the amount of time you spend with friend i. Formulate the problem as a constrained optimization problem. Is it convex?

Solution The problem might look like

$$\max_{x_1,\ldots,x_n} u(x_1,\ldots,x_n)$$

subject to

$$x_1 + \cdots + x_n \leq B$$

$$x_i \geq 0$$

that is, like a standard consumer problem where the prices of all goods are 1.

If you like to see each friend as much as possible, u will be monotonically increasing. For the problem to be convex, you would like u to be quasi-concave. That is, consider two feasible time allocation vectors, $(x_1,\ldots,x_n)$ and $(y_1,\ldots,y_n)$. If each guarantees a utility value of c at least, so should

$$\lambda(x_1,\ldots,x_n) + (1-\lambda)(y_1,\ldots,y_n).$$

This is a reasonable condition if at any level of the variables, you prefer to mix and have some variety. But it may not hold if the values are too low. For instance, if you mainly derive pleasure from gossip, it seems that frequent changes among friends is a great thing. But if you wish to get into a deep conversation about your emotional life, you may find that one hour with one friend is better than six ten-minute sessions with different friends.

4. Show that in the presence of discounts for quantity (that is, the price per unit goes down as you buy large quantities) the feasible set of the consumer is not convex.

Solution Suppose that the prices of goods 1 and 2 are $p_1 = p_2 = 1$, and that you have income of $I = 200$. But if $x_1 > 100$, the price of good 1 drops to $1/2$. Then the feasible (budget) set is bounded above by the segment connecting $(0, 200)$ and $(100, 100)$ (for $0 \leq x_1 \leq 100$) and by the segment connecting $(100, 100)$ and $(300, 0)$. Thus, the points

$A = (0, 200)$ and $B = (300, 0)$ are in the feasible set, but the point $(100, 133\frac{1}{3})$, which is on the segment connecting them, is not in the feasible set.

5. Show that the intersection of convex sets is convex.

Solution Let there be two convex sets $A, B \subset \mathbb{R}^n$. Consider $C = A \cap B = \{x | x \in A \text{ and } x \in B\}$. To see that C is convex, consider $x, y \in C$ and $\lambda \in [0, 1]$. We need to show that

$$\lambda x + (1 - \lambda)y \in C.$$

By convexity of A (and since $x, y \in A$), we get $\lambda x + (1 - \lambda)y \in A$. Similarly, $\lambda x + (1 - \lambda)y \in B$. But this means that $\lambda x + (1 - \lambda)y$ is both in A and in B, that is, in their intersection, C.

6. A half-space is defined by a (weak) linear inequality. That is, for a linear function $f : \mathbb{R}^n \to \mathbb{R}$ and a number $c \in \mathbb{R}$, it is the set of points

$$H(f, c) \equiv \{x \in \mathbb{R}^n \mid f(x) \leq c\}.$$

Show that the intersection of (any number of) half-spaces is convex.

Solution First, we need to convince ourselves that a single half-space is convex. To see this, assume that

$$x, y \in H(f, c),$$

that is,

$$f(x), f(y) \leq c.$$

Because f is linear, for any $\lambda \in [0, 1]$ we have

$$f(\lambda x + (1 - \lambda)y) = \lambda f(x) + (1 - \lambda)f(y);$$

hence

$$f(\lambda x + (1 - \lambda)y) \leq c$$

and

$$\lambda x + (1 - \lambda)y \in H(f, c),$$

that is, $H(f, c)$ is convex.

Next, we show that any intersection of convex sets is convex. We follow the same reasoning that applied in exercise 5 for two sets to any collection of sets. That is, assume that $\{A_\alpha\}_\alpha$ is some collection of convex sets, where α is an index that ranges over a certain set. (If α assumes only finitely many values, you can apply the conclusion of exercise 5 inductively. But the fact is true even if there are infinitely many α's.) Then

$$A^* = \cap_\alpha A_\alpha = \{x | x \in A_\alpha \ \forall\alpha\}$$

is convex because for any $x, y \in A^*$ and any $\lambda \in [0,1]$, we have

$$x, y \in A_\alpha, \qquad \forall\alpha,$$

and by convexity of A_α,

$$\lambda x + (1-\lambda)y \in A_\alpha, \qquad \forall\alpha,$$

and this means

$$\lambda x + (1-\lambda)y \in \cap_\alpha A_\alpha = A^*.$$

Hence, $x, y \in A^*$ implies that $\lambda x + (1-\lambda)y \in A^*$ for any $\lambda \in [0,1]$, and this is the definition of a convex set.

Chapter 4 Expected Utility

1. A concave utility function can explain why people buy insurance with a negative expected value. And a convex utility function can explain why they buy lottery tickets, whose expected value is also negative. But how would you explain the fact that some people do both simultaneously?

Solution One explanation is that the utility function looks like an inverse S: concave up to a certain point and convex thereafter. Imagine that w is the inflection point, the wealth level above which u is convex and below which it is concave. Then, if the decision maker is at w, considering a major loss (as in the case of insurance), she behaves in a risk-averse manner, but considering a major gain (as in a lottery), she behaves in a risk-loving manner.

The problem with this explanation is that it seems unlikely that all the people who both insure their property and buy lottery tickets are at the inflection point of their utility function. Another explanation is that this inflection point moves around with the current wealth level: the utility function depends on the wealth the individual already has.

This is very similar to the idea of a reference point (Kahneman and Tversky 1979). They argued that people respond to changes in the wealth level rather than to absolute levels of wealth. Moreover, they suggested that people react differently to gains as compared to losses. However, they found in their experiments that people are risk-averse when it comes to gains and risk-loving when it comes to losses. This appears to be in contradiction to the S-shaped utility function. At the same time, the sums of gains and losses involved in lotteries and insurance problems are much larger than the sums used in experiments.

Another explanation of the gambling behavior is that gambling is not captured by expected utility maximization at all. Rather, gambling has an entertainment value (people enjoy the game) or a fantasy value (people enjoy fantasizing about what they will do with the money they win). And these cannot be captured by expectation of a utility function, which is defined over outcomes alone.

2. Assume that a decision maker's preference $\succsim$ is representable by median utility maximization. That is, for a function $u : X \to \mathbb{R}$, and a lottery $P \in L$, define

$$med_P u = \max\left\{ \alpha \;\middle|\; \sum_{u(x)<\alpha} P(x) \leq \frac{1}{2} \right\}$$

and

$$P \succsim Q \Leftrightarrow med_P u \geq med_Q u$$

for all $P, Q \in L$.

Show that $\succsim$ is a weak order but that it violates continuity and independence.

Solution To see that $\succsim$ is a weak order, it suffices to note that it is defined by maximization of a real-valued function. Since every lottery P is mapped to

$$V(P) = \max\left\{ \alpha \;\middle|\; \sum_{u(x)<\alpha} P(x) \leq \frac{1}{2} \right\},$$

and the decision maker maximizes $V(P)$, the relation is complete and transitive (as is the relation $\geq$ on the real numbers).

To see that $\succsim$ is not continuous, assume for simplicity that $X = \mathbb{R}$ and $u(x) = x$, and consider the lotteries P, Q, R defined as follows:

$$P = \begin{cases} 10 & 0.5 \\ 0 & 0.5 \end{cases}.$$

Q guarantees outcome 5 with probability 1; and R guarantees the outcome 0 with probability 1. Then

$$V(P) = 10, \qquad V(Q) = 5, \qquad V(R) = 0;$$

hence

$$P \succ Q \succ R.$$

However, for any $\alpha \in (0, 1)$,

$$V(\alpha P + (1 - \alpha)R) = 0 < V(Q)$$

and

$$Q \succ \alpha P + (1 - \alpha)R,$$

which contradicts the continuity axiom.

As for independence, consider the same P, Q, R, and observe that

$$P \succ Q,$$

but if we mix them with R and $\alpha = 0.7$, we get

$$V(0.7P + 0.3R) = 0,$$

$$V(0.7Q + 0.3R) = 5,$$

and

$$0.7Q + 0.3R \succ 0.7P + 0.3R$$

in violation of the independence axiom, which would have implied

$$\alpha P + (1 - \alpha)R \succ \alpha Q + (1 - \alpha)R.$$

3. If $\succsim$ is representable by median u maximization as in exercise 2, how unique is u? That is, what is the class of functions v such that median v maximization also represents $\succsim$?

Solution In this case, u is unique up to (any) monotone transformation. The median ranking depends only on the ordering of the various outcomes, and thus any transformation that preserves this ordering can also serve as the utility function.

4. Suppose that the utility function from money, u, is twice differentiable and satisfies $u' > \delta > 0$ and $u'' < 0$. Let X be a random variable assuming only two values, with $EX > 0$.
a. Show that for every wealth level W, there exists $\varepsilon > 0$ such that

$$E[u(W + \varepsilon X)] > u(W).$$

b. Show that there exists a wealth level W such that for all $w \geq W$,

$$E[u(w + X)] > u(w).$$

Solution Assume that

$$X = \begin{cases} a & p \\ b & 1 - p \end{cases}$$

with $a > 0 > b$ and $EX = pa + (1 - p)b > 0$. Denote this expectation by $c = EX$.

We know that if the utility function were linear ($u'' = 0$), the decision maker would prefer to add αX to her current wealth level w, for any $\alpha > 0$ and any w. This is so because for a linear u,

$$E[u(w + \alpha X)] = u(w) + \alpha EX > u(w).$$

For risk-averse decision makers, this may not hold in general. However, we should expect it to be true if u can be approximated by a linear function, that is, if the decision maker is roughly risk-neutral.

In parts 4a and 4b, we have different reasons for thinking of the decision maker as roughly risk-neutral, that is, to approximating her utility function by a linear one. In the first case, the approximation is local, with the tangent to the utility function's graph as the linear approximation. In the second case, the utility function has a decreasing but positive derivative, and it therefore has to converge to a constant derivative, that is, to a linear function. More details follow.

Part 4a Here we want to approximate $u(x)$ by

$$v(x) = u(W) + (x - W)u'(W),$$

that is, by the tangent to the curve of u at W.

To simplify notation, we may change the variable so that $W = 0$. (Formally, introduce a new variable $y = x - W$.) Also, since u is given up to a positive linear transformation, no loss of generality is involved in assuming that $u(0) = 0$ and $u'(0) = 1$. Under these assumptions, we also have

$$v(x) = x.$$

Thus, the expected v-value of αX is simply $\alpha c > 0$, for any $\alpha > 0$.

Differentiability of u means that

$$\left|\frac{u(x) - x}{x}\right| \to_{x \to 0} 0.$$

Now consider the expected utility of $w + \varepsilon X = \varepsilon X$. We have

$$E[u(\varepsilon X)] = pu(\varepsilon a) + (1 - p)u(\varepsilon b),$$

and we wish to approximate it by the expected utility of $v(x) = x$, which is

$$E[\varepsilon X] = p\varepsilon a + (1 - p)\varepsilon b = \varepsilon c > 0.$$

Explicitly,

$$\begin{aligned} E[u(\varepsilon X)] &= pu(\varepsilon a) + (1 - p)u(\varepsilon b) \\ &= p\varepsilon a + p[u(\varepsilon a) - \varepsilon a] + (1 - p)\varepsilon b + (1 - p)[u(\varepsilon b) - \varepsilon b] \\ &= \varepsilon c + p\varepsilon a\left[\frac{u(\varepsilon a) - \varepsilon a}{\varepsilon a}\right] + (1 - p)\varepsilon b\left[\frac{u(\varepsilon b) - \varepsilon b}{\varepsilon b}\right] \\ &= \varepsilon\left(c + pa\left[\frac{u(\varepsilon a) - \varepsilon a}{\varepsilon a}\right] + (1 - p)b\left[\frac{u(\varepsilon b) - \varepsilon b}{\varepsilon b}\right]\right) \end{aligned}$$

or

$$\frac{E[u(\varepsilon X)]}{\varepsilon} = c + pa\left[\frac{u(\varepsilon a) - \varepsilon a}{\varepsilon a}\right] + (1 - p)b\left[\frac{u(\varepsilon b) - \varepsilon b}{\varepsilon b}\right].$$

Since the two expressions in brackets converge to zero as $\varepsilon \to 0$, the expression converges to $c > 0$. This proves our claim.

The meaning of this result is that if a decision maker has a constant (risk-free) asset W, and she has the opportunity to invest in an asset X with positive expected value, she would invest at least some amount $\varepsilon > 0$ in the asset X, even if she is risk-averse.

This conclusion may not be entirely realistic because the expected utility gain, for a very small ε, may not exceed the transaction cost (say, of buying an asset), and it may be also just too small for the decision maker to notice.

Part 4b Since the derivative of the utility function, u', is positive and decreasing (because u is increasing and concave), we know that it converges to a limit:

$$u'(w) \searrow_{w\to\infty} d \geq \delta > 0$$

(the notation $\searrow$ means "converges from above").

Consider the expected utility of getting the asset X with initial assets fixed at w:

$$E[u(w+X)] = pu(w+a) + (1-p)u(w+b).$$

We wish to show that for a large enough w the expected utility is higher than $Eu(w) = u(w)$. That is, we wish to show that the following expression is positive (for w large enough):

$$E[u(w+X)] - u(w).$$

Observe that

$$\begin{aligned} &E[u(w+X)] - u(w) \\ &\quad = pu(w+a) + (1-p)u(w+b) - [pu(w) + (1-p)u(w)] \\ &\quad = p[u(w+a) - u(w)] + (1-p)[u(w+b) - u(w)]. \end{aligned}$$

We know that a difference of the values of a differentiable function between two points is equal to the distance between the points times the derivative at some point between them. That is, for w and $w+a$, there exists $w' \in [w, w+a]$ such that

$$u(w+a) - u(w) = au'(w'),$$

and there also exists $w'' \in [w+b, w]$ such that

$$u(w+b) - u(w) = bu'(w'').$$

Using these, we can write

$$\begin{aligned} &E[u(w+X)] - u(w) \\ &\quad = pau'(w') + (1-p)bu'(w'') \end{aligned}$$

$$= pau'(w') + (1-p)bu'(w') + (1-p)b[u'(w'') - u'(w')]$$

$$= cu'(w') + (1-p)b[u'(w'') - u'(w')]$$

$$= u'(w')\left[c + (1-p)b\frac{u'(w'') - u'(w')}{u'(w')}\right].$$

As $w \to \infty$, $w', w'' \in [w+b, w+a]$ also converge to infinity, and $u'(w')$, $u'(w'') \to d$. This implies that

$$u'(w') - u'(w'') \to 0,$$

and because the denominator $u'(w') \geq d > 0$, the expression in brackets above converges to $c > 0$. Hence, the entire expression converges to $dc > 0$, and for all w from that point on, $E[u(w+X)]$ will be strictly higher than $u(w)$.

The meaning of this result is that when one becomes very rich, one tends to be risk-neutral. This may not be realistic because, as Kahneman and Tversky pointed out, people react to changes in their reference point, not to absolute levels of overall wealth.

5. Show that if $\succsim$ satisfies the vNM axioms, then whenever $P \succ Q$,

$$\alpha P + (1-\alpha)Q \succsim \beta P + (1-\beta)Q \quad \text{iff} \quad \alpha \geq \beta.$$

Solution Assume that $P \succ Q$. Consider $\alpha \in (0,1)$. Use the independence axiom with P, Q, and $R = Q$ to obtain

$$\alpha P + (1-\alpha)Q \succ Q,$$

and the same axioms with P, Q, and $R = P$ to obtain

$$P \succ \alpha P + (1-\alpha)Q.$$

Thus, whenever $P \succ Q$,

$$P \succ \alpha P + (1-\alpha)Q \succ Q.$$

Next, consider $\alpha, \beta \in (0,1)$. If $\alpha = \beta$, then the equivalence $\alpha P + (1-\alpha)Q \sim \beta P + (1-\beta)Q$ is trivial (because it is precisely the same lottery on both sides). Assume, then, without loss of generality, that $\alpha > \beta$. The point to note is that $\beta P + (1-\beta)Q$ can be described as a combination of $\alpha P + (1-\alpha)Q$ and Q. Specifically, denote

$$P' = \alpha P + (1-\alpha)Q,$$

$Q' = Q,$

$\gamma = \dfrac{\beta}{\alpha} \in (0, 1).$

Then we have $P' \succ Q'$, and by the first part of the proof,

$P' \succ \gamma P' + (1 - \gamma)Q',$

but

$$\begin{aligned}\gamma P' + (1-\gamma)Q' &= \gamma[\alpha P + (1-\alpha)Q] + (1-\gamma)Q \\ &= \frac{\beta}{\alpha}[\alpha P + (1-\alpha)Q] + \left(1 - \frac{\beta}{\alpha}\right)Q \\ &= \beta P + (1-\beta)Q,\end{aligned}$$

and the conclusion $\alpha P + (1 - \alpha)Q = P' \succ \beta P + (1 - \beta)Q$ follows.

6. a. Show that if $x, y, z \in X$ satisfy $x \succ y \succ z$ (where we abuse notation and identify each $x \in X$ with the lottery $P_x \in L$ such that $P_x(x) = 1$), there exists a unique $\alpha = \alpha(x, y, z)$ such that

$\alpha x + (1 - \alpha)z \sim y.$

b. Assume that for some $x, z \in X$, we have $x \succsim y \succsim z$ for all $y \in X$. Define

$u(y) = 1 \quad \text{if} \quad y \sim x,$

$u(y) = 0 \quad \text{if} \quad y \sim z,$

$u(y) = \alpha(x, y, z) \quad \text{if} \quad x \succ y \succ z.$

Explain why maximization of the expectation of u represents $\succsim$. (Sketch the proof or, even better, write the complete proof formally.)

Solution

Part 6a Consider the sets

$A = \{\alpha \in [0, 1] \mid \alpha x + (1 - \alpha)z \succ y\}$

and

$B = \{\alpha \in [0, 1] \mid y \succ \alpha x + (1 - \alpha)z\}.$

We know that $0 \in B$ and $1 \in A$, and from exercise 5, we also know that both A and B are contiguous intervals. Obviously, they are disjoint. The question is, can they cover the entire segment $[0, 1]$, or does there have to be something in between?

The answer is given by the continuity axiom. It says that both A and B are open: if α is in A, then for a small enough ε, $\alpha - \varepsilon$ is also in A. Similarly, if $\alpha \in B$, then for a small enough ε, $\alpha + \varepsilon \in B$. Together, this implies that A and B cannot cover all of $[0, 1]$, and a point

$$\alpha \in [0, 1] \backslash (A \cup B)$$

has to satisfy

$$\alpha x + (1 - \alpha) z \sim y.$$

To see that this α is unique, it suffices to use exercise 5. The strict preference between x and z implies that no distinct α, β can yield an equivalence

$$\alpha x + (1 - \alpha) z \sim y \sim \beta x + (1 - \beta) z.$$

Part 6b Consider a lottery

$$P = (p_1, x_1; p_2, x_2; \ldots; p_n, x_n).$$

Since

$$x_1 \sim \alpha(x, x_1, z) x + (1 - \alpha(x, x_1, x)) z,$$

we can replace x_1 in the lottery by $\alpha(x, x_1, z)x + (1 - \alpha(x, x_1, z))z$. More precisely, the independence axiom says that when we mix x_1 (with probability $\alpha = p_1$) with

$$\left(\frac{p_2}{1 - p_1}, x_2; \ldots; \frac{p_n}{1 - p_1}, x_n \right),$$

we might as well mix $\alpha(x, x_1, z)x + (1 - \alpha(x, x_1, z))z$ with the same lottery (and same $\alpha = p_1$). This gives us a lottery that is equivalent to P but does not use x_1. (It uses x, z, though.)

Continuing this way n times, we replace all the other outcomes by x, z. If we then calculate the probability of x in this lottery, we find that it is precisely

$$\sum_{i=1}^{n} p_i \alpha(x, x_i, z) = \sum_{i=1}^{n} p_i u(x_i),$$

that is, the expected utility of the lottery according to the preceding utility function.

Chapter 5 Probability and Statistics

1. Explain what is wrong with the claim, "Most good chess players are Russian; therefore a Russian is likely to be a good chess player."

Solution As explained in chapter 5 of the main text, this is a classical case of ignoring base probabilities, that is, of confusing the probability of A given B with that of B given A. It is possible that $P(A|B)$ is high while $P(B|A)$ is low.

2. When one sails along the shores of the Mediterranean, it seems that much more of the shoreline has hills and cliffs than one would have thought. One theory is that the Earth was created with the tourism industry in mind. Another is that this is an instance of biased sampling. Explain why.

(Hint: Assume that the Earth is unidimensional and that its surface varies in slope. To be concrete, assume that the surface is made of the segments connecting $((0,0),(90,10))$ and $((90,10),(100,100))$, where the first coordinate denotes distance and the second height. Assume that the height of the water is randomly determined according to a uniform distribution over $[0,100]$. Compare the probability of the shore's being at a point of a steep slope to the probability you get if you sample a point at random (uniformly) on the distance axis.)

Solution The hint here basically gives the answer. Once you draw the curve, if you select a point at random (with a uniform distribution) on the x axis, the steep slope has probability of 10 percent of being chosen. If you select a random point on the y axis (again, using a uniform distribution), you get a probability of 90 percent for the steep slope. Thus, if you look around the Earth from a plane, it seems that mountains are very rare. But if you pour water (presumably a random quantity that generates a uniform distribution over the height of the water surface), you're much more likely to have the water height be at a steep slope.

Similarly, if I spend 11 hours and 50 minutes at home, run for ten minutes to get to my office, spend another 11:50 hours there, and run back, people who see me at a random point on the street might think that I'm running a very hectic lifestyle. But if you randomly sample me over time, you're most likely to conclude that I don't move at all.

3. Comment on the claim, "Some of the greatest achievements in economics are due to people who studied mathematics. Therefore, all economists had better study mathematics first."

Solution Again, this is the same issue. It's possible that the probability of mathematical background given achievements in economics is high, but this doesn't mean that the probability of economic achievements given a mathematical background is also high.

4. Consider exercise 5, following (even if you do not solve it), and explain how many prejudices in the social domain may result from ignoring base probabilities.

Solution Think of a social prejudice, say, associating an ethnic group with a certain characteristic, and ask whether the prejudice might be partly driven by ignoring base probabilities.

5. Trying to understand why people confuse $P(A|B)$ with $P(B|A)$, it is useful to see that qualitatively, if A makes B more likely, it will also be true that B will make A more likely.
a. Show that for any two events A, B,

$$P(A|B) > P(A|B^c)$$

iff

$$P(A|B) > P(A) > P(A|B^c)$$

iff

$$P(B|A) > P(B|A^c)$$

iff

$$P(B|A) > P(B) > P(B|A^c),$$

where A^c is the complement of A. (Assume that all probabilities involved are positive, so that all the conditional probabilities are well defined.)

b. If the proportion of Russians among the good chess players is higher than their proportion overall in the population, what can be said?

Solution

Part 5a Consider the first equivalence,

$$P(A|B) > P(A|B^c)$$

iff

$$P(A|B) > P(A) > P(A|B^c).$$

The second line clearly implies the first. So let us prove the converse: assume the first line, and then derive the second.

Bayes' formula tells us that

$$P(A) = P(A|B)P(B) + P(A|B^c)P(B^c).$$

Denoting $\beta = P(B)$, we have $P(B^c) = 1 - \beta$, and then

$$P(A) = \beta P(A|B) + (1 - \beta)P(A|B^c),$$

with $\beta \in [0, 1]$. That is, the unconditional probability $P(A)$ is a weighted average (with weights $P(B)$, $P(B^c)$) of the two conditional probabilities $P(A|B)$ and $P(A|B^c)$. The weighted average is necessarily between the two extreme points. Moreover, if the two are distinct, say, $P(A|B) > P(A|B^c)$, and if these are well defined (that is, $P(B), P(B^c) > 0$), then $0 < \beta < 1$ and $P(A)$ is strictly larger than $P(A|B^c)$ and strictly smaller than $P(A|B)$.

Next, we wish to show that if $P(A|B) > P(A|B^c)$, then we can reverse the roles of A and B and get also $P(B|A) > P(B|A^c)$. (Clearly, the last equivalence is the same as the first, with the roles of A and B swapped.)

Assume that the probabilities of intersections of A and B and their complements are given by

	B	B^c
A	p	q
A^c	r	s

so that

$$P(A \cap B) = p; \quad P(A \cap B^c) = q,$$

$$P(A^c \cap B) = r; \quad P(A^c \cap B^c) = s,$$

with $p + q + r + s = 1$. For simplicity, assume that $p, q, r, s > 0$.
Then

$$P(A) = p + q; \quad P(A^c) = r + s$$

$$P(B) = p + r; \quad P(B^c) = q + s$$

and

$$P(A|B) = \frac{p}{p+r}; \quad P(A|B^c) = \frac{q}{q+s}$$

$$P(B|A) = \frac{p}{p+q}; \quad P(B|A^c) = \frac{r}{r+s}.$$

The condition

$$P(A|B) > P(A|B^c)$$

is

$$\frac{p}{p+r} > \frac{q}{q+s},$$

and it is equivalent to

$$p(q+s) > q(p+r)$$

or

$$ps > qr,$$

which is equivalent to

$$ps + pr > qr + pr$$

$$p(r+s) > r(p+q)$$

$$\frac{p}{p+q} > \frac{r}{r+s},$$

that is, equivalent to

$P(B|A) > P(B|A^c)$.

Part 5b We have found that if A makes B more likely (that is, more likely than B used to be before we knew A, or equivalently, A makes B more likely than does A^c), the converse is also true. B makes A more likely (than A was before we knew B, or equivalently, B makes A more likely than does B^c).

In this case, if the proportion of Russians among the good chess players is larger than in the population at large, we can say that

- the proportion of Russians among chess players is higher than among the non–chess players;
- the proportion of good chess players among Russians is higher than among non-Russians;
- the proportion of good chess players among Russians is higher than in the population at large.

Importantly, we cannot say anything quantitative that would bring us from $P(A|B)$ to $P(B|A)$ without knowing the ratio $P(B)/P(A)$.

6. Consider a regression line relating the height of children to that of their parents. We know that its slope should be in $(0, 1)$. Now consider the following generation, and observe that the slope should again be in $(0, 1)$. Does this mean that because of regression to the mean, all the population will converge to a single height?

Solution The answer is negative. The regression to the mean is observed when we try to predict a single case, not the average of the population. Indeed, particularly tall parents will have children that are, on average, shorter than they are (but taller than the average in the population), and particularly short parents will have, on average, taller children. These would be the extremes feeding the mean. At the same time, there will be the opposite phenomena. Parents of average height will have children at both extremes. (In particular, a parent with tall genes who happened to have been poorly nourished might be of average height yet still pass on the tall genes.)

Moreover, if one regresses the height of the parents on the height of the children, one is also likely to find a positive correlation, again with the regression to the mean. (Recall that correlation does not imply causation: the parents' height is a cause of the height of the children, not vice versa, but the correlation goes both ways.) If you were to agree that the children's generation would have a lower variance than the parents' generation, you should also endorse the opposite conclusion.

Chapter 6 Aggregation of Preferences

1. In order to determine a unique utility function for each individual, to be used in the summation of utilities across individuals, it was suggested to measure an individual's vNM utility functions (for choice under risk) and to set two arbitrary outcomes to given values (shared across individuals). Discuss this proposal.

Solution The proposal is not without merit. Fixing two outcomes that are considered to be more or less universally agreed-upon values makes sense. Of course, nothing is ever objective. An individual who wishes to commit suicide might prefer death to life, so we can't even agree on what seems like an obvious ranking. Yet, we can hope that this is exceptional. Moreover, we can take a paternalistic point of view and decide to ignore such preferences even if they do exist, ascribing to the person a preference for life over death, or for more money over less money, independently of what he actually prefers.

There are, however, two other difficulties with this proposal. First, it is not clear that the utility function used for describing behavior under risk is the right one for social choice. Assume that one individual is risk-averse and another is risk-neutral. We have to share \$1 between them. Assume that we normalize their utility functions so that they both have $u_i(0) = 0$ and $u_i(1) = 1$. If u_1 is concave and u_2 is linear, the maximization of

$$u_1(x) + u_2(1 - x)$$

will be obtained where $u_1'(x) = 1$. If, for example,

$$u_1(x) = \sqrt{x},$$

we end up giving

$$x = 0.25 > 0.5$$

to individual 1. That is, being risk-averse, this individual gets less of the social resource, and it is not obvious that we would like to endorse this.

Finally, once such a procedure is put into place, we should expect individuals to be strategic about it. If one knows that the responses one gives to vNM questionnaires eventually determine social policy, one may choose to provide untruthful reports (say, pretend to be less risk-averse than one really is) in order to get a larger share of the pie.

2. The Eurovision song contest uses a scoring rule according to which each country ranks the other countries′ songs and gives them scores according to this ranking. It has been claimed that the given scores favor standard songs over more innovative ones. Does this claim make sense? Is it more convincing when the score scale is convex or concave?

Solution If the score scale is convex, say, $1, 2, 4, 8, 16, \ldots$, it is worthwhile to be half the time at the higher end of the scale and the other half at the lower end, as compared to being around the middle all the time. If you have a choice between a risky song, which might be loved by some and abhorred by others, or a less risky one, which is likely not to arouse strong emotions in anyone, you would prefer the riskier song.

By contrast, a concave scale such as $5, 9, 12, 14, 15, \ldots$ generates the opposite incentives for similar reasons.

3. It turns out that for two particular individuals, Pareto domination defines a complete relation. (That is, for every two distinct alternatives, one Pareto-dominates the other.) Assume that

$$u(X) = \{(u_1(x), u_2(x)) \mid x \in X\}$$

is convex. What can you say about the utility functions of these individuals?

Solution First, if the set

$$u(X) = \{(u_1(x), u_2(x)) \mid x \in X\}$$

is convex, it has to be a straight line segment (in $\mathbb{R}^2$, where the first coordinate is $u_1(x)$ and the second is $u_2(x)$). To see this, assume that $u(X)$ is not contained in a segment. Connect two points in $u(X)$. Since the latter is not included in the line defined by these two points, there are points off the line. By convexity, there is an entire nontrivial triangle (with positive area) in $u(X)$. But in such a triangle one can find two points that are not ranked by Pareto domination. Hence $u(X)$ is contained in a segment.

If the segment has a negative slope, there are again points that are not ranked by Pareto domination. Hence we conclude that this segment can be parallel to the x axis, or parallel to the y axis, or it has a finite but positive slope. In all these cases we conclude that the utility function of one individual is a linear function of the other (with the possibility of zero coefficient if the segment is parallel to one of the axes, making one individual indifferent among all alternatives).

4. Assume that individual i has a utility function u_i. For $\alpha = (\alpha_1, \ldots, \alpha_n)$ with $\alpha_i > 0$, let

$$u_\alpha = \sum_{i=1}^{n} \alpha_i u_i.$$

Show that if x maximizes u_α for some α, it is Pareto-efficient.

Solution Assume that x maximizes u_α for some $\alpha > 0$, but suppose, by negation, that x is not Pareto-efficient. Then there exists y such that $u_i(y) \geq u_i(x)$ for all i, with a strict inequality for at least one i, say $i = i_0$. Since we assume that all the coefficients are strictly positive, we know that $\alpha_{i_0} > 0$. This means that

$$u_\alpha(y) > u_\alpha(x),$$

contrary to the assumption that x is a maximizer of u_α.

5. Is it true that every Pareto-efficient alternative maximizes u_α for some α? (Hint: For $n = 2$, consider the feasible sets

$$X_1 = \{(x_1, x_2) \mid \sqrt{x_1} + \sqrt{x_2} \leq 1; \qquad x_1, x_2 \geq 0\}$$

and

$$X_2 = \{(x_1, x_2) \mid x_1^2 + x_2^2 \leq 1; \qquad x_1, x_2 \geq 0\},$$

where $u_i = x_i$.)

Solution The answer is negative, as suggested by the second example in the hint. If the feasible set X were convex, the answer would have been almost positive. To be precise, it would have been positive if we allowed some α_i to be zero. Indeed, for the convex feasible set X_2 all points where $x_1, x_2 > 0$ are optimal for a utility function

$$\alpha_1 u_1 + \alpha_2 u_2,$$

with $\alpha_1, \alpha_2 > 0$, but for the extreme points $(0,1)$, $(1,0)$ we need to set one of the α_i to zero.

However, with a nonconvex set such as X_1, none of the Pareto-efficient points can be described as a maximizer of a utilitarian function with $\alpha_1, \alpha_2 > 0$.

6. Show that under approval voting, it makes sense for each individual to approve of her most preferred alternative(s) and not to approve of the least preferred one(s) (assuming that the voter is not indifferent among all alternatives).

Solution Assume that a is my preferred alternative. Suppose I am about to cast a ballot that approves of some set of alternatives B, which does not contain a. (B may be empty, although under approval voting an empty set is equivalent to abstention.) Next, consider switching from the ballot B to $B \cup \{a\}$, that is, adding the most preferred alternative to the set of approved-of alternatives.

Consider two possibilities: (1) The other voters casts ballots that, together with my vote B, put a among the top alternatives; and (2) the others' votes, together with my B, does not put a at the top. In case 1, a will certainly be at the top; in fact, it will become the unique top alternative if it was not the unique one before. Thus, I only stand to gain from adding a: either it used to be the unique one at the top, and then my vote does not change that, or else it singles out a as the unique one among several that used to be equally popular.

In case 2, a was not among the winners of the vote. Adding it to my ballot B might not change anything or might add a to the set of winners. But in the latter case, it reduces the probability of alternatives I like less than a in favor of a, which is my most preferred alternative. Thus, in this case again I can only gain by switching from B to $B \cup \{a\}$.

Chapter 7 Games and Equilibria

1. Suppose prisoner's dilemma is played T times between two players. Show that playing D is not a dominant strategy but that only Nash equilibria still result in consecutive play of D.

Solution First, let us show that playing D at the first stage is no longer dominant. To see this, imagine that the other player's strategy is to respond to your first move as follows. If you play C in the first stage, he plays C throughout the rest of the game, but if you play D in the first stage, he plays D in the remaining $(T-1)$ stages. If T is large and you are not too impatient, it makes sense to forgo the gain in the first period in order to get the higher payoffs guaranteed by the other player's playing C after that period.

Second, we wish to show that at each Nash equilibrium the players play only D. We reason from the end back to the beginning. Consider the last stage, $t = T$. In this stage there is no future and playing D is dominant. To be precise, whatever the strategy the other player chooses, a player's only best response is to play D (with probability 1) at this last node.

Now consider the penultimate stage, $t = T - 1$. Can it be the case that a player plays C at this stage? The answer is negative. Suppose you decided to play C. Why wouldn't you switch to D, which gives you a strictly higher payoff at stage $t = T - 1$? The only reason can be that this switch will be punished by the other player's reply in the following stage ($t = T$). However, you already know that the other player will play D. Differently put, for there to be a punishment threat, it should be the case that if you do stick to the presumed equilibrium strategy (and play C), you will be rewarded by the other player's playing C (or at least C with a positive probability) in the last stage. But we have concluded that any node that can be reached by the equilibrium strategies is one in which the players play D (with probability 1).

In the same manner we continue and prove, by induction on $k \geq 1$, that at any node that is in stage $T-k, T-k+1, \ldots, T$ and that is reached (with positive probability) by the equilibrium play, the players play D. Applying the conclusion to $k = T - 1$ completes the proof.

2. Consider the following story. In a certain village there are n married couples. It is the case that if one married woman is unfaithful to her husband, all other men are told about it immediately but not the husband. This fact is commonly known in the village. The law of the land is that if a husband knows that his wife has been unfaithful to him, he must shoot her to death on the same night. But he is not allowed to hurt her unless he knows that for sure.

One day a visitor comes to the village, gets everyone to meet in the central square, and says, "There are unfaithful wives in this village." He then leaves.

That night and the following night, nothing happens. On the third night, shots are heard.

a. How many shots were heard on the third night?
b. What information did the visitor add that the village inhabitants did not have before his visit?

Solution

Part 2a There were three shots. The reasoning is as follows. Let k be the number of unfaithful wives. Assume first that $k = 1$, that is, there is exactly one unfaithful wife in the village. In this case, all men apart from her husband know that there are unfaithful wives in this village. But the husband doesn't know whether there are $(k = 1)$ or there aren't $(k = 0)$. Importantly, this husband knows that the other married women are faithful to their husbands because he knows that, were one of them unfaithful, he would know about her. But he knows of none, and he knows that he knows of none. So he can conclude that the other women are faithful. Hearing the news that some women are not faithful $(k \geq 1)$ proves to him that his wife isn't faithful to him, and he will kill her on the first night.

Next, assume that there are exactly two unfaithful wives $(k = 2)$, them A and B. The husband of each knows that there are some $(k \geq 1)$ unfaithful wives because he knows for sure that the other wife (not his wife) is unfaithful. That is, A's husband knows that B is unfaithful but doesn't know whether A is, and B's husband knows that A is unfaithful but doesn't know whether B is. Hence, the husbands of both A and B are not too excited when they hear that there are unfaithful wives in this village. Each should say to himself, "Well, I don't know about my wife, but the fact that *some* wives are unfaithful is not news to me." However, A's husband should also reason as follows: "If my wife, A, were faithful to me, then B would be the only unfaithful wife in the village (that is, $k = 1$). In this case, by the reasoning for the case $k = 1$, B's husband just learned that B is unfaithful to him, and he'll shoot her tonight." Anticipating the prospect of a sensational killing, A's husband goes to sleep. In the morning, he is awakened by the birds chirping rather than by the sound of a shot. And then he must reason, "B's husband didn't shoot her last night, so that means he's not sure that she's unfaithful. Therefore, he already knew there had been some unfaithful wives, that $k \geq 1$. But, not knowing about B herself, he could only have known that my wife, A, has been unfaithful to me." Equipped with this sad conclusion, A's husband waits until night falls and then shoots his wife. By the same reasoning, so does B's husband.

Similarly, one can prove by induction that if there are exactly k unfaithful wives, then all their husbands will know for sure that their wives are unfaithful on the kth night and shoot them on that night. So, if the shots were heard on the third night, there were exactly three unfaithful wives in the village.

Part 2b The information added by the visitor was not "there are unfaithful wives in this village." Indeed, with $k \geq 2$, all husbands know that there are some unfaithful wives. The additional piece of information was that "there are unfaithful wives in this village" is common knowledge. That is, by making a public declaration, the visitor ensured that an already known fact was also commonly known.

To see this, observe that if $k = 1$, everyone but one husband knows the proposition $p =$ "there are unfaithful wives in this village." In this case, this husband does learn something from the declaration. If, however, $k = 2$, everyone knows p, but it is not true that everyone knows that everyone knows p. As analyzed, if the only two unfaithful wives are A and B, then A's husband knows p, and he also knows that all other husbands apart from B's husband know p, but he does not know that B's husband knows p. As far as he knows, his wife, A, may be faithful, and then B's husband would not know whether p is true or not (whether B is faithful or not). Similarly, if $k = 3$, everyone knows p, and everyone knows that everyone knows p, but it is not true that (everyone knows that)^3 p. Thus, for any k, some hierarchy of knowledge is missing, and this level is what the visitor adds by his public announcement.

3. Consider an extensive form game, and show how a player might falsify common knowledge of rationality (by deviating from the backward induction solution). Show an example in which it may be in the player's best interest to do so.

Solution It will be helpful to draw a game tree. Consider a game in which player I can choose to play down (D) and end the game with payoffs $(11, 5)$, or to play across (A). If she plays A, it's player II's turn. He can choose to play down (d) and end the game with payoffs $(x, 9)$, or to play across (a), in which case it's player I's turn again. In the last stage, player I has to choose between down (δ) with payoffs $(9, 0)$ and across (α) with payoffs $(10, 10)$.

The backward induction solution is as follows. At the last node player I would play across (α) because $10 > 9$. Given that, at the second node player II should play across (a) because he gets 10 by the backward induction solution if he continues and only 9 if he stops the game. Given that, we conclude that at the first node player I should play down (D) because this guarantees 11 and the backward induction analysis says she will get only 10 if she plays across.

However, what should player II think if he finds himself playing? The backward induction solution says that he will not have to play at all. So there is something wrong in the assumptions underlying the backward induction solution. What is it? We don't know. Maybe player I is not rational? Maybe she's crazy? In this case, can player II trust that she will indeed prefer 10 to 9 at the last stage? Maybe she won't, and then player II will get only 0? So perhaps it is safer for player II to play down (d), guaranteeing 9, rather than taking a lottery with outcomes 10 and 0 with the unknown probability that player I is rational?

Indeed, if x is rather low, this would make a lot of sense. But what happens if $x = 15$? In this case, this payoff is the best outcome for player I throughout the game. In this case it is in player I's interest to sow doubt about her own rationality in player II's mind. If player II is sure that player I is rational, he will play across. But if player I manages to convince player II that she is crazy, she will be better off. But then again, perhaps player II will see through this ruse and not be scared? Maybe he'll conclude, "Oh, I know the game she's playing. She is trying to scare me in order to get the best payoff for herself. But I will not be tricked. I'll play across, and I'm sure that when it's her choice in the final node, she'll be rational.... Or will she?"

Indeed, it's not clear how players revise their theory of the game (and of the other players' rationality) in such situations. We can see such examples in real life, for instance, political situations where one may be better off if others think one is crazy, but pretending to be crazy is not easy if the motives for doing so are too transparent.

4. Compute the mixed strategies equilibria in the following games (see section 7.3 of main text):

Game 6. Pure Coordination 1

	R	L
R	$(1,1)$	$(0,0)$
L	$(0,0)$	$(1,1)$

Game 7. Pure Coordination 2

	A	B
A	$(3,3)$	$(0,0)$
B	$(0,0)$	$(1,1)$

Game 8. Battle of the Sexes

	Ballet	Boxing
Ballet	$(2,1)$	$(0,0)$
Boxing	$(0,0)$	$(1,2)$

Solution Consider game 6 first. Assume that player I plays R with probability p and L with probability $(1-p)$. Assume that player II plays R with probability q and L with probability $(1-q)$. If player I uses a truly mixed strategy, that is, if $0 < p < 1$, it has to be the case that the expected utility she gets from both pure strategies is the same. To see this, observe that the expected utility is linear in p:

$$EU^1((p,1-p)) = pEU^1((1,0)) + (1-p)EU^1((0,1)).$$

If the expected utility from playing R, $EU^1((1,0))$ were higher than the expected utility from playing L, $EU^1((0,1))$, the only optimal response for player I would be $p = 1$. Conversely, if $EU^1((1,0)) < EU^1((0,1))$, the only optimal response would be $p = 0$. Hence the only way that $p \in (0,1)$ can be optimal is if

$$EU^1((1,0)) = EU^1((0,1).$$

In this case player I is completely indifferent between playing $(p,1-p)$ and playing $(1,0)$, $(0,1)$ or any other mixed strategy. This may sound a little weird, and indeed some people are not completely convinced by the concept of mixed strategy Nash equilibria in games that are not zero-sum (where there exist other justifications of the concept). But let's acknowledge these doubts and move on.

Given that player II plays $(q,1-q)$, we can compute these expected utilities:

$$EU^1((1,0)) = q \times 1 + (1-q) \times 0 = q,$$

and

$$EU^1((0,1)) = q \times 0 + (1-q) \times 1 = 1-q,$$

and the equation $EU^1((1,0)) = EU^1((0,1))$ means that $q = 1 - q$, or $q = 0.5$. The same calculation applies to player II, given that player I plays $(p, 1-p)$, and it yields $p = 0.5$.

For game 7 the same type of calculations (with the same notation for p and q, though the names of the pure strategies are different) yield

$$EU^1((1,0)) = q \times 3 + (1-q) \times 0 = 3q,$$

and

$$EU^1((0,1)) = q \times 0 + (1-q) \times 1 = 1 - q,$$

and the equation $EU^1((1,0)) = EU^1((0,1))$ implies $3q = 1 - q$, or $q = 0.25$.

Similarly, we get $p = 0.25$.

In game 8 we have (again, with the same meaning of p and q)

$$EU^1((1,0)) = q \times 2 + (1-q) \times 0 = 2q,$$

$$EU^1((0,1)) = q \times 0 + (1-q) \times 1 = 1 - q,$$

and $2q = 1 - q$, or $q = \frac{1}{3}$, but for player II we get

$$EU^2((1,0)) = p \times 1 + (1-p) \times 0 = p,$$

$$EU^2((0,1)) = p \times 0 + (1-p) \times 2 = 2(1-p),$$

and $p = 2(1-p)$, or $p = \frac{2}{3}$. That is, each player chooses the strategy that corresponds to his/her preferred equilibrium with probability $\frac{2}{3}$.

5. Show that a 2×2 game in which all payoffs are different cannot have precisely two Nash equilibria.

Solution Let there be a game

	L	R
T	a, α	b, β
B	c, γ	d, δ

Since all payoffs are different, we may assume without loss of generality that $a > c$. Otherwise, $c > a$, and we can rename the strategies to make a the higher payoff.

Consider b and d. If $b > d$, then strategy T strictly dominates strategy B for player I. In this case, in each equilibrium player I will play T with probability 1. And then the only equilibrium will be obtained when player II plays L (with probability 1) if $\alpha > \beta$, or R (with probability 1) if $\alpha < \beta$. That is, if $b > d$, the game has a unique equilibrium in pure strategies and no equilibria in mixed strategies. The number of equilibria is then one.

Next, consider the case in which $b < d$.

Recall that the vNM utility functions are given up to multiplication by a positive constant and an addition of a constant. In fact, if we only consider this particular game, we can also add an arbitrary constant to the payoffs of player I in each column and an arbitrary constant to the payoffs of player II in each row. (Such a shift of the utility function in a given column for player I or in a given row for player II does not change the best response set. A strategy is a best response for a player after such a shift if and only if it used to be a best response before the shift.)

Hence, we can assume without loss of generality that $c = 0$ (by subtracting c from player I's payoffs in column L) and that $b = 0$ (by subtracting b from player I's payoffs in column R) and obtain the game

	L	R
T	a, α	$0, \beta$
B	$0, \gamma$	d, δ

with $a, d > 0$. (Technically speaking, it is now no longer true that all payoffs are different, but what is important is that the payoffs can be compared by a given player who considers switching a strategy. That there are two zeros in this game does not change the fact that there are no indifferences when players compare their payoffs, given different choices of their own but the same choice of the other.)

We now turn to consider player II's payoffs. If $\alpha < \beta$ and $\gamma < \delta$, then R is a strictly dominant strategy for player II and the unique equilibrium is (B, R). Similarly, if $\alpha > \beta$ and $\gamma > \delta$, then L is a dominant strategy and the unique equilibrium is (T, L). Thus we are left with the interesting case in which player II does not have a dominant strategy either. This means that either $\alpha < \beta$ and $\gamma > \delta$, or $\alpha > \beta$ and $\gamma < \delta$.

Note that these cases are no longer symmetric. If one switches the names of the columns, one changes some of the assumptions about player I's payoffs.

We may still simplify notation by assuming that at least one zero appears among player II's payoffs in each row. We can decide, for instance, that $\beta = 0$ and consider the cases in which α is positive or negative. Or we may choose to work only with non-negative payoffs, and set a different parameter to zero each time. Let's do this, one case at a time.

In case 1 we may assume without loss of generality that $\alpha = \delta = 0$ and we get the game

	L	R
T	$a, 0$	$0, \beta$
B	$0, \gamma$	$d, 0$

with $a, d, \beta, \gamma > 0$. In this case there is no pure strategy Nash equilibrium. An equilibrium in mixed strategies $((p, 1-p), (q, 1-q))$ will have to satisfy

$$qa = (1-q)d,$$

$$(1-p)\gamma = p\beta,$$

that is,

$$\left(\left(\frac{\gamma}{\beta+\gamma}, \frac{\beta}{\beta+\gamma}\right), \left(\frac{d}{a+d}, \frac{a}{a+d}\right)\right)$$

is the unique mixed strategy Nash equilibrium, and the unique Nash equilibrium overall.

In case 2 we may assume that $\beta = \gamma = 0$ and the game is as follows:

	L	R
T	a, α	$0, 0$
B	$0, 0$	d, δ

with $a, d, \alpha, \delta > 0$.

In this case, both (T, L) and (B, R) are pure strategy Nash equilibria. Are there any mixed ones? If $((p, 1-p), (q, 1-q))$ is a Nash equilibrium in mixed strategies, it will have to satisfy

$$qa = (1-q)d,$$

$$p\alpha = (1-p)\delta.$$

Indeed,

$$\left(\left(\frac{\delta}{\alpha+\delta},\frac{\alpha}{\alpha+\delta}\right),\left(\frac{d}{a+d},\frac{a}{a+d}\right)\right)$$

is a mixed strategy Nash equilibrium. Overall, there are three equilibria in the game.

To conclude, if one of the players has a dominant strategy, the game will have a unique Nash equilibrium, and it will be pure. Otherwise, there might be a unique Nash equilibrium in mixed strategies (if the game is of the type of matching pennies, or three equilibria of which two are pure (if the game is a coordination game or a game of battle of the sexes).

6. A computer sends a message to another computer, and it is commonly known that the message never gets lost and that it takes 60 seconds to arrive. When it arrives, it is common knowledge (between the two computers) that the message has indeed been sent and has arrived. Next, a technological improvement was introduced, and the message can now take any length of time between 0 and 60 seconds to arrive. How long after the message was sent will it be commonly known that it has been sent?

Solution Suppose that the message was sent at time 0 (measured in seconds) and arrived at time t, $0 \leq t \leq 60$. At time t, the receiver knows that the message has arrived. Does the sender know that it has arrived? If $t = 60$, the sender will know that the message has arrived because 60 seconds is the upper bound on the transmission time. But if $t < 60$, the receiver will know it sooner, but not the sender. The sender will have to wait until the 60th second to know that the message did indeed arrive.

When will the receiver know (for sure) that the sender knows (for sure) that the message has arrived? The receiver knows about the analysis in the previous paragraph, so he knows that the sender is going to wait 60 seconds from the time she sent the message until she can surely say that the message has arrived. When was the message sent? The receiver can't know for sure. Getting the message at time t, he has to consider various possibilities. It might have been a quick transmission, sent at t and arriving immediately, or a sluggish one, sent at $t - 60$ and taking the maximal length of time, 60 seconds. When will the receiver know that the sender knows that the message has been sent? The receiver will have to wait 60 seconds after transmission time, which is somewhere between $t - 60$ and t. The maximum is obtained at t. That is, after having received the message, the receiver has to wait another 60 seconds to know for sure that the sender knows for sure that the message has arrived.

When will the sender know that the receiver knows that the sender knows that the message has arrived? She knows the previous analysis, that is, she knows that the receiver has to wait 60 seconds from the time that the message has actually arrived until he (the receiver) knows that she (the sender) knows that it has arrived. Sending the message at time 0, she has to consider the maximal t, that is, $t = 60$, and add to it another 60 seconds, and then, only at $t = 120$, can she say that she knows that he knows that she knows that the message has arrived.

And when will the receiver know that the sender knows that the receiver knows that the sender knows that the message has arrived? The receiver has to wait 120 seconds from the time the message has been sent, which means, $120 + 60 = 180$ seconds from the time he received it. Taking this into account, the sender knows that she has to wait $180 + 60 = 240$ seconds from the time of transmission until she knows that he knows that.... In short, the fact that the message has arrived will never be common knowledge.

Chapter 8 Free Markets

1. Discuss the reasons that equilibria might not be efficient in the following cases:

a. A physician should prescribe tests for a patient.
b. A lawyer assesses the probability of success of a legal battle.
c. A teacher is hired to teach a child.

Solution All these cases are examples of principal agent problems with incomplete information. A physician is an expert hired by the patient (directly or indirectly). The physician knows more than the patient does about the patient's condition, possible treatments, and so on. Consider a test that the physician might prescribe, which is very expensive or unpleasant. If he bears no part of the cost, he might be overly cautious and prescribe the test simply because he would feel more comfortable with the additional information. The patient may prefer to forgo the test and avoid the cost or pain involved, but she does not have the information to make this decision. If, however, the physician does bear some of the cost, say, he has a budget for tests, then he has an incentive to save money even if the test is necessary.

Again, the patient can't directly check whether the physician's recommendation is the same recommendation the patient would have arrived at, given the information. Thus, in both systems, equilibria may not be Pareto-efficient.

Similar problems arise when the lawyer, an expert, has to advise the client whether to pursue a legal battle. If the only costs and benefits involved are monetary, it is possible to agree on a fee that is proportional to the client's outcome and thus to align the interests of the informed agent (the lawyer) with the uninformed principal (the client). But since there are other costs (such as the psychological cost of uncertainty), one may again find that equilibria are inefficient.

Finally, in education we find a double agent problem. The parent hires the teacher to teach, but both the child and the teacher would prefer to tell each other jokes rather than work. The parent may condition the teacher's compensation on the child's performance on a certain test, but it's hard to disentangle the child's talent and the teacher's efforts and to make the compensation proportional to the latter. Again, inefficiency is to be expected.

2. The dean has to decide whether to give a department an overall budget for its activities or split the budget among several activities such as conferences, visitors, and so forth. Discuss the pros and cons of the two options.

Solution The argument for an overall budget is the classical argument for free markets. Rather than a central planner, who dictates the details of economic activities, the free market intuition suggests that we decentralize the decision-making process. Thus, the dean might say, "Who am I to judge what's the best trade-off between inviting visitors and going to conferences? Let the department make these choices. I should trust that the department knows best how useful conferences are, which ones should be attended, which visitors should be invited, and so on."

However, this free market intuition should be qualified. First, there is a problem of incomplete information, as in any principal agent problem. The principal may not know whether the faculty members go to a conference on a charming Mediterranean island because it's the most important conference in the field or because its location is nice. Since the faculty's payoff is not precisely aligned with the school's, it's also not clear whether the right trade-off has been struck between traveling and inviting visitors, and whether the choice of visitors was perfectly objective, and so on.

Besides this, there may be problems of externalities involved. For example, inviting visitors may benefit other departments, and this externality may not be internalized by the department making the decision.

3. Consider the student course assignment problem described in section 8.4 of the main text. Show that for every n it is possible to have examples in which n is the minimal number of students that can find a Pareto-improving reallocation of courses.

Solution Let there be n students and n courses, denoted $\{a_1, \ldots, a_n\}$. Consider the preferences shown in the following table. Each column designates a student, and the courses in that column are listed from top (most preferred) to bottom (least preferred):

1	2	3	...	n
a_1	a_2	a_3	...	a_n
a_2	a_3	a_4		a_1
a_3	a_4	a_5		a_2
...	...			
a_n	a_1	a_2		a_{n-1}

That is, the preferences of individual i are obtained from those of individual $i-1$ by taking the best alternative in the eyes of $i-1$ and moving it to the bottom without changing the ranking of the other pairs of alternatives.

Now assume that the allocation is such that each individual has her second-best choice. That is, 1 has a_2, 2 has a_3, and so on (with a_1 in the hands of individual n). Clearly, there is a Pareto-improving trade by which each gets her most preferred alternative instead of her second most preferred. However, no proper subset of the individuals can obtain a Pareto-improving trade. To see this, assume that a particular individual is not among the traders. Without loss of generality, assume that this is individual n. In this case, individual 1 cannot get a_1, which is the only alternative she is willing to trade for what she has, namely, a_2. This means that individual 1 will also not be part of the trade. This, in turn, means that individual 2 cannot be convinced to give up her current holding, a_3, and so on.