Binary-Choice Constraints
and Random Utility Indicators

JACOB MARSHAK
Yale University

1. The Problem

This paper contains suggestions for combining various available types of
information on economic choice. There are in the main three types.

(i) Controlled experiments on choices from small sets (usually pairs) of
alternatives.

(ii) Surveys of consumers’ choices from large sets of alternatives, each
set being determined by the market prices and by each consumer’s monetary
resources; similar surveys (mutatis mutandis) of producers’ choices are also
available, especially in the case of farmers.

(iii) Time series of total consumption (or production) of individual
commodities, total incomes, and market prices, over a given geographical area.

Various researchers (Marschak [16], and more recently Tobin [19], Wold
[22], and Farrell [9]) have tried to combine information on (ii) and (iii). The
present paper deals with combining the materials in (i) and (ii), albeit only
to the extent of outlining the general problem and of concluding with an
example.

Suppose that, given a set of consumers’ budgets at varying prices, we
want to estimate a function on the commodity space that might be used as
a utility indicator (defining a system of “indifference surfaces”) to describe
the tastes of the sample group of people. The desired statistical estimation
will usually be preceded by the “specification” of a class of eligible functions
on the commodity space. The continuity and convexity (“diminishing
marginal satisfaction”) properties postulated in economic theory, partly in
connection with its requirement of “consumer’s equilibrium,” permit such
specification only within very broad limits.

Dedicated to Ragnar Frisch, an old and esteemed friend, on the occasion of his sixty-
fifth birthday.

The paper was completed during a year’s visit at the Carnegie Institute of Technology.
The research was carried out partly under a contract of the Office of Naval Research
with the Cowles Foundation for Research in Economics at Yale University, Task
NR047-006. Previous collaboration with H. D. Block and discussions with Herman
Chernoff, Gerard Debreu, Morris DeGroot, Tjalling Koopmans, and R. Duncan Luce were
most helpful.
Suppose, however, that, in addition to the consumers' survey data, laboratory data are available that imply that the choices being made by people of a given socio-economic group obey certain constraints. The knowledge of these constraints might provide help in further restricting the set of eligible utility indicators.

It will be realistic to assume that a man's choice from a given set of alternatives is not unique but obeys some probability distribution. This agrees with the approach usual in psychometrics and psychophysics, and may also explain a part of the statistical variations in economic survey data. Accordingly, our approach will be "stochastic" throughout. We shall study the logical connections between different behavior assumptions made in stochastic theories of decision.

2. Binary-Choice Probabilities

Unless specified to the contrary, all concepts in this paper are associated with a single given person, the "subject" of experiments. The only experiments considered in this paper—except for cursory remarks on other kinds of experiments, especially in Sections 4, 5, and 11—are binary choices: the subject is forced to choose one of the pair \((x, y)\) of alternatives. On the other hand, the data resulting from a consumer survey are, of course, multiple choices since a consumer's monetary resources permit him, given the prices of goods, to choose from among more than a pair of alternative budgets.

The following notation will be used:

- \(X\), the set of alternatives, with generic elements \(x, y, \ldots\).
- \(n\), the number of elements of \(X\) when \(X\) is finite; in this case the elements will be identified by integers: \(X = \{1, \ldots, n\}\).
- \(p\), the class of functions \(p\) on \(X \times X\) (or, with \(X\) finite, the set of matrices \(p = [p_{xy}]\)) such that
  \[
  p_{xy} = 1 - p_{yx} \geq 0.
  \]

Hence \(p_{xx} = \frac{1}{2}\). When \(x \neq y\), the function \(p_{xy}\) is called the probability of a binary choice, viz., of the choice of \(x\) out of the subset \((x, y)\) of \(X\). A region \(p_{c} \subseteq p\) defines a condition \((C)\) on all \(p_{xy}\) and is called a binary-choice constraint. It will be convenient to identify such constraints, and other conditions, by symbols in parentheses, thus: \((C)\).

Binary-choice constraints involving no other quantities but the \(p_{xy}\) are called directly testable: it will be assumed that they can be accepted or rejected, in the sense of statistical inference, on the basis of binary-choice experiments. (However, except in the case of finite \(X\) and very large samples permitting the approximation of each \(p_{xy}\) by the corresponding observed frequency, difficult statistical problems that cannot be treated here present themselves, but see [1, Section 10].) Constraints on the \(p_{xy}\) that are not "directly testable" involve in particular the existence of the variously defined "utilities," a theoretical construct. It will be important to find which of the
"directly testable" constraints imply or are implied by the existence of "utilities" in the various senses of this term.

**Example.** Let \( X = (1, 2, 3) \). A function \( p \) is completely described by any six numbers \( p_{xy} \) obeying (1). The relation \( 1 \leq p_{12} + p_{23} + p_{31} \leq 2 \) exemplifies a binary-choice constraint. It is "directly testable" and is implied by certain "utility" statements, as will be seen in Section 6.

3. **Random Utility Indicator**

A probability measure on the set of real-valued functions on \( X \) defines a real-valued random function on \( X \) (a random vector if \( X \) is finite); its values are random numbers.

**Definition.** \( U^{(2)} \), a real-valued random function on \( X \), is called a random utility indicator “in the binary sense” if, for every \( x \neq y \),

\[
\text{Pr} (U^{(2)}_{x} \geq U^{(2)}_{y}) = p_{xy}
\]

(The idea goes back to Thurstone’s “law of comparative judgment” [18].)

It follows that if \( F \) is a random utility indicator in the binary sense, and \( G \) is a strictly increasing transform of \( F \) (that is, \( F_x \geq F_y \) if and only if \( G_x \geq G_y \), for all \( x, y \)), then \( G \) is also a random utility indicator in the binary sense.

The following theorem was suggested by M. DeGroot:

**Theorem 1.** If \( U^{(2)} \) is a random utility indicator in the binary sense, then \( \text{Pr} (U^{(2)}_{x} = U^{(2)}_{y}) = 0 \) for every \( x \neq y \).

**Proof.** Let \( x \neq y \) and suppose \( \text{Pr} (U^{(2)}_{x} = U^{(2)}_{y}) > 0 \). Then

\[
P_{xy} + P_{yx} = \text{Pr} (U^{(2)}_{x} \geq U^{(2)}_{y}) + \text{Pr} (U^{(2)}_{y} \geq U^{(2)}_{x})
= \text{Pr} (U^{(2)}_{x} > U^{(2)}_{y}) + \text{Pr} (U^{(2)}_{y} > U^{(2)}_{x}) + 2 \text{Pr} (U^{(2)}_{x} = U^{(2)}_{y})
= \text{Pr} (U^{(2)}_{x} > U^{(2)}_{y}) + \text{Pr} (U^{(2)}_{y} > U^{(2)}_{x}) + \text{Pr} (U^{(2)}_{x} = U^{(2)}_{y}) + \text{Pr} (U^{(2)}_{y} = U^{(2)}_{x})
= 1 + \text{Pr} (U^{(2)}_{x} = U^{(2)}_{y}) > 1,
\]

contradicting (1).

**Examples.** (i) Let \( X = (1, 2, 3) \); then \( U^{(2)} \) is a random utility indicator if the six numbers \( \text{Pr} (U^{(2)}_{x} > U^{(2)}_{y}) \) are respectively equal to the six numbers \( p_{xy} \) which (or some constraints on which) are obtained from observations.

(ii) Let \( X \) denote the real \( m \)-space with generic point \( x = (x_1, \ldots, x_m) \); let \( T = (T_1, \ldots, T_m) \) be a random vector such that \( 0 < T_i < 1 \) (\( i = 1, \ldots, m \)) and \( \Sigma T_i = 1 \). Let

\[
U^{(2)}_{x} = \prod_{i=1}^{m} x_i^{T_i}
\]

(the "Cobb-Douglas function" frequently used in economics, with \( X \) being the commodity space). If \( \text{Pr} (U^{(2)}_{x} \geq U^{(2)}_{y}) = p_{xy} \), then \( U^{(2)} \) is a random utility
CHOICE CONSTRAINTS AND UTILITY INDICATORS

indicator "in the binary sense." This example will be used again in Section 12.

**Condition (U(3)).** The function \( p \) is such that \( U(3) \) exists.

4. **A Remark on Multiple-Choice Probabilities**

The concepts and the condition just defined can be easily generalized to multiple-choice probabilities. For every \( x \in X^* \) and \( X^* \subseteq X \), one defines \( p_e(X^*) \), obeying

\[
p_e(X^*) \geq 0, \quad \sum_{x \in X^*} p_e(X^*) = 1.
\]

The number \( p_e(X^*) \) is interpreted as the probability that the subject, forced to choose from the subset \( X^* \), will choose \( x \). Clearly, \( p_{xy} \) is a special case, with \( X^* = (x, y) \). The following condition exemplifies an important and plausible constraint that would be directly testable if multiple choices from subsets of various sizes were observable:

\[
\text{if } x \in X^{**} \subseteq X^* \subseteq X, \text{ then } p_e(X^*) \leq p_e(X^{**}).
\]

In [1], Block and Marschak dealt with multiple-choice probabilities, and had accordingly used the following definition of the random utility indicator, stronger than the one given in Section 3:

**Definition.** \( U \), a random function on \( X \), is a random utility indicator if, for all \( X^* \subseteq X \),

\[
\Pr (U_e \geq U_y, \ y \in X^*) = p_e(X^*).
\]

Correspondingly, Condition (U(3)) is replaced by the stronger

**Condition (U).** The set of probabilities \( p_e(X^*) \), where \( x \in X^* \) and \( X^* \subseteq X \), is such that \( U \), in the sense of (6), exists.

Clearly, Condition (U) implies (U(3)) and also implies (5).

Some other properties of \( U \) will be used later, in Section 11. However, because of experimental difficulties with large subsets of alternatives, it seems worthwhile to devote a special and intensive study to binary-choice probabilities. They cover much of the existing materials on human responses, and many practical situations ("pairwise comparisons"). Moreover, the following needs no proof:

**Theorem 2.** If the \( p_{xy} \) admit the existence of a function \( U(3) \) satisfying Condition (U(3)), then, for every \( x \in X^* \) and \( X^* \subseteq X \), there exists a number \( q_e(X^*) \geq 0 \) such that, putting \( p_e(X^*) = q_e(X^*) \) and \( U = U(3) \), condition (6) is satisfied.

That is, Condition (U(3)) guarantees the existence of numbers which, if they were the observed multiple-choice probabilities, would satisfy Condition (U).

Except in Section 11, it will not be assumed in this paper that constraints on non-binary-choice probabilities—such as (5)—can be inferred from observations.
5. Probability of Ranking

Similarly, and possibly on still stronger grounds, no reference will be made in this paper to experiments in which the subject is requested, not to make an actual choice but to rank three or more alternatives according to his preferences (with two alternatives, choosing and ranking may be regarded as identical). It will not be assumed that responses obtained in such experiments are in any sense consistent with choices of single alternatives from subsets. To be sure, we shall use the term “probability of ranking,” but not to denote a number that is estimated (or, more generally, whose properties are inferred) from experiments on ranking. Rather, we give the following definition:

**Definition.** If \(U^{(x)}\) is a random utility indicator, then, for any \(x_1, x_2, \ldots, x_m\) in \(X\), the number

\[
P(x_1, x_2, \ldots, x_m) = \Pr(U_{x_1}^{(x)} > U_{x_2}^{(x)} > \cdots > U_{x_m}^{(x)})
\]

is called the probability of the ranking (permutation) \(x_1, x_2, \ldots, x_m\). It follows that \(P(xy) = p_{xy}\).

**Theorem 3.** If \((U^{(x)})\) is satisfied and \(X^* = M = (1, \ldots, m)\) is a finite subset of \(X\), then for any \(i, j \in M\),

\[
p_{ij} = \sum_{r \in R_{ij}} P(r),
\]

where \(R_{ij}\) is the set of all rankings \(r\) on \(M\) in which \(i\) precedes \(j\).

**Proof.** By (2), (7).

**Theorem 4.** \((U^{(x)})\) is satisfied for \(X\) finite, \(X = N = (1, \ldots, n)\), if there is a non-negative function \(P\) on the set of rankings \(r\) on \(N\) such that equation (8) holds, provided \(R_{ij}\) is redefined as the set of all rankings on \(N\) in which \(i\) precedes \(j\).

**Proof.** Denote by \(r = 1, 2, \ldots, n\) the ranking in which the \(i\)th place is occupied by \(i\). Then, by (7)

\[
P(r) = P(1, 2, \ldots, n) = \Pr(U_{1}^{(r)} > U_{2}^{(r)} > \cdots > U_{n}^{(r)}),
\]

and by (8) the condition (2) is satisfied.

6. A Condition for the Existence of a Random Utility Indicator

**Condition (d) ["triangular"].** For all \(x, y, z\) in \(X\),

\[
p_{xy} + p_{yz} \geq p_{xz}.
\]

By virtue of (1), this can be rewritten more symmetrically as

\[
1 \leq p_{xy} + p_{yz} + p_{zx} \leq 2.
\]

Define

\[
\Delta_{xyz} = p_{xy} + p_{yz} + p_{zx} - 1;
\]
then by (1), $A_{xy} + A_{yx} = 1$; and (9a) can be rewritten as

$$0 \leq A_{xy} \leq 1.$$  

Condition (d) was probably first noticed by Guilbaud [13].

**Theorem 5.** $(U^{(3)})$ implies (d). [And therefore, clearly, $(U)$ also implies (d).]

**Proof.** Denote by $x, y, z$ the generic elements of the subset $(1, 2, 3)$ of $X$. Then, by the definition (7), the six $P(xyz)$ add up to 1, and each is greater than or equal to 0. If $(U^{(3)})$ is true, then, by Theorem 3,

$$p_{12} = P(132) + P(123) + P(312),$$

$$p_{23} = P(123) + P(213) + P(231),$$


(11)

In (11) the sum of the left sides is $A_{13} + 1$; on the right side, the three off-diagonal terms of the symmetric matrix add up to $s$, say, where $0 \leq s \leq 1$; and all nine terms add up to 1. Hence

$$0 \leq A_{123} = s = P(123) + P(231) + P(312) \leq 1,$$

and (9b) is satisfied.

**Theorem 6.** If $X = (1, 2, 3)$, then (d) implies $(U^{(3)})$.

**Proof.** Denote by $x, y, z$ the generic elements of $X$. We shall show that, for any six numbers $p_{xy}$ satisfying (1) and such that $0 \leq A_{123} = p_{12} + p_{23} + p_{31} - 1$, one can find six numbers $P(xyz)$, non-negative, adding up to 1, and satisfying (11). We have just seen that if $\sum_{xy} P(xyz) = 1$, then (11) implies

$$A_{123} = P(123) + P(231) + P(312).$$

Subtracting (12) from each of the equations (11), we obtain

$$p_{12} - A_{123} = P(132) - P(123),$$

$$p_{23} - A_{123} = P(213) - P(231),$$

$$p_{31} - A_{123} = P(321) - P(312).$$

(13)

Put $0 = P(231)$ or $P(132)$ according as $p_{12} \geq p_{23}$ or $\leq A_{123}$; put $0 = P(312)$ or $P(213)$ according as $p_{23} \geq p_{31} \leq A_{123}$. Substituting into (13) and (12), one obtains the remaining four $P(xyz)$, non-negative and adding up to 1. Hence, by Theorem 4, $(U^{(3)})$ is satisfied.

An alternative proof of Theorem 5. A probability distribution on the set of rankings on $X = (1, \ldots, n)$ is a vector $P$ whose components $P(r^k)$, where $k = 1, \ldots, n!$, are non-negative and add up to 1. The set $P$ of all such vectors is convex and has $n! - 1$ dimensions and $n!$ vertices. The vertex $P_k$ shall represent the distribution in which $P(r^k) = 1$. Suppose $(U^{(3)})$ holds and let $[p_{xy}(P)]$ be the $n \times n$ matrix of binary probabilities generated by $P$ in $P$. Consider the vertex $P_k$ of $P$ and denote by $r_x^k$ the rank assigned to $x$ when the ranking is $r^k$. Then, for $x \neq y$,

---

1 See also [1], Theorems 5.7, 5.1.
Let \( x, y, z \) be all distinct. Since \( r_x^z < r_y^z \) and \( r_y^z < r_x^z \) imply \( r_x^z > r_y^z \), the triple \( \{p_x(z), p_y(x), p_z(y)\} \) cannot consist of 1's only or of 0's only. Hence the sum of these three probabilities, for every ranking \( r^k \), is either 1 or 2. Now, every vector \( P \) is a convex combination of the unit vectors \( P_k \):

\[
P = \sum P_k \cdot P(r^k) \quad \left( \sum P(r^k) = 1; P(r^k) \geq 0; k = 1, \ldots, n! \right)
\]

and

\[
p_{xy}(P) = \sum p_{xy}(P_k) \cdot P(r^k).
\]

Hence every sum \( p_{xy}(P) + p_{yx}(P) + p_{zx}(P) \) is a weighted average of the numbers 1 and 2, so that \((d)\) is satisfied.

This proof was obtained after a conversation with Herman Chernoff, who had suggested the following way to prove (or disprove) the converse of Theorem 5.

**Converse of Theorem 5.** Denote by \( S_P \) the set of all matrices \( [p_{xy}] \) generated by the distributions \( P \) in \( P \); denote by \( S_n \) the set of all matrices \( [p_{xy}] \) satisfying \((d)\). Each of these sets has \( n(n - 1)/2 \) dimensions and is convex. \( S_P \) has \( n! \) vertices, each corresponding to precisely one vertex of \( P \), and such that for \( x, y, z \) distinct, \( p_{xy} = 1 \) or 0, and \( p_{xy} + p_{yx} + p_{zx} = 1 \) or 2. Hence all vertices of \( S_P \) are elements and are, in fact vertices, of \( S_n \), and (another way to state Theorem 5) \( S_n \) contains \( S_P \). If it were possible to prove that all vertices of \( S_n \) are vertices of \( S_P \), the converse of Theorem 5 would be proved for any \( n \), thus extending Theorem 6. We therefore register (but do not assert to be true!) the following

**Conjecture 1.** For any finite \( X \), \((d)\) implies \((U^{(2)})\).\(^*\)

While \((d)\) is necessary for \((U^{(2)})\) and therefore for \((U)\), and even if it were true that \((d)\) is sufficient for \((U^{(2)})\), it is certainly not sufficient for \((U)\) since \((d)\) does not put any constraints on non-binary choices while \((U)\) does [we have seen that \((U)\) implies \((5)\)]. Some sufficient conditions for \((U^{(2)})\) and \((U)\), respectively, will be given in Sections 10 and 11.

7. **Transitivity of Stochastic Preferences**

**Definition.** Alternative \( x \) is stochastically preferred (stochastically indifferent) to \( y \) if \( p_{xy} > \frac{1}{2} \) \( (p_{xy} = \frac{1}{2}) \).

In what follows, the words "transitivity of stochastic preferences" will be abbreviated to "transitivity."

**Conditions (t\_\_o) [weak transitivity], (t\_\_m) [mild transitivity], and (t\_\_) [strong

\(^*\) I am told by Professor T. Motzkin that he has proved the conjecture for \( n \leq 5 \). Because of applications to the commodity space or to the space of wagers, the continuous case (as in Theorems 19 and 14) would deserve study.
CHOICE CONSTRAINTS AND UTILITY INDICATORS

transitivity). If \( \min(p_{x_2}, p_{x_3}) \geq \frac{1}{2} \), then

\[
p_{x_1} \leq \begin{cases} \frac{1}{2} & \text{Condition (}t_6)\text{,} \\ \min(p_{x_2}, p_{x_3}) & \text{Condition (}t_m)\text{,} \\ \max(p_{x_2}, p_{x_3}) & \text{Condition (}t_n)\text{.} \end{cases}
\]

Equivalently, by (1): if \( \max(p_{x_2}, p_{x_3}) \leq \frac{1}{2} \), then

\[
p_{x_1} \leq \begin{cases} \frac{1}{2} & \text{Condition (}t_6)\text{,} \\ \max(p_{x_2}, p_{x_3}) & \text{Condition (}t_m)\text{,} \\ \min(p_{x_2}, p_{x_3}) & \text{Condition (}t_n)\text{.} \end{cases}
\]

A more symmetric equivalent form for (\(t_6\)) is, by (1),

\[
\min(p_{x_2}, p_{x_3}, p_{x_4}) \leq \frac{1}{2} \leq \max(p_{x_2}, p_{x_3}, p_{x_4})\,.
\]

**Theorem 7.** Condition (\(t_6\)) is equivalent to the following conditions: if \( p_{x_2} \geq \frac{1}{2} \), then \( p_{x_2} \geq p_{x_3} \), for all \( x \in X \); and if \( z \in X \) and \( p_{x_2} \geq p_{x_3} \), then \( p_{x_2} \geq \frac{1}{2} \).

**Proof.** See [1], Theorem 4.1.

**Theorem 8.** Condition (\(t_6\)) is strictly stronger than (\(t_m\)), and (\(t_n\)) is strictly stronger than (\(t_6\)).

**Proof.** Obvious.

**Theorem 9.** Condition (\(t_6\)) is neither sufficient nor necessary for (\(A\)).

**Proof.** Let \( n = 3 \), and \( p_{13} = 0.6 \). If \( p_{13} = p_{32} = 0.9 \), then (\(t_6\)) is satisfied but (\(A\)) is not. If \( p_{13} = p_{23} = 0.4 \), then (\(A\)) is satisfied but (\(t_6\)) is not.

**Theorem 10.** Condition (\(t_m\)) is strictly stronger than (\(A\)).

**Proof.** (i) Sufficiency. Consider (1, 2, 3) \(\leq X\). Using the notation of Section 6, we have to prove that (\(t_m\)) implies \( 0 \leq A_{x_2} \leq 1 \). By (1) we can let \( p_{13} \leq p_{12} \leq \frac{1}{2} \leq p_{22} \leq p_{11} \), without loss of generality; then

\[
\begin{align*}
p_{12} + p_{22} + 1 & \geq p_{13} + p_{23} \\
p_{12} + p_{23} + p_{11} - 1 & = A_{123} \geq 0 \\
1 - A_{132} & = p_{12} + p_{23} + p_{11} - 1 = A_{231} \leq 1.
\end{align*}
\]

Now assume (\(t_m\)). Then

\[
p_{13} \leq \max(p_{12}, p_{13}) = p_{12}, \quad A_{123} \leq p_{12} + p_{23} + p_{11} - 1 = p_{12} \leq 1
\]

and

\[
p_{13} \geq \min(p_{12}, p_{13}) = p_{12}, \quad A_{231} \leq p_{21} + p_{23} + p_{11} - 1 = p_{12} \geq 0.
\]

(ii) No necessity. By Theorems 8, 9.

**Theorem 11.** Condition (\(t_m\)) is strictly stronger than the conjunction of (\(t_6\)) and (\(A\)).

---

\(^3\) Conditions (\(t_m\)) and (\(t_n\)) were formulated by S. Valavanis-Vail [21] and originated in his work with C. Coombs. Condition (\(t_m\)) was formulated by N. Georgescu-Roegen [12] and by John Chipman [4]. Experimental tests of stochastic transitivity conditions were undertaken by Papandreou et al. [17] and by Davidson and Marschak [5].
PROOF. (i) Sufficiency. By Theorems 8, 10. (ii) No necessity. Let $X = \{1, 2, 3\}$, $p_{13} = .8$, $p_{32} = .7$, and $p_{21} = .4$. Then $(t_w)$ and $(d)$ are satisfied. But since $p_{13} > p_{32} > .5$ while $p_{31} = .6 < p_{32}$, Condition $(t_m)$ is violated.

Write $\rightarrow$ for "implies", $\leftrightharpoons$ for "implies and is implied by", $\not\rightarrow$ for "implies but is not implied by", and $\not\leftrightharpoons$ for "does not imply nor is implied by". Then our results can be summarized in

**Theorem 12.**

$$
(t_s) \quad \not\rightarrow \quad (t_m) \quad \leftrightharpoons \quad (t_w, d)
$$

$$
(d; n = 3) \quad \not\rightarrow \quad (U^{(1)}) \quad \not\rightarrow \quad (d) \quad \not\leftrightarrow \quad (t_w) \quad \not\leftrightarrow \quad (U)
$$

In political science the fact that Condition $(U^{(1)})$ does not imply Condition $(t_w)$ is interpreted as "intransitivity of the majority rule." It was pointed out by Kenneth Arrow. We have shown in [1, Theorem 3.2] that even the stronger Condition $(U)$ does not imply $(t_w)$.

8. Weak Utility Function

**Definition.** A real-valued function $w$ on $X$ is called a weak utility function if $w_x \geq w_y$ when and only when $p_{xy} \geq \frac{1}{2}$.

**Condition (w).** The function $p$ is such that $w$ exists. It follows that $w$ is unique up to an increasing monotone transformation.

**Theorem 13.** $(w) \rightarrow (t_w)$.

**Proof.** (i) Sufficiency. Let $p_{ab} \geq \frac{1}{2}, p_{bc} \geq \frac{1}{2}, a, b, c \in X$. If $(w)$ is true, then $w_a \geq w_b \geq w_c$ and $p_{ab} \geq \frac{1}{2}$. (ii) No necessity. Let $X$ be a real 2-space with generic points $x = (x_1, x_2)$ and assume that $p_{xy} > \frac{1}{2}$ when either $x_1 > y_1$ or $x_1 = y_1$ and $x_2 > y_2$. (lexicographical ordering) Compare Debreu [6].

**Condition (D).** $X$ is perfectly separable (i.e., contains a countable and dense subset); and for every $a$ in $X$, the sets $(x : p_{ax} \geq \frac{1}{2})$ and $(x : p_{xa} \leq \frac{1}{2})$ are closed. This is adapted from Debreu's paper just cited, which also contains, *mutatis mutandis*, the proof of

**Theorem 14.** If $(D)$ is satisfied, then $(w)$ and $(t_w)$ are equivalent.

**Remark.** Condition $(D)$ is trivially satisfied on $X$ finite. Moreover, every subset of a finite-dimensional real space is perfectly separable; and it is reasonable to assume that Condition $(D)$ is satisfied on a commodity space.

We summarize some of the previous results in

**Theorem 15.**

$$
(t_w; n \text{ finite}) \quad \not\rightarrow \quad (t_w, D) \quad \not\rightarrow \quad (w) \quad \not\rightarrow \quad (t_w)
$$

$$
(d; n = 3) \quad \not\rightarrow \quad U^{(1)} \quad \not\rightarrow \quad (d) \quad \not\leftrightarrow \quad (t_w, d) \quad \not\leftrightarrow \quad (t_m) \quad \not\leftrightarrow \quad (t_s)
$$
9. Strong Utility Function

**Definition.** A real-valued function \( v \) on \( X \) is called a strong utility function if there exists a monotone increasing function \( \phi_v \) such that
\[
\phi_v(v_x - v_y) = p_{xy}; \quad \phi_v(0) = \frac{1}{2}.
\]

**Condition (v).** The function \( p \) is such that \( v \) exists.

**Theorem 16.** If \( X \) is continuous and \( \phi_v \) is strictly monotone, then (i) \( v \) is unique up to an increasing linear transformation, and (ii) for any real number \( \lambda \),
\[
\phi_v(\lambda) + \phi_v(-\lambda) = 1.
\]

**Proof.** For (i), see [1, Section 2]; (ii) is obvious.

Equation (15) implies that \( \phi_v \) is anti-symmetrical about \( (0, \frac{1}{2}) \); the median and the mean (if it exists) are 0.

The strong utility \( v_x \) of alternative \( x \) corresponds to the "sensation" produced by a stimulus \( x \), as defined by Fechner [10] in 1859. It also corresponds to the relative position of a gene in the chromosome when inferred from the probability of a "crossover": this generation of a real line from a set of probabilities was noticed by D. Hilbert.²

**Condition (q) [quadruple condition].** For any \( x, y, z, t \) in \( X \),
\[
\text{if } p_{xy} \geq p_{zt}, \quad \text{then } p_{xt} \geq p_{ys}.
\]

**Theorem 17.** Condition (v) implies Condition (q) [Proof. By monotonicity of \( \phi_v \)]; Condition (q) implies Condition (t) [Proof. In (16) put \( z = t \) and use Theorem 7]; Condition (v) implies Condition (w) [Proof. In (14), put \( v = w \)].

**Theorem 18.** If \( X \) is finite, \( X = \{1, \cdots, n\} \), then (q) implies (v) only for \( n \leq 4 \).

**Proof.** See [1], Proof of Theorem 4. 1.

For \( X \) continuous it is useful to consider the following

**Condition (s-c) ["stochastic continuity"];** If \( p_{xy} < q < p_{zt} \), then there is a \( t \) in \( X \) such that \( p_{xt} = q \). In [7] Debreu has proved

**Theorem 19.** If Condition (s-c) is satisfied, then (q) implies (v).

If both \( (U^{(s)}) \) and \( (v) \) are satisfied, we can define a random function \( V_x \) by
\[
U^{(s)}_x = v_x + V_x, \quad EV_x = 0,
\]
and set up the following three conditions, each stronger than the preceding one:

There exist two functions on \( X \), a real-valued \( v \) and a random-valued \( V \), and a distribution function \( \phi_v \) such that
\[
p_{xy} = \phi_v(v_x - v_y), \quad \text{Pr}(V_y - V_x \leq \lambda) = \phi_v(\lambda),
\]

² In a paper [14] pointed out to me by Hans Rademacher.
where

**Condition (V):** \( \lambda = v_x - v_y \);

**Condition (V'):** \( \lambda = \text{any real number} \);

**Condition (V''):** \( \Pr (V_x \leq \alpha, \ V_y \leq \beta) = \Pr (V_x \leq \beta, \ V_y \leq \alpha), \ \lambda, \alpha, \beta = \text{any real numbers} \).

In Condition (V) the distribution of the differences \( V_x - V_y \) is defined only for each pair of alternatives (possibly finite in number); in (V') this distribution is defined for any real argument; in (V''), in addition, the (multivariate) distribution function of \( V \) is symmetrical in its arguments. From (17) and Theorem 16 one obtains

**Theorem 20.** (V'') \( \rightarrow \) (V') \( \rightarrow \) (V) \( \rightarrow \) (\( U_x \), \( v \)) \( \rightarrow \) if \( X \) is finite;

(V'') \( \rightarrow \) (V') \( \rightarrow \) (V) \( \rightarrow \) (\( U_x \), \( v \)) \( \rightarrow \) if \( X \) is continuous.

In the work of Thurstone and his school a condition even stronger than (V'') is used: the distribution function of \( V \) is assumed to be normal and symmetric (zero means, equal variances, equal covariances). Actually, a normal distribution of differences \( \phi \) (as assumed by Fechner) is consistent also with a normal distribution of \( V \) non-symmetrical in its arguments.

For possible application to the commodity space some of our results can be summarized in

**Theorem 21.** If \( X \) is continuous and the conditions (D) and (s-c) are satisfied, then

\[
(V'') \rightarrow (V') \rightarrow (U^{(2)}, v) \rightarrow (v) \rightarrow (q) \rightarrow (t_q) \rightarrow (t_m) \]

\[
(U^{(2)}) \rightarrow (j) \rightarrow (w) .
\]

10. **Strict Utility Function (In the "binary" sense)**

**Definition.** A positive-valued function \( u^{(2)} \) on \( X \) is called a strict utility function (in the "binary" sense) if

\[
(18) \quad p_{xy} = \frac{u^{(2)}_x}{u^{(2)}_x + u^{(2)}_y} .
\]

Clearly, every positive multiple of a strict utility function is also a strict utility function.

**Condition (u^{(2)}).** The function \( p \) is such that \( u^{(2)} \) exists. Applying (18) one obtains

**Theorem 22.** Condition (u^{(2)}) is true only if (i) \( 0 < p_{xy} < 1 \) for every \( x, y \) in \( X \), and (ii) for any subset \( \{1, \cdots, m\} \subseteq X \) \( (m > 1) \), the following equation holds:

\[
(19) \quad p_{12} \cdot p_{23} \cdots p_{m-1,m} \cdot p_{1m} = p_{12} \cdot p_{23} \cdots p_{m-1,m} \cdot p_{1m} .
\]

Note that (19) is "directly testable." Condition (u^{(2)}) was formulated by
Törnqvist [20], Bradley and Terry [3], and L. R. Ford [11]. It is a weak form of a postulate that has been proposed by Luce [15] and that involves multiple-choice probabilities: see Section 11 below.

**Theorem 23.** \((u^{(2)})\)\(\implies\) (v).

**Proof.** (i) Necessity. Put \(u_x^v = e^v;\) then by (18) \(p_{xy} = \phi_x(v_x - v_y),\) where

\[
\phi_x(\lambda) = \frac{1}{1 + e^{-\lambda}},
\]

the "logistic curve." (ii) No sufficiency. For \(X\) continuous, by the above proof of necessity (due to Luce) and by Theorem 16, we see that Condition \((u^{(2)})\) is not implied by (v), since \((u^{(2)})\) restricts the function \(\phi_x\) to the form (20). For \(X\) finite, let

\[
X = (1, 2, 3); \quad p_{12} = .6, \quad p_{23} = .7, \quad p_{13} = .8.
\]

Then (v) is satisfied\(^5\) by \(v_1 = .3, v_2 = .2, v_3 = .0,\) and \(\phi_x(\lambda) = .5 + \lambda.\) But (19), and therefore \((u^{(2)})\), is not.

**Theorem 24.** If \(X\) is finite, then \((u^{(2)})\) implies \((U^{(2)})\).

**Proof.** We shall use an arithmetical identity proved in [1, Theorem 3.6] and [8]:

**Lemma.** Let \(N = (1, \ldots, n)\) and denote by \(r = (1, 2, \ldots, n)\) the permutation on \(N\) in which the \(k\)th place is occupied by the element \(k,\) denote by \(R_{ik}\) the set of permutations on \(N\) in which \(i\) precedes all other elements of \(M \subseteq N.\) Then, for any positive \(u_1, \ldots, u_n,

\[
\frac{u_i}{\sum_{h \in M} u_h} \equiv \sum_{r \in R_{iM}} \prod_{j=1}^{n-1} \frac{\sum_{h \in R_{jM}} u_{hr}}{\sum_{h \in R_{jM}} u_{hr}}.
\]

To prove Theorem 24 from this Lemma, let \(X = N, M = (i, j), u_x = u_x^M;\) then by (18) the left side of (22) becomes \(p_{ij},\) and the set \(R_{iM} = R_{ij},\) as defined in Theorem 3. Since each of the products on the right side is positive and they add up to \(p_{ij},\) we can equate each of them to a probability of ranking in the sense of (7), so that, for any ranking \(r\) on \(N,

\[
P(r) = P(1, 2, \ldots, n) = \frac{u_{1r}}{u_{1r} + \cdots + u_{nr}} \cdot \frac{u_{2r}}{u_{2r} + \cdots + u_{nr}} \cdots \frac{u_{(n-1)r}}{u_{(n-1)r} + u_{nr}}.
\]

Since we interpret each \(u_i\) as \(u_i^{(2)}\), then by Theorem 4 \((U^{(2)})\) is satisfied.

If \(X\) is continuous, the reasoning just used can be applied to any finite subset of \(X.\) Hence the following

**Conjecture 2.** Theorem 24 applies to any \(X.\)

\(^5\) It is noteworthy that for \(n = 3,\) Condition (4) is not only necessary but also sufficient for (v). This follows from Theorems 17 and 18.
Remark. The proof of Theorem 24 could be much simplified if Conjecture 1 (Section 6) were proved to be true. For it would then suffice to prove that \((U^{(2)})\) implies \((D)\). This is indeed the case: for, writing \(u_y^{(2)}/u_x^{(2)} = a\), \(u_z^{(2)}/u_y^{(2)} = b\), and \(u_x^{(2)}/u_z^{(2)} = c\), we have \(abc = 1\), and by (10), (18), we obtain

\[
0 < d_{xyz} = \frac{1}{1 + a} + \frac{1}{1 + b} + \frac{1}{1 + c} - 1 = \frac{1}{1 + k} < 1,
\]

where \(k = (1 + ab + bc + ca)/(1 + a + b + c) > 0\), since \(a, b, c > 0\).

Theorem 25. \((U^{(3)})\) does not imply \((U^{(2)})\).

Proof: Consider again example (21). Since \(p_{21} = .2\), it follows that \(d_{123} = .5\); hence Condition \((D)\) and, by Theorem 6, Condition \((U^{(2)})\) are satisfied But (19) is not.

Since example (21) was seen to satisfy \((u)\), Theorem 25 can be replaced by the even stronger

Theorem 26. The conjunction of \((U^{(3)})\) and \((u)\) does not imply \((U^{(2)})\).

Some of the results of this and the preceding section can be summarized in

Theorem 27. For \(X\) finite,

\[
(U^{(3)}) \quad \longrightarrow \quad ([U^{(3)}], (v)) \quad \longrightarrow \quad (p) \quad \longrightarrow \quad (t_{(S)}) \quad \longrightarrow \quad (t_{(W)}) \quad \longrightarrow \quad (w)
\]

Remark on the commodity space. If Condition \((D)\) is satisfied on the continuous commodity space and if Conjecture 2 is correct, Theorem 27 applies also to the commodity space; if, in addition, \((s-c)\) is satisfied, then \((v) \quad \longrightarrow \quad (q)\).

11. Some Results on Multiple-Choice Probabilities

In Section 4, the Condition \((U)\) was stated. Clearly, \((U)\) implies \((U^{(3)})\). Similarly, \((U^{(3)})\) can be replaced by the following stronger

Condition \((u)\). There exists a positive-valued function \(u\) on \(X\) such that for any finite subset \(M = (x_1, \ldots, x_m) \subseteq X\), the multiple-choice probability

\[
p_{x_1}(M) = \frac{u_1}{\sum_{j \in M} u_j}.
\]

The function \(u\) may be called a strict utility function in the multiple-choice sense.

As was proved in [1], \((u)\) implies that the probability of each ranking in the multiple-choice sense, defined as

\[
P(r) = \Pr(U_{1r} > U_{2r} > \cdots > U_{nr}), \quad \text{with} \quad \sum_{r \in \mathcal{L}} P(r) = p(M),
\]

\[
P(r) = \Pr(U_{1r} > U_{2r} > \cdots > U_{nr}), \quad \text{with} \quad \sum_{r \in \mathcal{L}} P(r) = p(M),
\]
exists and is precisely the expression on the right side of (23). For example, if \( X = (1, 2, 3) \), then

\[
P(xyz) = \frac{u_x}{u_x + u_y + u_z} \cdot \frac{u_y}{u_y + u_z} = p_x(x) \cdot p_y(y, z),
\]

by (25). Thus \((u)\) implies (and, as shown in [1], is equivalent to) the proposition that the probability of a given ranking of a finite set of alternatives is the product of probabilities of multiple choices made successively, from the set itself, then from the subset remaining after eliminating the first choice, then from the subset remaining after eliminating the second choice, etc. Applying the Lemma of Section 10 we obtain, using (25),

**Theorem 28.** \((u) \rightarrow (U)\) when \( X \) is finite.

This result is discussed in more detail in [1]. We can prove the even stronger

**Theorem 29.** \((u) \rightarrow [(U), (v)]\) when \( X \) is finite.

**Proof.** (i) Necessity. By Theorems 28 and 23, remembering that \((u)\) of course implies \((u^{(2)})\). (ii) No sufficiency. Note that for \( X = (1, 2, 3) \), condition \((u)\) implies, by (19) and (26) respectively,

\[
p_{13} \cdot p_{31} \cdot p_{33} = p_{13} \cdot p_{33} \cdot p_{31},
\]

and

\[
P(123) \cdot P(231) \cdot P(312) = P(213) \cdot P(132) \cdot P(321).
\]

Moreover, \((U)\) implies \( p_{xyz} = P(xyz) + P(xzy) + P(zyx) \) and \( p_x(X) = P(xyz) + P(xzy) \). It is easy to construct examples satisfying \((v)\) and \((U)\) but not (27) or (28). For example, use (21), which we have seen satisfies \((v)\) and which contradicts (27); and put, in addition, the following values: for the probabilities of rankings, \( P(123) = P(132) = .3, P(231) = P(213) = .2, P(312) = P(321) = 0 \); for the ternary probabilities, \( p_t(x) = .6, p_t(x) = .4, \) and \( p_t(x) = 0 \). The foregoing results, together with some obtained in the previous section, can be summarized in

**Theorem 30.**

\[
\begin{align*}
(u) & \iff [(U), (v)] \iff (U) \\
\downarrow & \downarrow & \downarrow \\
(u^{(2)}) & \iff [(U^{(2)}), (v)] \iff (U^{(2)}) \\
\downarrow & \downarrow \\
(v) & \iff (d)
\end{align*}
\]

Luce [15] proposed a condition still stronger than \((u)\), being a conjunction of \((u)\) and the following

**Condition \((u')\).** Denote by \( p_x^*(x) \) the probability that \( x \) is the last choice out of the subset \( X^* \subseteq X \). There exists a real-valued function \( u' \) on \( X \) such that for any finite subset \( M = (x_1, \ldots, x_n) \subseteq X \), the probability

\[
\begin{align*}
&
\end{align*}
\]
\[ p'_x(M) = \frac{u'_i}{\sum_{j \in M} u'_j} . \]

The function \( u' \) may be called a strict disutility function. However, apart from trivial cases, the functions \( u \) and \( u' \) cannot both exist.

**Theorem 31.** Conditions \((u)\) and \((u')\) cannot both hold only if all first choices are equally probable and all last choices are equally probable, or if \( X \) has only two elements.

**Proof.** If \( n \geq 3 \), let \( M \subseteq X \) consist of three elements, to be denoted generically by \( x, y, z \). Assume both Conditions \((u)\) and \((u')\). Without loss of generality, let \( u_x + u_y + u_z = 1 = u'_x + u'_y + u'_z \), so that \( p_x(M) = u_x, \ p'_x(M) = u'_x \), for all \( x \) in \( M \). Then

\[ u_x \cdot \frac{u'_i}{1 - u'_i} = u'_x \cdot \frac{u_x}{1 - u_x}, \quad \text{all } x, z \text{ in } M, \]

since both the left and the right side express the probability of the same event, viz., of \( x \) being the first, and \( z \) the last choice. Hence \( u'_x = u_x, \ p'_x(M) = p_x(M) \) for any pair \( x, z \) of the arbitrary three-elements subset \( M \). Interchanging variables \( x, y, z \), and applying \((u), (u')\),

\[ p'_x(X) = p'_y(X) = p'_z(X) = p_x(X), \quad \text{all } x, y \text{ in } X. \]

This proves the Theorem for \( n \geq 3 \). When \( n = 2 \), the same reasoning yields trivially the identity \( p_x(x, y) = p'_x(x, y) \). For a more detailed proof see [1, Theorem 3.8].

It has been pointed out by Luce (in private correspondence) that while in experiments on choices the probabilities of last choices may or may not be accessible to observations, they definitely are in experiments on perceptions: the subject’s decision as to which is the loudest of three sounds is as observable as his decision as to which sound is the softest. If one accepts \((u)\), symmetry requires one to accept \((u')\) also. But, as we have shown, these two conditions are mutually inconsistent.

Therefore, we must limit ourselves to the weaker Condition \((u^{(2)})\). It involves binary choices only, so that, when identifying the loudest sound, the subject identifies also the softest. That is, if the subset \( X^* \) consists of two elements, then Conditions \((u^{(2)}), (u)\), and \((u')\) are identical. A strict utility function in this limited sense—\( u^{(2)} \) rather than \( u \)—is compatible with the existence of the corresponding strict disutility function and of a random utility indicator \( U^{(2)} \) involving binary choices only.

**12. Binary-Choice Constraints and the Form of the Random Utility Indicator**

Consider Example (ii) of Section 3, and let \( m = 2, \ 0 < T_1 = A < 1, \) and \( T_2 = 1 - A \). Then

\[ U^{(2)} = x_1^r x_1^{1-r}. \]

the function \( U^{(2)} \) is monotone increasing in \( x \) in the sense that if \( x = (x_1, x_2), \ y = (y_1, y_2), \ x_1 \geq y_1, \) and \( x_1 > y_1 \), then \( U^{(2)}(x) > U^{(2)}(y) \). \( U^{(2)} \) is a random func-
tion depending on the random parameter $A$. If $U^{(z)}$ is a utility indicator in the “binary” sense of (2), then $p_{xy} = Pr\left(U^{(y)}_x \geq U^{(x)}_y\right) = Pr(x^4y^{-4} \leq x^4x^{-4})$. Let $y_1 > x_1$ and $y_3 < x_3$ (so that $x$ does not dominate $y$, and conversely). Then $\log(y_1/x_1) > 0$, $\log(x_3/y_3) > 0$, and

$$p_{xy} = Pr\left(\frac{A}{1-A} \leq \frac{\log(x_3/y_3)}{\log(y_1/x_1)}\right) = F(\mu_{xy}),$$

where $\mu_{xy} = \log(x_3/y_3)/\log(y_1/x_1)$ and $F$ is a monotone non-decreasing function depending on the distribution of $A$. We shall now show that the form (29) of the random utility indicator is not consistent with some binary-choice constraints:

**Theorem 32.** If the random utility indicator has the form (29), then the strong transitivity condition (t4) is, in general, not satisfied.

**Proof.** Consider three points $x, y, z$ with $z_1 > y_1 > x_1$ and $z_3 < y_3 < x_3$. Then

$$p_{xy} = F(\mu_{xy}), \quad \mu_{xy} = (\log x_3 - \log y_3)/(\log y_1 - \log x_1),$$

$$p_{yz} = F(\mu_{yz}), \quad \mu_{yz} = (\log y_3 - \log z_3)/(\log z_1 - \log y_1),$$

$$p_{zx} = F(\mu_{zx}), \quad \mu_{zx} = (\log z_3 - \log z_1)/(\log z_1 - \log x_1),$$

$$\mu_{xz} = \frac{\mu_{xy}(\log y_1 - \log x_1) + \mu_{yz}(\log z_1 - \log y_1)}{(\log y_1 - \log x_1) + (\log z_1 - \log y_1)},$$

a convex combination of $\mu_{xy}$ and $\mu_{yz}$. Hence $\min(\mu_{xy}, \mu_{yz}) \leq \mu_{xz} \leq \max(\mu_{xy}, \mu_{yz})$, and since $F$ is non-decreasing, $\min(p_{xy}, p_{yz}) \leq p_{xz} \leq \max(p_{xy}, p_{yz})$. This is consistent with the Weak and the Mild Transitivity Conditions, ($t_a$) and ($t_a$) of Section 7, but not with the Strong Transitivity Condition ($t_4$). All three conditions are directly testable.

Thus (29) is in general inconsistent with the existence of strict or strong utility functions, since those imply strong transitivity (Theorems 17 and 27).

Consider now another random form of the “Cobb-Douglas function”:

$$U^{(z)}_x = x^\alpha x^{-\alpha} + V_x,$$

where $\alpha$ is a constant, $0 < \alpha < 1$, and $V$ is a random function on $X$. If $U^{(z)}$ is a random utility indicator, then

$$p_{xy} = Pr(U^{(y)}_x \geq U^{(x)}_y) = Pr(V_x - V_y \leq x^\alpha x^{-\alpha} - y^\alpha y^{-\alpha}).$$

Suppose we have found $V$ such that $Pr(V_x - V_y \leq \lambda) = \phi(\lambda)$ for any real $\lambda$, with $\phi$ being a distribution function independent of $x, y$. Then Condition ($V''$) of Section 9 is satisfied, and

$$U^{(z)}_x = V_x = x^\alpha x^{-\alpha}.$$

---

*Maximizing $U^{(z)}$ with respect to $x$, subject to the constancy of income $x_1 + x_2$, we see that $A$ is the (random) proportion spent on the 1st commodity. Hence the observed frequency distribution of this proportion should help to estimate the distribution of $A$—assuming that members of the sampled population were characterized by the same random function $U^{(z)}$. (But we shall not pursue this matter here.)*
is a strong utility function consistent with the assumed random utility indicator.\footnote{One may want to use indifference lines to picture some relevant aspect of a given random utility indicator, and in particular, its mean. In the case just discussed, described by (30) and (31), we may choose $V$ to have 0 expectation; then $EU_x^{(2)} = v_x$. Accordingly, an indifference line through $x$ consists of all points $y$, with $v_y = v_x$ and $p_x^y = \phi_x(0) = \frac{1}{2}$ as in (14); it is then the common boundary of the two sets used in Condition (D), Section 8. On the other hand, if the random utility indicator is described by (29), equation $EU_y^{(3)} = EU_x^{(2)}$, or possibly $E \log U_y^{(3)} = E \log U_x^{(2)}$, gives quite a different representation.}

If we could accept the still stronger condition $(V'')$, i.e., let the (infinite-dimensional) distribution of $V$ be symmetrical in its arguments $(x, y, \ldots)$, we would have to search for such distributions. However, the symmetry assumption is hardly realistic: for example, in the case of normality it implies that deviations from the average budget composition have not only the same variances everywhere but also the same correlations for any two points of the $X$-plane, however close or remote in relation to each other.

The examples used in this section have the purpose merely of illustrating the problem: to find a class of random utility indicators consistent with a given set of binary-choice constraints.

REFERENCES


CHOICE CONSTRAINTS AND UTILITY INDICATORS


