DOCUMENT RESUME

ED 086 723 TM 003 373

| | |
|---|---|
| AUTHOR | Wallsten, Thomas S. |
| TITLE | The Psychological Concept of Subjective Probability: A Measurement-Theoretic View. |
| INSTITUTION | North Carolina Univ., Chapel Hill. L.L. Thurstone Psychometric Lab. |
| SPONS AGENCY | National Inst. of Mental Health (DHEW), Bethesda, Md.; National Science Foundation, Washington, D.C. |
| REPORT NO | 115 |
| PUB DATE | Feb 73 |
| NOTE | 39p. |
| EDRS PRICE | MF-$0.65 HC-$3.29 |
| DESCRIPTORS | *Decision Making; *Expectation; Measurement; *Models; Probability; *Probability Theory; Psychological Characteristics; *Psychometrics |
| IDENTIFIERS | *Subjective Probability |

ABSTRACT

A point of view is presented concerning the psychological concept of subjective probability, both to study its relation to the corresponding mathematical and philosophical concepts and to provide a framework for the rigorous investigation of problems unique to psychology. In order to do this the empirical implications of axiom systems for measurement are discussed first, relying primarily on Krantz's work, with special emphasis, however, on some similarities and differences between psychological and physical variables. The psychological variable of uncertainty is then examined in this light, and it is concluded that few, if any, current theories are satisfactory when viewed from this perspective, particularly those deriving from the mathematical work in the axiomatic foundations of probability. This might appear to pose difficulties for applications to real problems of normative decision theory when those applications require numerical probability judgments from individuals. Two possible solutions are discussed briefly. (Author)





# The Psychological Concept of Subjective Probability: A Measurement-Theoretic View

Thomas S. Wallsten

February, 1973

Number 115

THE L. L. THURSTONE PSYCHOMETRIC LABORATORY
UNIVERSITY OF NORTH CAROLINA

CHAPEL HILL, N. C. 27514

# The Psychological Concept of Subjective Probability: A Measurement-Theoretic View[1]

A point of view is presented concerning the psychological concept of subjective probability, both to study its relation to the corresponding mathematical and philosophical concepts and to provide a framework for the rigorous investigation of problems unique to psychology. In order to do this the empirical implications of axiom systems for measurement are discussed first, relying primarily on Krantz's work, with special emphasis, however, on some similarities and differences between psychological and physical variables. The psychological variable of uncertainty is then examined in this light, and it is concluded that few, if any, current theories are satisfactory when viewed from this perspective, particularly those deriving from the mathematical work in the axiomatic foundations of probability. This might appear to pose difficulties for applications to real problems of normative decision theory when those applications require numerical probability judgments from individuals. Two possible solutions are discussed briefly.

The concept of subjective probability, or expectancy, has been used variously in psychology, often in ways strongly influenced by the mathematical or philosophical meanings of that term. Psychologists interested in behavioral decision theory have concentrated primarily on three related problems. One has been that of how subjective probability combines with other variables, especially utility, to determine decisions. (see Rapoport and Wallsten, 1972, for a review of recent literature). A second area of research has centered around the process of subjective probability, focusing on questions such as what independent variables influence subjective probability, or how is subjective probability revised with new information (see, e.g., Edwards, 1968; Rapoport and Wallsten, 1972; Tversky and Kahneman, 1972; Wallsten, 1972; Wise, 1970)? Finally, investigation by psychologists and others has been directed to the experimental measurement of subjective probability (see, e.g., Beach, and Phillips, 1967; Staël von Holstein, 1970; Winkler, 1967).

The term "subjective probability" is not used identically in all cases, although it generally is taken to refer to some aspect of an individual's (or a group's) uncertainty or expectation concerning which subset of a set of events is going to occur, or is going to occur most frequently, or is true under specific conditions. We will attempt some clarifications of the psychological concept in this paper, particularly to present a point of view concerning the relation between certain mathematical work and psychological research. It is suggested that this might provide a framework within which those questions of subjective probability unique to psychology may be formulated and investigated.

The ideas in this paper revolve closely around questions of measurement, and a considerable amount of space is devoted to its empirical justification. This is particularly important, since so much recent psychological work on subjective probability has depended on measurement in one form or another, often treating numbers emitted by subjects as measures of subjective probability or odds. However, nothing said here should be taken to imply that any theory of behavior under uncertainty must have measurement, in the sense to be defined, as one of its ends. If this is not one of its purposes, then clearly that theory need not worry about its justification. But, in the absence of evidence or reason to support metric assumptions about the data, the theory should be qualitative, or ordinal, in nature, as it will be argued should measurement-oriented formulations. An excellent and important recent development in this spirit, which allows the concept of expectancy to be used meaningfully with infrahumans as well as humans, is by Irwin (1971).

The approach to be advocated in this paper has been suggested before (Wallsten, 1970, 1971), but can be expanded and made considerably more clear now in light of Krantz's (1972a,b) analysis of measurement foundations as qualitative empirical laws. First, we will discuss axiom systems for measurement and their interpretation as empirical statements. Special attention will be paid to some similarities and differences between physical and psychological variables in terms of methods for their definition and empirical realization. These similarities and differences have strong implications for data interpretation and theory construction in general. Following this we will be in a position to consider the psychological concept of subjective probability. The paper will end with some comments concerning the relation between theoretical and applied research in this area.

## Empirical Implications of Axiom Systems for Measurement

Research in the foundations of measurement is concerned with the conditions required of a set of elements ordered with respect to a particular qualitative property such that that property may be represented numerically in a meaningful fashion, i.e., measured. For example, it may be desired to represent the masses of objects, individual's utilities of objects, or individual's subjective probabilities of events numerically. The conditions are stated in the form of axioms about the ordered set, which taken together are at least sufficient for the existence of an isomorphic (or homorphic) mapping from the set of objects into the real numbers. A proof establishing such existence is called a representation theorem. A uniqueness theorem establishes the relation that exists between any two permissible mappings.

Recently Krantz, et al (1971, p. 26ff) have pointed out that the search for conditions leading to measurement scales is a search for lawfulness (see also Krantz 1972a,b; Krantz and Tversky, 1971). In that sense, the axioms in reference to a particular set of elements and particular operations are empirical statements, some of which are subject to empirical verification. Thus, the construction of measurement scales de novo is accomplished only with the development of an appropriate set of laws, or an appropriate theory.

To discuss the empirical implications of axiom systems for measurement, following Krantz's views, consider a set of objects possessing some qualitative property of interest, for example, the physical property of mass. The set can be empirically ordered with respect to that property, in the case of mass with a pan balance. The empirical ordering in general may be denoted by the symbol $\precsim$. Thus if rock a in one pan tips the balance when rock b is in the other pan, we say b is no heavier than a, or $b \precsim a$. And often, but not always, two elements may be combined, or concatenated with respect to the property of interest and compared with a third element. In the present example this would be done by placing two rocks in one pan of a balance and a third rock in the other. The concatenation operation in general may be denoted $\circ$. Thus, if rocks b and c together tip the pan balance over a, we write $a \precsim (b \circ c)$.

It is usually most useful to consider the ordering relation and the concatenation operation, respectively, to correspond to the relation "less than or equal to", denoted $\leq$, and the operation "addition", denoted +, in the real number system. Given a set

of reasonable axioms, an isomorphic mapping associates each empirical element $\underline{a}$ (or class of equivalent empirical elements) with a number which we may call $\phi(\underline{a})$ such that for all $\underline{a}$, $\underline{b}$ in the empirical set, $\underline{b} \precsim \underline{a}$, iff[2] $\phi(\underline{b}) \leq \phi(\underline{a})$. When concatenation is empirically defined and given an appropriate axiom system, the isomorphism also assures that $\underline{a} \precsim (\underline{b} \circ \underline{c})$ iff $\phi(\underline{a}) \leq \phi(\underline{b}) + \phi(\underline{c})$. When the mapping exists we may work with the numbers instead of the elements, confident that within (often unspecifiable) limits of error we are correctly predicting the relevant aspects of the qualitative property.

As already mentioned, the qualitative conditions, or axioms, which must be satisfied by the elements for the mapping to exist, thereby allowing us the convenience of numbers, are empirical statements. Some of these are formulated in such a way that they may be actually subjected to test. For example one of the assumptions is that of transitivity, if $\underline{a} \precsim \underline{b}$, and $\underline{b} \precsim \underline{c}$, then $\underline{a} \precsim \underline{c}$, the empirical test of which is clear. If it systematically fails the desired mapping does not exist, and unless another can be established, real numbers can not be used to represent the particular property of the elements.

Other axioms, although empirical in principle, are formulated in such a way that they usually can not be tested satisfactorily. An example of this is the Archimedian axiom, which essentially states that for any pair of elements, $\underline{a} \precsim \underline{b}$, $\underline{a}$ can be concatenated with a sufficient number of identical copies of itself (say $\underline{n}$ copies, written for convenience as $\underline{na}$) so that $\underline{b} \prec \underline{na}$.[3] Clearly for a particular set of elements this axiom may not hold for a variety of uninteresting reasons, such as, for example, there not

being sufficient copies of a. Occasionally, however, it may fail on more substantive grounds. Thus, if we are working with velocities and b is the velocity of an electromagnetic wave, then regardless of the size of n, it will not be the case that $b \prec na$. Some of the problems involved in empirically testing axioms are discussed in Krantz et al (1971, p. 28ff), and others in Rapoport and Wallsten (1972).

As qualitative laws, the axioms required for the measurement of intermediate values of mass, length, and time intervals are so obviously true or uninteresting that there is no reason to experimentally investigate them (see, e.g., Krantz, 1968). These properties were successfully measured long before the procedures were theoretically justified. However, the same set of axioms is not valid for empirical relational sets with different attributes of interest. It does not apply, for example, when the property is utility, intelligence, anxiety, brightness, or almost any other likely to arise in the social sciences.

There are numerous reasons why this set of axioms does not generally apply to such properties, but the most important is the lack of an empirically defined concatenation operation. It is this lack which led Campbell (1920) and others to claim that fundamental measurement will never be possible in the social sciences. This is clearly wrong, as evidenced by developments in the theory of simultaneous conjoint measurement (Luce and Tukey, 1964; Krantz, et al, 1971), which provides for the simultaneous measurement of two or more variables given certain conditions. A general lesson from this work is that qualitative axioms embodying different empirical properties, but necessary and sufficient to prove

representation and uniqueness theorems can be developed and tested.

These axioms, which should suggest experiments to be performed, will constitute a theory concerning how the property or properties under consideration are ordered and, perhaps, how each property combines with itself or with other properties. Or to put it differently, these axioms will constitute a theory concerning the qualitative behavior of a set of elements subjected to certain operations, when the members of that set are presumed to differ among each other in the property or properties of interest. Indeed, for the purposes of measurement the properties are defined only in terms of the elements' behaviors in response to certain empirical operations.

In that sense, going back to the previous example, for purposes of measurement mass is defined only in terms of the behavior of rocks in pan balances. The fact that the variable so defined can be related to behaviors of many other objects as well and that units of mass can be algebraically combined with units of other variables in meaningful fashions attests to its vast generality and usefulness.

Similarly in the social sciences, especially psychology, variables may be defined in terms of qualitative laws, or axioms, concerning the behavior of elements presumed to differ in the particular variable or variables. The elements here, however, are organisms under different circumstances, or one organism under various circumstances, and, potentially, the variables are any from achievement to zoophilism. Perhaps the earliest work in this vein is the axiomatization of utility theory by Von Neumann and Morganstern (1944).[4]

Luce (1972) has argued that the degree of stability and generality obtained with measurement of physical variables, particularly the ability to combine measures of different variables in meaningful algebraic structures, has not been shown yet in psychophysics and doubts, therefore, that measurement as it axists in physics will ever exist in psychophysics. He has explicitly not extended his argument to other areas of psychology, not because of evidence to the contrary, but because of a dearth of evidence.

His argument is probably valid for other areas of psychology, such as learning, motivation, or decision theory, which consider intervening variables (see footnote 4). What are the reasons? Certainly not that the types of psychological variables under consideration need be any less well defined than the physical variables. This can be done analogously in both cases, with specification of ordinal empirical laws.

A possible well known reason is that psychology has yet to determine a small set of variables that in some sense is basic to understanding all aspects of behavior, and whose inter-relationships may be specified. Perhaps such a set does not exist.

There is another reason, which has implications for the use that can be made of resulting measurement scales, and the kinds of interpretations that can be attached to them. Although the definitions of physical and psychological intervening variables, or properties, may be equally fundamental, their empirical realizations are not equally simple. Physical variables can be made manifest and studied with more-or-less simple apparatus under well defined conditions. Thus, elements and combinations of

elements are ordered with respect to mass by means of a pan balance, or other instrument calibrated to reflect the information which would be given by a pan balance. And importantly, although obvious once mentioned, the variables being investigated are independent of the apparatus used to investigate them. A set of rocks could be weighed on any suitable pan balance or corresponding instrument. Ignoring relativistic consideration, it is generally assumed that manipulations on the variables leave unaltered the equipment through which the effects are observed. For example, one may observe the pressures of various gases at different temperatures, assuming that the temperature changes affect the gases and not the indicating instrument. It is assumed that readings on the instrument will reflect only pressure, and not other factors, regardless of the temperature.

The situation is different in psychology. Luce (1972) has suggested that psychophysics might profitably be considered the study of a very complicated measuring device that transduces various inputs into common neural units, rather than considering it measurement comparable to physical measurement. For intervening variables studied in other branches of psychology matters may be more complex yet. Here the variables originate within the organism in response to events internal and external to it. That is, to repeat from above, the organisms are the elements, and the properties that they embody and we desire to measure vary depending on circumstances internal and external to the elements.

But more than that, the organism is also the instrument that conveys information to us regarding the qualitative ordering of

the property or properties of interest. Thus, the human in whose uncertainties we are interested tells us by his behavior whether his uncertainty under condition A is greater than or less than that under B. If habit strength and incentive are the variables being conjointly measured, the dog's behavior tells us whether one combination is ordered above or below another. Unlike physics we can not separate the elements in whose properties we are interested from the device which makes those properties manifest. They are one and the same, namely the living organism. Manipulations intended to affect the variables of necessity also affect the device used to order the variables. Clearly, if the latter is not invariant, simple algebraic relations between the former will not emerge. Indeed, theories about the variables will often depend on theories about the device.

This, it is claimed, is a fundamental difference between the empirical realization and measurement of variables in psychology and physics. And it is for this reason that resulting measurement scales will not have the same generaltiy in psychology as in physics.

## Subjective Probability

Two arguments were developed in the previous section. The first was that for purposes of original measurement psychological variables should be defined in terms of qualitative empirical statements which are at least sufficient to prove a representation theorem. The second point was that since any behavior which makes the variable manifest also reflects other variables, changes in behavior over different situations often can not be understood without embedding that theory in a more general one encompassing the other variables. It is with these considerations before us

that we look at the relation between the psychological and mathematical/philosophical concepts of subjective probability.

The mathematical work on subjective probability has centered on formulating axioms concerning an ordered set (technically, concerning an algebra of sets, to be defined below), which together are sufficient to prove the existence of a function $\underline{P}$ from the set into the real interval [0,1] such that the three properties of a probability measure hold, viz: $\underline{P}(A) \geq 0$; $\underline{P}(X)=1$; and if $A \cap B=\phi$, then $\underline{P}(A \cup B)=\underline{P}(A)+\underline{P}(B)$; where X is the sample space, or sure event; $A, B \subset X$; and $\phi$ is the null event.[5] As Fishburn (1967) has pointed out, the axiomatizations have been of two forms. In one the elements are ordered by a binary relation denoted here $\precsim$, and in which $A \precsim B$ is read "A is not more likely than B." Axiomatizations of this sort are found in de Finetti (1937), DeGroot (1970), Koopman (1940), Villigas (1964, 1967), Luce (1967), and others. They are collectively discussed in Chapter 5 of Krantz, et al (1971). DeGroot's (1970) formulation, based on that by Villigas (1964), is especially clear.

In axiom systems of the second form, often called subjectively expected utility (SEU) theory, the elements are ordered by a binary relation denoted here $\precsim^*$, and in which $a \precsim^* b$ is read "a is not preferred to b." Here a and b are conceived of as sure commodities or as probability mixtures of outcomes, i.e., commodities conditional on uncertain events. A representation (utility function) is established from the certain outcomes or commodities into the real numbers, and a probability function is simultaneously established from the uncertain events into the

interval [0,1]. Axiom systems of this form are in Savage (1954), Fishburn (1970), and Luce and Krantz (1971).

The philosophical impact of this work has been to establish a foundation r probability theory as the "opinion of rational man," or as "rational opinion." That is, either binary relation $\precsim$ or $\precsim^*$, depends on judgments of the sort humans often make, and the axiom systems contain "rational" statements about relative likelihood or preference, respectively. If one does not disagree with the axioms, one can not disagree with their implications and, therefore, probability theory describes rational opinion. (The fact that mortals may exist who accept the axioms but not their conclusions is immaterial here.)

Furthermore, de Finetti (1937) proved that a coherent set of preferences among probability mixtures of outcomes is a necessary and sufficient condition for the derivation of a mapping which satisfies the requirements of a probability measure from the uncertain events into the interval [0,1]. One's set of preferences is coherent if they do not allow his gambling opponent to select options which leave him simultaneously happy (i.e., do not violate any of his preferences) and guaranteed to lose. Clearly, this is a weak requirement for the existence of a probability measure!

Research in psychology has responded to this work in two ways. Various studies have attempted to assess the descriptive validity of one or more of the axioms, especially those concerned with the preference relation. Others have, at least implicitly, found the axioms so compelling that their concern has been with the measurement of subjective probability distributions, assuming their existence. Much of the latter work is reviewed in Staël von Holstein

(1970) and Murphy and Winkler (1970).

If one accepts the arguments in the first part of this paper, then the psychological variable of subjective probability is defined for purposes of measurement by an appropriate set of qualitative empirical laws which taken together represent a theory about that variable. The axiom systems for relative likelihood and for preference are obvious candidates for such laws, and knowledge of their descriptive validity becomes important. Unless either of the systems is empirically valid or another can be found which is, measurement of subjective probability, in the present sense of that term, is impossible, and theories of that form are not useful for describing decision behavior under uncertainty.

However, as discussed earlier, even when a system is valid, interpretation of the derived scale values, or of the measurements, is still problematical. It will often depend on the more general theory in which the specific set of empirical laws is embedded. This presents interesting challenges for the theoretician and serious problems for the practitioner. We will consider the latter following the theoretical discussion.

Empirical validity of preference based axioms. Research since 1965 relevant to the descriptive validity of SEU theory has been reviewed by Rapoport and Wallsten (1972) and need not be discussed here. It will suffice to reproduce their conclusion:

> "...It seems then that the conflicting evidence pertaining to SEU theory is presently irreconcilable. Consequently, the basic experimental question should not be whether to accept or reject SEU theory as a whole, but rather to systematically discover the conditions under which it is or is not valid (Rapoport and Wallsten, 1972, p. 141)."

Clearly, when SEU theory does not describe subjects' choices under uncertainty, this does not mean that they do not experience uncertainty, nor even that they could not rank order the uncertainty associated with the various events. But it does mean, since as pointed out before we are simultanesouly studying elements differing in certain properties (here the subject faced with various gambles) and the device (also the subject) that makes the effects of these properties manifest, that either our empirical laws concerning the former are fundamentally wrong or our theory concerning the latter is wrong or incomplete. In either case the operational definition of subjective probability is inappropriate and fundamental measurement is rendered impossible.

Since there are situations in which SEU theory is valid (e.g., Tversky, 1967; Wallsten, 1971), it would seem that the qualitative empirical laws concerning how subjective probability and utility conjoin are not fundamentally wrong, but rather that they must be embedded in a more general theory which encompasses other variables as well and allows a priori prediction of when these other variables will affect the observed choice behavior. We are not prepared currently to offer such a theory, and can only suggest that it would represent a major step towards relating subjective probability to choice and understanding individual decision behavior.

Empirical validity of likelihood based axioms. To the best of our knowledge there have been no extensive empirical tests of the descriptive validity of these axiom systems, and with good reason, since they are probably virtually untestable for any interesting sample space. In view of our argument that subjective

probability, like other psychological variables, ought to be axiomatically defined, and the fact that elegant, compelling axiomatizations for a probability representation exist, this statement deserves amplification.

Although most axiom systems require infinite sample spaces, some exist for finite spaces (e.g., Kraft, Pratt, and Seidenberg, 1959; Fishburn, 1969). We will discuss the former.

Consider a nonempty sample space, or set, X, and a nonempty family of subsets of X, $\xi$, in which for every $A \varepsilon \xi$, $\bar{A} \varepsilon \xi$ ($\bar{A}$ is the complement of A), and for every A, $B \varepsilon \xi$, $A \cup B \varepsilon \xi$. $\xi$ is called an algebra of sets. In addition if for all $A_i \varepsilon \xi$, i=1,2,..., it is the case that $\bigcup_{i=1}^{\infty} A_i \varepsilon \xi$, then $\xi$ is called a $\sigma$-algebra or $\sigma$-field.

The ordering relation $\precsim$, "is not more likely than," is defined over $\xi$. Considering axiomatizations primarily for infinite X, if the triple $\langle X, \xi, \precsim \rangle$ satisfies five axioms, a probability function, $\underline{P}$, exists from $\xi$ into the real interval [0,1]. Various statements of the five axioms exist, some of which were referenced above. They differ from each other primarily in terms of the fifth axiom. The set given here is due to Luce (1967) and requires $\xi$ to be an algebra, not necessarily a $\sigma$-algebra. The five axioms state that for all A, B, C, D, $A_1 \ldots, A_i, \ldots \varepsilon \xi$:

(1) $\langle \xi, \precsim \rangle$ is a weak order.[6]

(2) $\phi \prec X$ and $\phi \precsim A$

(3) If $A \cap B = A \cap C = \phi$, then $B \precsim C$ iff $A \cup B \precsim A \cup C$.

(4) Archimedian: If $A_{\underline{i}} \cap A_{\underline{j}} = \phi$ for all $\underline{i} \neq \underline{j}$, $\phi \prec A \prec B$, $A_{\underline{i}} \sim A$ for all $\underline{i}$, then the set of positive integers $\underline{N} = \{\underline{n} \mid \bigcup_{\underline{i}=1}^{\underline{n}} A_{\underline{i}} \prec B\}$ is finite.[7]

(5) If $A \cap B = \phi$, $C \lesssim A$, $D \lesssim B$, then there exist, $C'$, $D'$, $E \varepsilon \xi$. Such that $E \sim A \cup B$, $C \sim C'$, $D \sim D'$, $C' \cup D' \subseteq E$ and $C' \cap D' = \phi$.

The first three axioms are clear, but the fourth and fifth require comment. The Archimedian axiom states that for an event A strictly less likely than B, but with $\underline{P}(A)>0$, only a finite number of disjoint events equally likely as A may be joined by the union operation in a subset which is also strictly less likely than B. Or in other words, if $\underline{P}(A)>0$, then any subset containing $\underline{i}+1$ "identical copies of A" has probability greater than a subset containing $\underline{i}$ copies. Luce's actual formulation of this axiom avoids the problems involved in there being an insufficient number of $A_{\underline{i}}$ and assures that the sequence of $\bigcup_{\underline{i}=1}^{\underline{n}} A_{\underline{i}}$ formed as $\underline{n}=1,2,\ldots,$ is bounded by X.

Axiom 5 is best paraphrased by Luce himself: "...if A and B are disjoint and dominate C and D respectively, then there are disjoint subsets of $A \cup B$ that are equivalent in probability to C and D (Luce, 1967, p. 781)."

Axioms 1-4 are necessary for a probability representation, but not sufficient. That is, given the existence of a representation, 1 through 4 will not be violated. But the converse is not necessarily true. An example of an ordering satisfying Axioms 1-3 (4 does not apply), but not admitting of a probability representation has been provided by Kraft, Pratt, and Seidenberg (1959). Thus, a fifth axiom is needed to limit the structures to which 1 through 4 will be applied. The one above from Luce (1967) is weaker than most in that it applies to some finite X as well as infinite X. Most others apply only to infinite X.

For example, the first four axioms by DeGroot (1970) are equivalent to those proposed by Luce (1967). But his system requires $\xi$ to be a $\sigma$-algebra, and the fifth axiom is that there exists a random variable which has a uniform distribution on the interval [0,1]. Then the original space is enlarged by composing it with that of the new random variable and the continuous, uniformally distributed random variable is used to establish the probability mapping from the original $\xi$ into the [0,1] interval.

It should be obvious why the systems shown here, which are typical, will not easily lend themselves to empirical study. The first three axioms will rarely fail, although with some ingenuity one could probably arrange failures of the first. Furthermore, if we were to test them on sets to which the system is restricted by axiom 5, we would in general be required to use infinite sets, and a complete test would be impossible. This latter point is relatively minor, since systematic failure of axioms 1, 2, or 3 would suffice to reject the probability representation for any set. However, if it is agreed beforehand that failure in any circumstance is unlikely, then success is not very interesting. An empirical test of axiom 4 would also be very unlikely to fail, but, as with any Archimedian axiom, any extensive test would require huge numbers of observations, and is quite infeasible.

Finally, although it may be easy to create circumstances in which Luce's axiom 5 is rejected, that does not imply the nonexistence of a probability representation. And in general it would be impossible to establish its success, since an infinite number of judgments would be required.

At first blush, DeGroot's axiom 5 would appear to be empirically sound; we might, for example, introduce a perfectly balanced spinner which randomly picks points around the unit circle.[8] But in light of the difficulties Davidson, Suppes, and Siegel (1957) had in establishing a binary event with subjective probability one-half, it is doubtful that a continuous variable with a uniform subjective distribution could be found.

If one can not test the axioms, one might look for other necessary conditions with more empirical content. Thus, Ellsberg (1961) has demonstrated, and Becker and Brownson (1964) have firmly substantiated, that ambiguity affects whether human subjective probability can be represented by a probability function. Specifically, using only binary choices in an informal experiment Ellsberg (1961) demonstrated that for many subjects $R_U \sim B_U$ and $R_A \sim B_A$, approximately, but $R_A \nsim R_U$ and $B_A \nsim B_U$, where $R_U$ is the event of drawing a red ball on a single trial from an urn unambiguously containing 50 red and 50 black balls, and $B_U$ is the event of obtaining a black ball from that urn on one draw. $R_A$ is the event of drawing a red ball in a single trial from an ambiguously constituted urn in which it is only known that there are 100 red and black balls. $B_A$ is the corresponding event for the black ball. Clearly that set of binary judgments can not be represented by a probability measure.

Undoubtedly, with sufficient skill and insight there can be discovered other necessary conditions for the representation and other variables which cause them to be violated. However, the first question is whether under appropriate conditions one can find

an interesting sample space which is not so large as to preclude an experiment, and for which linear inequalities based on binary judgments and the assumption of a probability measure are solvable. As far as we know such an experiment has not been done. If the results of such an experiment were to be positive, then we would know that at least under some circumstances the psychological variable of uncertainty would result in behavior with ordinal properties which are consistent with a probability measure. The fact that SEU theory has been occasionally validated provides only the mildest support for such a statement about human uncertainty, because the chance events in these experiments have rarely been more than binary.

The question whether there are situations in which human behavior is consistent with probability theory is important, but only of limited usefulness to psychology. It is important, because we would clearly like to know the conditions under which human and "rational" opinion agree. Furthermore a considerable amount of research is concerned currently with developing methods, such as proper scoring rules (Staël von Holstein, 1970), for measuring an individual's "true" subjective probability, and it would be well to know when that concept is well defined.

But the question is only of limited interest to psychology, because by itself it has the potential of explaining only a very narrow segment of decision behavior. The more fruitful psychological questions concern how uncertainty arises and is affected by other factors and how it combines with other psychological variables to determine choices. It is within theories designed to answer these questions that one would like to define and perhaps measure

uncertainty. That is one reason why SEU theory has been extensively studied.

Other Systems. Note that when the matter is put this way it is still required that the psychological variable of uncertainty be defined in terms of qualitative laws relating it to behavior. Assuming that the laws are such as to allow fundamental measurement, and there is no compelling reason why they should be, there is no requirement that the measurement conform to the rules of probability theory. The properties of the scales will depend on representation and uniqueness theorems, and their interpretation will depend on the nature of the general theory. Unfortunately there are very few theories concerning behavioral aspects of uncertainty which meet the criteria discussed here.

As an example of one that does at least in part, using the theory of simultaneous conjoint measurement (Krantz, et al, 1971; Krantz and Tversky, 1971), Wallsten (1972) presented a very general additive (and under some conditions distributive) model for revision of opinion in the presence of probabilistic information. This is the usual Bayesian probability revision task. Although the required experiments are complicated, the model is easy to state and will be given here in the distributive form, which applies when choosing between two alternative hypotheses, $X_i$ and $Y_j$, on the basis of a sample of $n$ identical events, E. These restrictions have both advantages and disadvantages which need not be discussed here. This form of the model states that

$$R(X_i, Y_j | nE) = f\{\phi_3(n)[\phi_1(E|X_i) - \phi_2(E|Y_j)]\}, \quad [1]$$

where the left side of the equation refers to a response with at least ordinal properties concerning the likelihood of $X_i$

relative to $Y_j$ on the basis of $\underline{n}$ events. $\phi_3$ is a real valued function whose domain is sample size and which refers to the subjective diagnostic value of that number of replicates of E. $\phi_1$ is a real valued function representing the subjective conditional likelihood of E given $X_i$. Given a particular E, the domain of $\phi_1$ is traced out by varying $X_i$. $\phi_2$ is the corresponding subjective function for E given $Y_j$.

Relying only on the ordinal properties of the data one may check the empirical validity of certain of the axioms necessary for the model. Assuming neither gross nor systematic failures, numerical representations for the scale values can be derived. However, the scales themselves are of very little import. Of greater interest is isolation of those variables which cause one or more of the axioms to fail and interpretation of relations between the scale values when the model does not fail. It is worth describing some of the experimental results to clarify these concepts.

Thus, Wallsten (1972) found the model to be reasonably accurate for eight of his 12 subjects and to fail in well defined ways for two others. When the model was valid the derived values for $\phi_3$ decisively showed that information samples of two identical events carried considerably less than twice the diagnostic weight of one of those events alone. $\phi_1$ and $\phi_2$ could be plotted against objective measures for the likelihood of E given $X_i$ and $Y_j$, respectively. The ratio of the slope of the least squares' best fitting line for $\phi_1$ to that for $\phi_2$ was invariant under all permissible transformations

of the two functions, and was greater than 1.0 for all subjects. This was interpreted either in terms of attentional factors or an aversion towards uncertainty, the latter interpretation being at variance with the former results mentioned.

Wallsten,[9] in a much more extended experiment scoring responses with the spherical scoring rule, but still relying only on ordinal properties of the data, found the model to hold for roughly the same percentage of subjects. The results concerning values of $\phi_3$ were replicated and extended to samples of size three. Looking at the effects of payoffs determined by the scoring rule on the relation between $\phi_1$ and $\phi_2$, the aversion of uncertainty interpretation was rendered much less likely, since the ratio of slopes was less than 1.0 for some subjects. If the attention interpretation is acceptable, then the effects of payoffs were to increase between subject variability in that factor, since there was considerably greater between subject variability in the slope ratios than there was in the previous study.

Finally, Wallsten and Delaney[10] using a Marschak bidding procedure, but still only the ordinal data properties, again found the model reasonably accurate for about two-thirds of the subjects. The results for samples of two or three identical events were again replicated. But now, upon analyzing samples of size three with two identical events and one different, the effect virtually disappeared. Here the two identical events were just about twice as diagnostic as that event appearing singly. Clearly the composition of the information sample affects the subjective values of its components.

In this experiment $\phi_1$ could be plotted directly as a function of $\phi_2$, without regard to other values. Biased payoff matrices resulted in differential utilities for the alternatives between which the subjects were deciding on the basis of the information. This differential utility slightly affected the derived values for $\phi_1$ and $\phi_2$ in a way consistent with the attention interpretation and with the subjects' probabilities of choice between the alternatives. The choice probabilities themselves were very strongly affected.

Certainly this work is of limited scope. For example, it says nothing about how the revised opinions combine with other factors to determine final decisions. And so far the work has been confined to simple alternative hypotheses and simple information samples. However, it does illustrate remarks in the first half of this paper and shows some of the difficulties in applying them to real data. Thus, an important unanswered question concerns those factors responsible for the model's success with some subjects and failure with others. Also, the concept of attention seemed to be useful above, but it still remains to make that concept more precise and bring it under better experimental control. This series of experiments also demonstrates that with appropriate designs ordinal data can be very rich.

An example of another investigation of the determinants of subjective probability relying primarily on ordinal data is that by Kahneman and Tversky (1972) and Tversky and Kahneman (1972). They showed that subjects' judgments of probability were strongly influenced by the degree of similarity between a sample and the

population from which it was drawn. In other circumstances the judgments were determined in part by the ability to recall past instances of the event in question. These demonstrations are potentially very important and it is to be hoped they will be replicated in rigorous experiments. Note that, as the authors themselves point out, they have not proposed a theory, but have provided qualitative data for which, if replicated, any theory will have to account.

## Applied Decision Theory

Conspicuous by its absence thus far has been any mention of data involving subjects' numerical estimates of probabilities. This is consistent with the philosophy of this paper, that the psychological variable of uncertainty is qualitative, like any other variable, and only when certain conditions are met can it be represented by numbers. And then the numbers must reflect the behavior induced by that variable, not be generated more or less independently of that behavior. However, probability (and utility) numbers are needed in the application of normative decision theory to real problems, and if the applications are to make any sense at all it is necessary that they reflect something of the opinions and values of the people involved.

It is primarily for this reason that many ingenious methods have been devised to assist subjects in generating probability numbers which reflect their "true" opinions. Chief among these methods is the family of strictly proper scoring rules, which requires subjects to give estimates identical to their subjective probabilities in order to maximize their subjectively expected gain. One of these rules is used routinely to score weather

forecasters (Murphy and Epstein, 1967). There are also various other methods involving fractionation of sample spaces and hypothetical experiments (see Pratt, et al, 1965; Staël von Holstein, 1970, Chapter 5; Winkler, 1967). But since it has not yet been established that human opinion can be properly represented by a probability function, and we concluded that at best rarely will that be the case, there would appear to be a basic conflict. We can suggest two possible solutions.

One of them is easier to state than to execute. It is desirable to have practice conform as closely to theory as possible. Thus initially one might attempt to calibrate responses obtained using a particular method with scale values derived from a theory valid in that situation. After the calibration has been painstakingly established the method could be used routinely. This is what is done when a spring scale is used instead of a pan balance to weigh objects. Of course, if calibrated scales can be established, they will not in general be probabilities, although this should only occasionally be a problem. However, considering the numerous variables which influence either verbal responses or the validity of a theory, such calibration procedures are quite unlikely to be successful.

One was tried recently by Wallsten (see footnote 9) to evaluate a strictly proper scoring rule. In that study subjects' probability estimates in the revision of opinion task were scored with a spherical scoring rule. It will be recalled that the distributive model for opinion revision provided a reasonably accurate description of most subjects' behavior, and scale values

could be derived from that model. Considering only those subjects for whom it was concluded that the distributive model was valid, if the maximization principle upon which the scoring rule rests also described their behavior, there should exist an identity transformation relating their probability estimates to the scale values combined according to the distributive model. Of those subjects whose behavior was judged to be described by the distributive model we selected the one whose responses were best described by the scoring rule and the one whose responses were most poorly described by the scoring rule, and present in Figure I the monotonic transformation of their estimates that best fits the model values. In this framework, subject 3PP was reporting his "true" opinion and 2PN was not. Clearly, it is not an easy matter to evaluate when the scoring rule is forcing subjects to be "honest," much less to establish a general calibration function.

The other solution is to treat applied decision theory in a manner similar to classical test theory. That, it appears to us, is possible and appropriate. That is, the question whether a set of responses reflects "true"opinion or not is irrelevant. The aim is to obtain reliable responses which are valid, i.e., correlate highly with other external criteria. The criteria might be other measures of uncertainty or they might be ultimately satisfactory decisions. Hoffman and Peterson (1972) put the matter well when they described the use of a proper scoring rule to assist the assessor in learning "what kinds of numbers are warranted by different states of knowledge (Hoffman and Peterson, 1972, p. 2)."

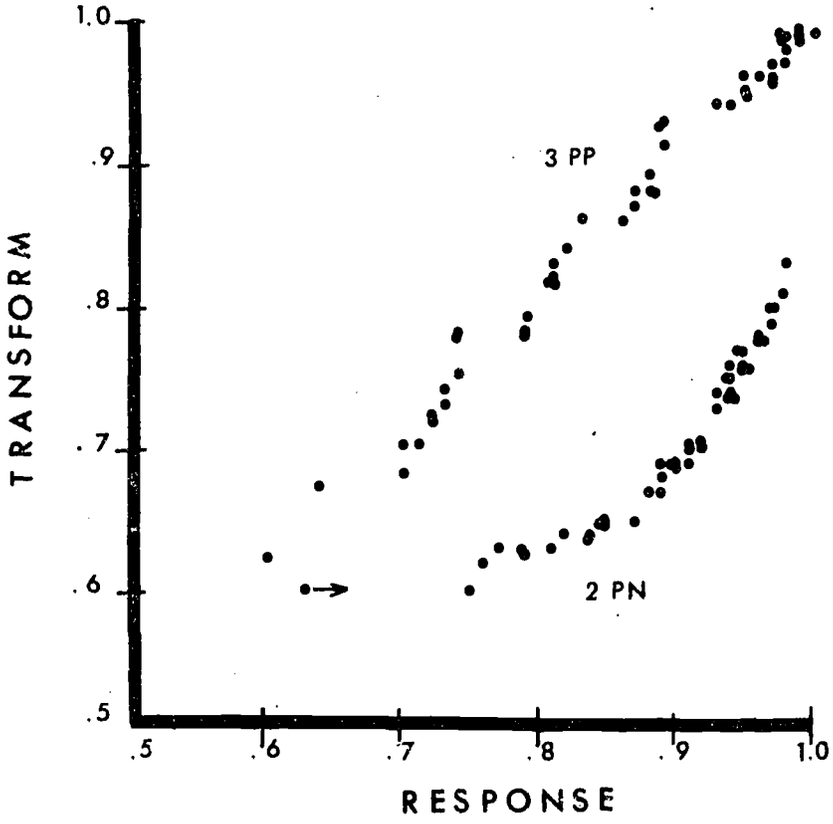


Figure I: Data from two subjects showing for each the monotonic transformation of the original probability estimates (response) that best fits the distributive model of opinion revision (Adapted from Wallsten, see footnote 9).

In essence, however, this is the approach already being followed by some researchers. Thus Murphy and Winkler wrote that "...perhaps the most important attribute of the (verbally estimated) probabilities is their 'validity', i.e., the association between the probability statements and the actual outcomes (Murphy and Winkler, 1970, p. 28)." Winkler and Murphy (1968) and Murphy and Winkler (1971) discuss further the association between an estimate and the occurence of an event on a single occasion and the correspondence between a collection of similar estimates and the appropriate relative frequencies. Alpert and Raiffa (1969) reported an experiment evaluating properties of verbally assessed probability distributions which were necessary for them to represent the actual distributions of various uncertain quantities. A brief general discussion of external validity appears in Chapter 4 of Staël von Holstein (1970).

Although logically prior to validity, the question of reliability does not appear to have been treated in this literature. There has been some worry about how to elicit probability distributions from various assessors in a manner designed to reduce between assessor variability (Winkler, 1968). But so far as we know, there has been no attempt to discover which assessment techniques result in the most highly correlated estimates within a single subject when he is in similar circumstances two or more times. This is clearly important, since unreliable estimates will not systematically correlate with any other reliable criterion and may, therefore, lead to low validity.

None of this is to claim that the theoretical approaches discussed above should hold no interest for practioners, nor that theorists should ignore non-laboratory problems. Neither is the case. A theory is relatively useless if it can not predict behavior outside the experimental laboratory, and the practioner will be considerably aided by knowing the psychology of the decision situation in which he is working.

References

Alpert, M. and Raiffa, H. A Progress Report on the Training of Probability Assessors, Harvard University Unpublished Manuscript, 1969.

Beach, L.R. and Phillips, L. Subjective Probabilities Inferred from Estimates and Bets. Journal of Experimental Psychology, 1967, 75, 354-359.

Becker, S.W. and Brownson, F.O. What Price Ambiguity? Or the Role of Ambiguity in Decision Making. The Journal of Political Economy, 1964, 72, 62-73.

Campbell, N.R. Foundations of Science: the Philosophy of Theory and Experiment. New York: Dover, 1957. (Originally published as Physics: the Elements. Cambridge: Cambridge University Press, 1920).

Davidson, P., Suppes, P., and Siegel, S. Decision Making: An Experimental Approach. Stanford: Stanford University Press, 1957.

de Finetti, B. La Prévision: Ses Lois Logiques, Ses Sources Subjective, Annales de l'Institut Henri Poinceré, 1937, 7, 1-68 (English translation in H.E. Kyburg and H.E. Smokler (eds.) Studies in Subjective Probability, New York and London: Wiley, 1964, pp. 93-158.)

DeGroot, M.H. Optimal Statistical Decisions. New York: McGraw-Hill, 1970.

Edwards, W. Conservatism in Human Information Processing. In Formal Representation on Human Judgment (ed. by B. Kleinmuntz) New York: Wiley, 1968, pp. 17-52.

Ellsberg, D. Risk, Ambiguity, and the Savage Axioms. Quarterly Journal of Economics, 1961, 75, 643-669.

Fishburn, P. Preference-Based Definitions of Subjective Probability. The Annals of Mathematical Statistics, 1967, 38, 1605-1617.

Fishburn, P. Weak Qualitative Probability on Finite Sets. The Annals of Mathematical Statistics, 1969, 40, 2118-2126.

Fishburn, P. Utility Theory for Decision Making. New York: Wiley, 1970.

Hoffman, J. and Peterson, C. A Scoring Rule to Train Probability Assessors. Engineering Psychology Laboratory Technical Report. University of Michigan: 1972.

Irwin, F.W. Intentional Behavior and Motivation: A Cognitive Theory. Philadelphia: Lippencott, 1971.

Kahneman, D. and Tversky, A. Subjective Probability: A Judgment of Representativeness. Cognitive Psychology, 1972, 3, 430-454.

Koch, S. (ed.) Psychology: A Study of a Science, vol. 2, New York: McGraw-Hill, 1959.

Koopman, B.O. The Axioms and Algebra of Intuitive Probability. Annals of Mathematics, 1949, 41, 269-292.

Kraft, C.H., Pratt, J.W., and Seidenberg, A. Intuitive Probability on Finite Sets. The Annals of Mathematical Statistics, 1959, 30, 408-419.

Krantz, D.H. A Survey of Measurement Theory in Mathematics of the Decision Sciences, part 2, vol. 12. Lectures in Applied Mathematics (ed. by G.B. Danzig and A.F. Weinott, Jr.) Providence: 1968, pp. 314-350.

Krantz, D.H. Measurement Structures and Psychological Laws. Science, 1972a, 175, 1427-1435.

Krantz, D.H. Measurement Theory and Qualitative Laws in Psychophysics. Michigan Mathematical Psychology Program Technical Report MMPP 72-3, University of Michigan: 1972b.

Krantz, D.H., Luce, R.D., Suppes, P., and Tversky, A. Foundations of Measurement, vol. I. Additive and Polynomial Representations. New York and London: Academic Press, 1971.

Krantz, D.H. and Tversky,A. Conjoint Measurement Analysis of Composition Rules in Psychology. Psychological Review, 1971, 78, 151-169.

Luce, R.D. Sufficient Conditions for the Existence of a Finitely Additive Probability Measure. The Annals of Mathematical Statistics, 1967, 38, 780-786.

Luce, R.D. What Sort of Measurement is Psychophysical Measurement? American Psychologist, 1972, 27, 96-106.

Luce, R.D. and Krantz, D.H. Conditional Expected Utility. Econometrika, 1971, 39, 253-271.

Luce, R.D. and Tukey, J.W. Simultaneous Conjoint Measurement: A New Type of Fundamental Measurement. Journal of Mathematical Psychology, 1964, 1, 1-27.

MacCorquodale, K. and Meehl, P.E. On a distinction between hypothetical constructs and intervening variables. Psychological Review, 1948, 55, 95-107.

Marx, M.H. Theories in Contemporary Psychology. New York and London: McMillian, 1964.

Murphy, A.H. and Epstein, E.S. Verification of Probabilistic Predictions: A Brief Review. Journal of Applied Meteorology, 1967, 6, 748-755.

Murphy, A.H. and Winkler, R.L. Scoring Rules in Probability Assessment and Evaluation. Acta Psychologica, 1970, 34, 273-286.

Murphy, A.H. and Winkler, R.A. Forecasts and Probability Forecasts: Some Current Problems. Bulletin of the American Meteorological Society, 1971, 52, 239-247.

Pratt, J.W., Raiffa, H. and Schlaifer, R. Introduction to Statistical Decision Theory (Preliminary Edition). New York: McGraw-Hill, 1965.

Rapoport, A. and Wallsten, T.S. Individual Decision Behavior. Annual Review of Psychology, 1972, 23, 131-176.

Savage, L.J. The Foundations of Statistics. New York: Wiley, 1954.

Staël von Holstein, C.A.S. Assessment and Evaluation of Subjective Probability Distributions. Stockholm School of Economics: 1970.

Tolman, E.C. Operational Behaviorism and Current Trends in Psychology. In Proceedings of the 25th Anniversary Celebrating the Inauguration of Graduate Studies, University of Southern California, Los Angeles, 1936, pp. 89-103.

Tversky, A. Additivity, Utility, and Subjective Probability, Journal of Mathematical Psychology, 1967, 4, 175-201

Tversky, A. and Kahneman, D. Availability as a Determinant of Frequency and Probability Judgments. Oregon Research Institute Technical Report 12 No. 1: 1972.

Wallsten, T.S. Subjects Probability Estimates and Subjectively Expected Utility theory: Their Relationship and Some Limitations. Psychometric Laboratory Report No. 85. University of North Carolina: 1970.

Wallsten, T.S. Subjectively Expected Utility Theory and Subjects Probability Estimates: Use of Measurement-Free Techniques. Journal of Experimental Psychology, 1971, 88, 31-40.

Wallsten, T.S. Conjoint-Measurement Framework for the Study of Probabilistic Information Processing. Psychological Review, 1972, 79, 245-260.

Winkler, R.L. The Assessment of Prior Distributions in Bayesian Analysis. Journal of the American Statistical Association. 1967, 62, 776-800.

Winkler, R.L. The Consensus of Subjective Probability Distributions. Management Science, 1968, 15, B-61 - B-75.

Winkler, R.L. and Murphy, A.H. 'Good' Probability Assessors, Journal of Applied Meteorology, 1968, 7, 751-758.

Wise, J.A. Origins of Subjective Probability. Acta Psychologica. 1970, 34, 287-299.

Villigas, C. On Qualitative Probability σ-Algebras. Annals of Mathematical Statistics, 1964, 35, 1787-1796.

Villigas, C. On Qualitative Probability. American Mathematical Monthly, 1967, 74, 661-669.

Von Neuman, J. and Morganstern, O. Theory of Games and Economic Behavior. New York: Wiley, 1944.

Footnotes

[1]This work was supported in part by a PHS Grant No. MH-10006 from the National Institute of Mental Health and in part by a University Science Development Program Grant No. GU-2059 from the National Science Foundation. Both Harold E. Delaney and Amnon Rapoport critically read an earlier draft of this paper and their comments are gratefully acknowledged.

[2]read "if and only if".

[3] $\prec$ is defined from $\preceq$ in the same way that < is defined from $\leq$. That is, $x \prec y$ iff $x \preceq y$ and not $y \preceq x$.

[4]It is interesting to note that the concept of <u>intervening variable</u>, first introduced in psychology in 1936 by Edward C. Tolman, is similar in many respects to the concept of a measurable variable discussed above. Tolman (1936) formalized this notion so that mentalistic properties such as <u>expectancy</u>, <u>valence</u>, or <u>representation</u> could be operationally defined and take their place in rigorous, behavioristic psychological theory. Intervening variables were operationally defined in terms of the relations they predicted between independent variables and dependent behavior.

For many years the concept of an intervening variable was both influential and controversial in research on learning and motivation (see Koch, 1959, and Chapter 5 of Marx, 1964, for a glimpse of the later stages of debate.) A definition that was central in the debate, and appears to capture the essence of intervening variables, was offered by MacCorquodale and Meehl (1948) when they wrote:

> "...First, the statement of such a concept does not contain any words which are not reducible to the empirical laws. Second, the validity of the empirical laws is both necessary and sufficient for the 'correctness' of the statements about the concept. Third, the quantitative expression of the concept can be obtained without mediate inference by suitable groupings of terms in the quantitative empirical laws" (MacCorquodale and Meehl, 1948, p. 107)."

Note the first two characteristics outlined by MacCorquodale and Meehl, that the concept is defined only in terms of empirical laws, and that their validity is necessary and sufficient for the statements about the concept to be correct. This provides an excellent description of the approach advocated in the present paper.

The third characteristic, concerning quantification of the concept, differs considerably from the present approach. A feature of using measurement foundations as empirical laws is that the laws themselves are qualitative, not quantitative, but taken together they allow the concept to be expressed numerically.

[5]The usual set theoretic notation is used here and throughout the rest of the paper:

$a \in A$ means "a is an element of A"

$A \cap B$ means "A intersect B"

$A \cup B$ means "A union B"

$A \subset X$ means "A is a subset of X"

[6]The term "weak order" means that the ordering is connected, i.e., either $A \precsim B$ or $B \precsim A$; reflexive, i.e., $A \precsim A$; and transitive, defined earlier.

[7]This formulation is different from Luce's (1967). However, the simplicity gained for the present discussion is well worth the subtle, although important, conceptual problems introduced. The interested reader is referred to Luce (1967).

[8]DeGroot (1970, p. 77) suggests that the statistician might imagine such an ideal device for the purpose of comparing relative likelihoods of other events $A \varepsilon \xi$.

[9]Manuscript in preparation entitled "A simultaneous evaluation of a conjoint-measurement model for revision of opinion and a strictly proper scoring rule."

[10]Manuscript in preparation entitled "The effects of a biased payoff matrix on probabilistic information processing."

L. L. Thurstone Psychometric Laboratory Report Series, 1972-1973.

103. Barik, H.C. Simultaneous interpretation: temporal and quantitative data. January, 1972.

104. Young, F.W. POLYCON Users Manual. A FORTRAN-IV program for polynomial conjoint analysis. February, 1972.

105. Johnson, E.S. and Baker, R.F. Some effects of computerizing an experiment in human problem solving. February, 1972.

106. Conger, A.J. Estimating profile reliability and maximally reliable composites. February, 1972.

107. Young, F.W. and Dobson, R. On the perception of a class of bilaterally symmetric forms. July, 1972.

108. Young, F.W. Polynomial conjoint analysis: Some second-order partial derivatives. July, 1972.

109. Conger, A.J. and Conger, Judith C. Maximally reliable profile dimensions for the WISC. August, 1972.

110. Rapoport, A., Kahan, J.P., and Stein, W.E. Decisions of Timing in pure conflict III: the silent duel with equal accuracies. October, 1972.

111. Rapoport, A. and Stein, W.E. Minimum risk decision policies in multistage betting games. October, 1972.

112. Kahan, J.P. and Rapoport, A. Test of the bargaining set and kernel models in three-person games. December, 1972.

113. Horowitz, A.D. and Rapoport, A. Test of the kernel and two bargaining set models in four- and five-person games. December, 1972.

114. Fillenbaum, S. Problems in the use of similarity data for the study of semantic structures. February, 1973.

115. Wallsten, T.S. The psychological concept of subjective probability: A measurement-theoretic view. February, 1973.