Reprinted from PHILOSOPHY OF SCIENCE
Vol. 22, No. 2, April, 1955
Printed in U.S.A.

OUTLINES OF A FORMAL THEORY OF VALUE, I

DONALD DAVIDSON, J. C. C. McKINSEY, AND PATRICK SUPPES[1]

1. Introduction. Contemporary philosophers interested in value theory appear to be largely concerned with questions of the following sort:

What is value?

What is the meaning of the word 'good'?

Does the attribution of value to an object have a cognitive, or merely an emotive, significance?

The first question is metaphysical; to ask it is analogous to asking in physics:

What is matter?

What is electricity?

The others are generally treated as semantical questions; to ask them is analogous to asking in statistics:

What is the meaning of the word 'probable'?

Does an attribution of probability express an objective fact, or merely a subjective attitude?

Physics has advanced, however, without answering metaphysical questions; statistics has advanced without answering semantical questions; and it is our opinion that similar substantial progress in the theory of value can be made independent of metaphysics and semantics. As in other disciplines, theory in the domain of value can proceed along formal lines without waiting upon a solution to the grand questions; indeed even the most modest constructive progress might result, here as elsewhere, in putting what have been considered the fundamental problems in a new light.

We take it as the general function of formal value theory to provide formal criteria for rational decision, choice and evaluation. Our conception of this aspect of value theory is in one way similar to Kant's, for like him we believe it possible to state in purely formal terms certain necessary conditions for rationality with respect to value. Unlike Kant, however, we do not suggest that any particular evaluations or value principles can be derived from purely formal considerations. Value theory, as here conceived, is associated with another venerable, and at present rather unfashionable, tradition, for it seems to us that there is a sense in which it is perfectly correct to say that just as logic can be used to define neces-

[1] The unexpected death of Professor J. C. C. McKinsey after the completion of an earlier and much shorter draft of the present paper means that although he played a major part in formulating the fundamental ideas he cannot be held accountable for any of the shortcomings of the final version.

We are indebted to Professor David Sachs and Dr. Leo Simons for helpful comments on the first draft. An early version of the paper was read by Professor McKinsey at the University of California at Los Angeles on May 23, 1953, and a useful critique was given by Dr. Alexander Sesonske. We have benefited from numerous discussions with members of the Value Theory Project at Stanford University. Most of the content of this paper plus some additional material was issued as Report No. 1 of the Value Theory Project, 10 February 1954.

sary formal conditions for rational belief, so it is a use of value theory to define necessary formal conditions for rational choice.

The subject matter of the theories to be considered here consists in the events, acts, objects, or goals to which value is attributed, and such ordering relations as preference, equivalence in value, and difference in value. By rejecting linguistic entities like words and sentences as the sole or chief subject matter of value theory, we turn back, in one more way, to an earlier view of the philosophy of value.

In this and two subsequent papers ((I), (II), (III)) we shall give the outlines of a formal method for dealing with certain questions in value theory. In (I) we attempt an explication of the notion of a rational preference pattern, and discuss the problem of measuring preference. In (II) we propose to clarify the concept of a value principle: on the basis of notions analyzed in (I) we lay down some conditions of adequacy for value principles; then we consider the formal aspects and logical consequences of accepting a number of alternative value principles; the paper will conclude with an investigation of the mutual consistency and relative priority of sets of value principles. In (III), where the fundamental problem discussed is the determination of social and political principles to be used for the organization of society and the construction of a set of social values, we consider two main questions: first, supposing the preference patterns of the individuals composing a social group are given, to find principles for determining a preference pattern for the social group as a whole; second, to consider the consequences for the determination of social values of the acceptance of given substantive value principles.

We wish here to make some remarks about the axiomatic method of which we shall make large use in these papers. It is generally recognized that a set of axioms provides *implicit* definitions of the primitive terms which appear in it. It is equally important to realize that any axiomatization may be formulated as an explicit definition of a set-theoretical property. Thus the following two axioms (for *quasi-orderings*) might be said to give an implicit definition of the primitives K and R:

A1. The relation R is reflexive in the set K;

A2. The relation R is transitive in the set K.

But we could also formulate this axiomatization as an explicit definition of the set-theoretical property of being a quasi-ordering:

χ *is a* QUASI-ORDERING $=_{\mathrm{df}}$ *there exists a set* K *and a binary relation* R *such that* χ *is the ordered couple* $\langle K, R\rangle$ *and* R *is reflexive and transitive in the set* K.

In some cases it is possible to regard such set-theoretical definitions as partial explications. Thus Peano's axioms for arithmetic can be viewed as a partial explication of the notion of natural number. His axioms isolate certain formal properties of numbers although they do not tell us what numbers are. (A more complete explication was provided by Dedekind and Frege's construction of the natural numbers as certain sets, or properties, of sets.) In this manner the partial explication provided by an axiomatization can pave the way for a more complete explication.

It is an important feature of a theory axiomatically formulated that the primitive and defined notions which are implicitly or explicitly defined *within* the theory have a meaning only as part of the theory. The axiomatic approach can therefore eliminate for value theory as it has for other theories the need to depend on the inadequate and inflexible resources of ordinary language, and can help overcome the frustration that results from an attempt to explicate the basic value concepts in isolation from a coherent theory.

2. Rationality and Preference. Let us consider a set K of alternatives which are to be ranked in preference. It is fairly well agreed among philosophers that it is reasonable to speak of such a preferential ranking of a set of alternatives[2], and there seems even to be substantial agreement regarding the formal properties of the ranking relation—thus many philosophers would probably agree with Perry ([19], p. 636) that preference is transitive and asymmetric. For the sake of example, suppose that the set K comprises just three alternatives, timocracy, oligarchy and democracy. We will express the fact that x is preferred to y by writing 'x P y'. The following six patterns of preference will then be consistent with the assumption that P is transitive and asymmetric:

timo P olig olig P demo timo P demo	olig P timo timo P demo olig P demo	demo P olig olig P timo demo P timo
olig P demo demo P timo olig P timo	demo P timo timo P olig demo P olig	timo P demo demo P olig timo P olig

On the analogy of the explicit definition of a quasi-ordering given above, we might now define a Rational Hegemonic Ranking as any ordered couple whose first member is the set K consisting of timocracy, oligarchy and democracy and whose second member is one of the above six relations.

It is a limitation of a Rational Hegemonic Ranking that each alternative must be ranked above or below every other. In order to admit the possibility that two or more alternatives are equal in preferential status, let us introduce the relation E which holds between two alternatives when they are equivalent in preference. It seems natural to consider E transitive and symmetrical. The possible patterns of preference will now include such rankings as the ones indicated by the following sets of statements:

demo E timo timo E olig demo E olig timo E demo olig E timo olig E demo	olig P timo timo E demo demo E timo olig P demo

[2] See, for example, [13], p. 490; [19], p. 606 ff.; and [23], p. 124.

Without some further axiom, we can also see that there is nothing to exclude such a pattern as the following:

demo E timo
timo E demo
demo P timo
timo P olig
demo P olig

These cases are ruled out by an axiom which has the effect of asserting that if x and y are alternatives, then no more than one of the following: x P y, y P x, x E y. (It now becomes redundant to stipulate the asymmetry of P.)

It has been assumed in the discussion and examples that every pair of alternatives in K is related either by P or by E. If we want to be able to say that in a Rational Hegemonic Ranking every alternative is assigned some rank in relation to the other alternatives, we must strengthen the previous axiom to provide also that if x and y are alternatives, then at least one of the following: x P y, y P x, or x E y. (This makes it unnecessary to stipulate the symmetry of E.)

Let us now generalize the ideas we have been considering to apply to *any* set K of alternatives and any binary relations P and E whose fields are contained in K.[3]

Definition 1. The ordered triple $\langle K, \mathrm{P}, \mathrm{E}\rangle$ *is a* RATIONAL PREFERENCE RANKING *if and only if:*

P1. *The relation* P *is transitive;*

P2. *The relation* E *is transitive;*

P3. *If x and y are in K, then exactly one of the following:* x P y, y P x, x E y.

From Definition 1 we see that *all* ordered triples which satisfy certain conditions are RPR's. Among examples of RPR's, then, we would find Rational Hegemonic Rankings, and also such triples as the following:

K	P	E
Economic systems	preferred-to	equal-in-preference-to
Automobiles	more-expensive-than	the-same-price-as
Pains	more-painful-than	just-as-painful-as
Mountains	higher-than	the-same-height-as
Positive integers	less-than	equal-to

Obviously many things are RPR's which have nothing to do with preference or value; this points up the fact that the definition at best provides necessary conditions for rational ranking. It is the general intention of the remarks which

[3] Some gain in perspicuity has been made at the expense of formal precision in this and the definitions to follow. A more precise version of Definition 1, on the model of the definition of a Quasi-Ordering given in Section 1, is:

Definition 1. $\mathfrak{L}$ *is a* RATIONAL PREFERENCE RANKING $=_{df}$ *there exists a set K, and binary relations* P *and* E *the field of each of which is contained in K, such that* $\mathfrak{L}$ *is the ordered triple* $\langle K, \mathrm{P}, \mathrm{E}\rangle$ *and:*

A1. *The relation* P *is transitive;*

A2. *The relation* E *is transitive;*

A3. *If x and y are in K, then exactly one of the following:* x P y, y P x, x E y.

follow to show that the definition does provide such necessary conditions, and that it can therefore serve as a basis for a partial explication of rationality in the field of value.

It will be noticed, to begin with, that Definition 1 sets no limitations on the kind of entities which are to be ordered. Set K may contain objects, properties, experiences, events, commodities, acts, ends-in-view, purposes, world states or cooking apples. Nor is there necessarily any reason to consider the members of K as exhaustive alternatives or as mutually exclusive, although these possibilities are not ruled out. So far as such questions are concerned the definition is neutral.

In another important respect the definition is again neutral, and this point is probably worth emphasizing in order to prevent misunderstanding. The definition imposes limitations only on the patterns of preference and equivalence among sets of alternatives, but it says nothing whatsoever about when, or under what conditions, one alternative is preferred or equivalent to another. The definition of an RPR is as indifferent to particular rankings between pairs as the theory of the syllogism is to the truth of individual premises. And just as the universal validity of syllogistic inference implies nothing about the status of premises, so the universal applicability of the definition of rational ranking does not imply, for example, that value judgments are objective, or absolute, or timeless. This is not to say that such a definition throws no light on the status of value judgments, but only that it does not imply any particular values, any particular standard of value, or any particular view about standards of value.

To deny that the definition implies any particular judgments of preference or equivalence is not to deny the fact that it is intended to perform a normative function, however. If a, b and c are alternatives, for example, and a P b and b P c, then in an RPR it follows that a P c. This cannot be taken simply as a description of how people order their preferences; on any normal interpretation of preference, we would expect to find cases where people preferred a to b, b to c and c to a. By refusing to call such a pattern of preferences *rational* we in effect establish a formal condition for rationality. It should be noted that its normative character lies in the way the definition is used, not in anything it directly says. It would be misleading to interpret the definition as saying that if a P b and b P c, then a *should* be preferred to c. The definition allows us to deduce no normative statements from non-normative premises; it does not *say* what we should believe, prefer, or choose. An analogy may help illustrate this point. Suppose that 'X', 'Y' and 'Z' are names of sentences. If X is a consequence of Y, and Y a consequence of Z, then X is a consequence of Z; this is a truth of logic. Logic does not say that we should reason in accordance with this truth, nor that if we believe the antecedent of the sample truth, we should believe the consequent. But we can *use* this truth of logic to explain what we mean in part by reasoning rationally, and when we do this we use a truth of logic normatively.

The words 'rational' and 'preference' serve to emphasize the use of Definition 1 in stating formal conditions for rational patterns of preference. It is possible that this aspect of the definition would be made clearer for some people, however, if the defined entities were called, say, 'Comparative Value Rankings,' P were interpreted as the relation better-than, of-greater-value-than, or preferable-to, and E as the relation of-equal-value-with. Such re-labeling, while leaving the intended use of the definition untouched, would call attention to its function of displaying certain formal properties of basic comparative value concepts.

Now we may consider the three axioms of the definition with an eye to their claim to embody necessary conditions for a rational pattern of preferences. We hope the reader who (justifiably) feels that here we treat a subtle and difficult topic with cavalier brevity will bear in mind that our primary aim is the programmatic one of persuading philosophers that our method, if not our results, is useful; and that the real vindication of constructive formal systems is apt to lie more in their ability to reformulate old problems than in their close fit with the concepts and categories of common sense.

The first axiom provides that if x, y and z belong to a rationally ranked set of alternatives, and x is valued above y and y above z, then x is valued above z. Perhaps the most common objection to this condition is that while x P y and y P z might hold at time t_1, z P x might with reason hold at time t_2. It seems easy to answer this objection, however, for if valuations change with time, this may be considered a change in the alternatives themselves: thus x at t_1 is simply a different alternative from x at t_2. If we consider that y and z remain unchanged with time, and call x at t_1 'x_1', and x at t_2 'x_2', then the above apparent exception to Axiom P1 becomes the innocuous situation: x_1 P y, y P z, z P x_2. Many other apparent objections to Axiom P1 can be handled in an analogous way.

A more serious difficulty seems to be raised by the following example: Mr. S. is offered his choice of three jobs by a cynical department head (never mind what department): He can be a full professor with a salary of $5,000 (alternative a), an associate professor at $5,500 (alternative b) or an assistant professor at $6,000 (alternative c). Mr. S. reasons as follows: a P b since the advantage in kudos outweighs the small difference in salary; b P c for the same reason; c P a since the difference in salary is now enough to outweigh a matter of rank. What arguments can be given to show that this is an irrational set of preferences? Obviously the reasons for each of the paired comparisons may be good ones (bearing in mind that there may be good reasons for wrong judgments). The following considerations appear to indicate, however, that the reasons could never be good enough to justify acceptance of such a set of preferences. Presumably an important function of an RPR is to serve as the basis for rational choice. The obvious principle would appear to be this: a rational choice (relative to a given set of alternatives and preferences) is one which selects the alternative which is preferred to all other alternatives; if there are several equivalent alternatives to which none is preferred, then any of these is selected. In short, a rational choice is one which selects an alternative to which none is preferred. But it is clear that the set of Mr. S's preferences makes a rational choice impossible, for whichever alternative he chooses there will be another alternative which is preferred to it.

We may imagine a scene in which the point becomes obvious.[4] The department head, advised of Mr. S's preferences, says, 'I see you prefer b to c, so I will let you have the associate professorship—for a small consideration. The difference must be worth something to you.' Mr. S. agrees to slip the department head \$25. to get the preferred alternative. Now the department head says, 'Since you prefer a to b, I'm prepared—if you will pay me a little for my trouble—to let you have the full professorship.' Mr. S. hands over another \$25. and starts to walk away, well satisfied, we may suppose. 'Hold on,' says the department head, 'I just realized you'd rather have c than a. And I can arrange that—provided . . .'

Many objections which may be raised to the second axiom will be similar to those raised to the first and they can be answered, if at all, by considerations of the same sort. For example, it may be argued that under certain circumstances x_1 is clearly more valuable than x_n although there is no direct way to detect a difference in value between the adjacent members of the sequence $x_1, x_2, \cdots, x_n$, so that $x_1 \mathrm{E}\, x_2, x_2 \mathrm{E}\, x_3, \cdots, x_{n-1} \mathrm{E}\, x_n$. Since the transitivity of E would imply in this case that $x_1 \mathrm{E}\, x_n$, and hence not $x_1 \mathrm{P}\, x_n$, it may be said that Axiom P2 unnecessarily demands infinite discrimination between alternatives. But the man who believes the adjacent members of the sequence equivalent will exchange x_1 for x_n in a series of equal swaps, and this is inconsistent with the belief that x_1 is better than x_n. Definition 1 does not imply that if we believe we can see a difference between x_1 and x_n that then, to be rational, we must be able to see a difference between at least two adjacent members of the sequence $x_1, x_2, \cdots, x_n$; it provides merely that if x_1 is held more valuable than x_n, it must rationally be held that there is some difference in value between at least two adjacent members of the sequence.

The third axiom stipulates that every two alternatives in a rational ranking must be comparable. This may be thought excessive, since it is at least possible that alternatives belonging to different categories are simply incomparable (does it make sense to ask whether the Venus de Milo is better than *Don Giovanni*?). But the axiom does not imply that all alternatives whatsoever must be comparable, only those in the given set K; the axiom may therefore be considered as limiting the set K to alternatives which *are* comparable. It will be noted that it is possible for two sets K_1 and K_2 of alternatives, each ranked rationally (in the sense of Definition 1), to overlap without necessarily implying that every member of K_1 is comparable with every member of K_2. Thus it might be held that taking a wife (a) is better than taking a mistress (b), and taking a mistress better than buying a yacht (c); but that taking a wife and buying a yacht were incomparable. In this case, a, b and c could not belong to the same RPR, but a and b might belong to one and b and c to another.

We have so far considered the normative aspects of the definition of an RPR. Now we may speak briefly of its potentialities for description and prediction. Economists, anthropologists, sociologists, psychologists and others need a conceptual framework within which questions about patterns of preference and

[4] We owe the inspiration for this example to Dr. Norman Dalkey of the Rand Corporation.

valuation of individuals, cultures, economic classes, and so forth may intelligently be asked. Empirical findings which are organized by appropriate schemes will also be useful in making predictions. Economists have found in fact that such axiomatizations as Definition 1 (and its more sophisticated modifications to be considered subsequently) can serve as a basis for prediction of behavior in certain areas.

Before such a question as 'Is Mr. R's preference ranking rational?' can be asked or answered in a meaningful way, it is clear that an empirical interpretation of an RPR must be given. It is not our purpose to argue for any particular interpretation. It seems reasonable to suppose, in fact, that different interpretations will be suitable to different problems and domains. The economist may, for certain purposes, find it satisfactory to interpret 'x P y' in such a way that a man's willingness to work longer for x than for y is conclusive evidence that x P y; the psychologist may decide that for an individual faced with a set of cards of different colors, he will consider the remark 'I prefer card x to card y' suitable evidence that x P y.

For some purposes, it may serve to interpret preference and equivalence in terms of actual particular choices, or particular statements about preferences. But then the following difficulty may arise. Consider again the case of Mr. S. who chose a over b, b over c and c over a. It may be argued that since the choices necessarily occur at different times, each particular choice may spring from a momentary preference ranking which is rational. Thus if we take unique events like choices as identical with preferences, we can never prove what a man's preference ranking is over more than two alternatives.

For this and other reasons, it would seem generally more plausible to interpret preference and equivalence of preference as *dispositions* which characterize individuals (or firms, families, countries) over a period of time. In this case, we may consider particular choices as *evidence* for the disposition, but not identical with it. Mr. S's choices are then evidence, so far as it goes, that his preference ranking is not rational; but we would reconsider this verdict if we learned that he had changed his mind about the relative ranking of a and c after his first two choices.

Just what is to count as evidence for the dispositions of preference and equivalence is a matter to be decided in terms of many considerations. Whether to admit only behavioristic evidence, for example, or to count the data of introspection, would depend among other things on the character of the alternatives, the structure of the relevant laws of psychology, the testing procedures available, and the predictive power of the results.

It would be an error to hold that the call for suitable empirical interpretation of the axiomatization of Definition 1 (and of the axiomatizations to follow) is relevant only to its descriptive use. Before we can say that a man's beliefs are logically consistent or inconsistent, we require an interpretation of belief. In the same way, before we can characterize the pattern of a man's preferences or judgments of value as rational or irrational, we need an interpretation of preference or judgment of value. This is true even when the preferences under consideration are our own. A good deal of the content of moral philosophy is concerned with the achievement of a rational pattern of preferences.[5] Psychoanalysis may be considered, in part, as a technique for the discovery of preferences, and the resolution of irrational into rational patterns of value.[6] Thus both moral philosophy and psychoanalysis must attempt to supply workable empirical interpretations of preference.

It is not however part of the formal theory of value (as conceived in these papers) to make detailed decisions concerning empirical interpretation. In particular it should be remarked that the possibility of establishing necessary formal conditions for patterns of preferences has no direct relation with the question whether preference is best interpreted as belief (as some cognitive theories of value hold) or as a non-cognitive attitude (as emotive theories of value hold).[7]

3. The Measurement of Preference. For many questions of value theory it is sufficient to use Rational Preference Rankings. However, there are important and significant contexts in which it would be useful to have at hand a stronger measurement of preference or comparative value than is given by an RPR. Many applications of value theory in statistical, economic and political theory may be adduced to show the usefulness of such a measure. Certain of these applications will be discussed in (II) and (III). For the present the relevance of such a stronger measure to the behavior of a rational man may be illustrated by a simple example.

Suppose that Wright is a congressman interested in federal aid for education. He has introduced a bill authorizing such aid; the bill includes a provision that money shall go only to school districts which agree not to practice racial discrimination. When the bill comes on the floor, it becomes clear that it would pass if the anti-segregation clause were dropped. On the other hand, Wright estimates the bill has only one chance in three of passing with the anti-discrimination clause. Wright ranks the three possible outcomes, in order of preference, as follows: *a*) passage of the full bill; *b*) passage of the bill without the anti-discrimination clause; *c*) defeat of the bill. It is obvious that his preference ranking and his estimate of the chances for each of the alternatives cannot alone serve as a basis for decision. Should he press for the full bill with a substantial chance of defeat (Action 1), or accept the weakened bill which has a practical certainty of passing (Action 2)? What is needed is clearly some measure of how *much more* Wright values *a* than *b* and *b* than *c*. If, for instance, dropping the anti-discrimination clause matters very little, Wright will not hesitate to present the bill in its curtailed form (that is, he chooses Action 2). But if he sets great store on the anti-discrimination condition, he may care very little whether the bill passes or

[5] Examples from recent literature may be found in [19], Chapters VII and XIII, and [9], Part Three, Section II.

[6] See [25], p. 105.

[7] Here, and throughout these papers, the words and phrases 'preference,' 'is preferred to,' 'equivalence,' 'equal in preferential status,' as well as the symbols 'P' and 'E', will be used with the understanding that they allow for the large range of interpretations indicated in this section. We have retained the word 'preference' (and its cognates) partly from lack of a better or more neutral term, partly from deference to the economic literature which inspired the formal developments here outlined.

not without it. In this case, he chooses Action 1. Suppose next Wright judges that b is as much better than c as a is than b (b is "halfway" between a and c in value). Then he will choose Action 2. We can represent his reasoning as follows: assign the number 2 to a, the number 0 to c, and the number 1 to b (since what these numbers measure is only the *relative* values of a, b, and c, the actual numbers chosen here are arbitrary except for the relative magnitude of their differences). Wright can now calculate the relative merits of the two actions. Since Action 2 ensures getting b, and b has been assigned a relative value of 1, Action 2 is worth 1. Action 1 entails one chance in three of getting a with its relative value of 2, and two chances in three of getting c with its relative value of 0. It is natural to evaluate Action 1, then, as worth $(1/3)\cdot 2 + (2/3)\cdot 0 = 2/3$. The relative scores of Actions 1 and 2 clearly determined Wright's choice. Similar calculations will show that as long as the ratio of the difference in value between a and b, and b and c, is less than 2 (that is, $\frac{n(a) - n(b)}{n(b) - n(c)} < 2$), Wright will take Action 2, while if this ratio is greater than 2, he will take Action 1. This way of calculating simply means that, having decided the relative merits of the alternatives, Wright tempers the weight he will assign each of them in a practical situation by his opinion of its probability. A course of action which has only one chance in three of getting him his first choice is worth only a third as much as a course of action which is sure to obtain it; when the course of action means one chance of getting a to two chances of getting c, it should be weighed at a third the value of a and two-thirds of the value of c.

Congressman Wright's particular problem illustrates one kind of need for some stronger form of measure than that of ranking. But philosophers have long appreciated this need in dealing with preferences, values, interests, desires, pleasures and pains and other value phenomena. It would seem especially desirable for philosophers with certain naturalistic views of value to find a positive solution: for example, for utilitarians (Bentham, Mill, Sidgwick), those with interest theories (Perry), and pragmatists (Dewey, Lewis). On the other hand, a large number of philosophers, including some of these naturalists themselves, believe that it is not possible to measure preference in a stronger sense than ranking. Unfortunately the discussion of this problem in the philosophical literature has been uniformly inadequate and confused. On the one hand the possibility of an articulate theory of the measurement of value is rejected for mistaken reasons; on the other the positive proposals are often unnecessarily naive.

The following quotations are typical of the skeptical attitude towards the possibility of a strong form of measurement of value. Wheelwright ([24], p. 87) asserts:

> We may on a particular occasion *prefer* reading a book to taking a walk: the former, then, we say, would give us (on this occasion) the greater pleasure. But is there any conceivable sense in which we could say that the intensity of the pleasure to be got from reading is twice rather than three times or one and a half times, the intensity of the pleasure to be got from walking? Would we not, by trying to make our comparison of intensities mathematically exact, reduce it to meaninglessness?

Similarly, Perry ([19], p. 643) maintains that:

> . . . magnitudes must be of the extensive type . . . they must be divisible into equal units. We can say that the work done in lifting twelve pounds one foot is equal to that done in lifting six pounds two feet, only because twelve pounds are twice six pounds and two feet are twice one foot; because, in other words, a foot is a foot, and a pound a pound, whatever the existing magnitude to which it is added or from which it is subtracted. It is assumed not only that the two magnitudes can be multiplied by one another, but that the reduction of the one can be offset by the increase of the other in some constant ratio throughout the scale . . .
>
> But this condition is not fulfilled in the case of the magnitudes of intensity and preference. We can say that one interest is more intense than another, and that one object is preferred to another; but we cannot say, in either case, that the successive increments are equal, or that one interest is more intense or preferred *by so much*, or that one interest is twice or one-half as intense or preferred as the other.

And Lewis ([13], p. 490) holds that:

> . . . numerical measure cannot be assigned to an intensity of pleasure, or of pain, unless arbitrarily. Intensities have degree, but they are not extensive or measurable magnitudes which can be added or subtracted. That is; we can—presumably—determine a serial order of more and less intense pleasures, more and less intense pains, but we cannot assign a measure to the interval between two such.

The skepticism expressed in these and similar passages may in part spring from a far too restricted notion of strong measurement. Thus, for example, Broad ([6], pp. 246–248), and Leys ([14], p. 18), can be interpreted as saying that measurement of preference is possible only by assigning a unique number to each object, analogously to the way numbers are assigned to classes to measure their cardinality. But even in physics most measurements do not satisfy this rigid requirement of absolute uniqueness. For example the number assigned to a body to measure its mass is uniquely determined only after a unit of mass has been arbitrarily selected (thus it is not a consequence of physical laws, but of convention, whether the mass of a body is measured in pounds or kilograms).

Wheelwright and Perry (in the above quotations), although they apparently do not restrict the notion of strong measurement as severely as Broad and Leys, do assume that measurement is possible only if a meaning can be given to the ratio of the measures assigned different objects, as when we say that the mass of one body is twice that of another. That this ratio requirement is too rigid is attested by the fact that there are significant kinds of physical measurement, such as longitude, which do not satisfy it: the number assigned to measure the longitude of a point on the earth's surface is uniquely determined only after the zero meridian and the unit of longitude have been arbitrarily chosen. If geographers and astronomers were to decide to put the zero meridian through Palomar, rather than Greenwich, the ratio of the longitude of Greenwich to that of Palomar would change from 0 to ∞.

In our opinion the erroneous view of the nature of measurement implicit in these quotations is not peculiar to writings on the theory of value, but infects much of the literature of the philosophy of science. The root of this error lies in the assumption that the only things which are measurable in a strong sense are extensive magnitudes; i.e., magnitudes for which there exists a natural operation corresponding closely to the addition of numbers. Thus masses of bodies are extensive magnitudes, since the mass of the combination of two bodies is equal to the sum of their masses. On the other hand, temperatures of bodies are not extensive magnitudes: for example, the temperature of a mixture of two liquids is very seldom the sum of the temperatures of the original liquids.[8]

From the above discussion it is clear that there are at least *three* significant types of strong measurement in addition to the weak measurement exemplified by an RPR. The four cases, in order of descending strength, are classified in the following table:

Type	Uniqueness Characteristic	Example
1. absolute scale	absolutely unique	cardinality of classes
2. ratio scale	arbitrary unit	mass, length
3. interval scale	arbitrary zero and unit	longitude, time
4. ordinal scale	order preserving	RPR, Beaufort wind scale

In terms of these notions, Wheelwright and Perry can be taken to say that preference (or value) cannot be measured in the sense of a ratio scale. But even if this point be granted, the possibility is not excluded that preference can be measured in the sense of an interval scale. We shall see later that there are substantial arguments to support the view that preference can be measured in this sense.

If preference is measurable in any of the above senses, then it will be possible to summarize the results in numerical statements such as:

1. Event *a* (passage of the full bill) has value 2 for Wright. Many philosophers have objected on one ground or another that it is impossible to give a reasonable direct explication of such numerical assertions (see, e.g., Bohnert in [5]). We agree that statements like (1) cannot be given a sharp meaning independent of a coherent theory of measurement. A coherent theory of measurement is given by specifying axiomatically conditions imposed on a structure of empirically realizable operations and relations. The theory is formally complete if it can be proved that any structure satisfying the axioms is isomorphic to a numerical structure of a given kind. In such a theory it is not expressions like (1) which are given a direct empirical interpretation but assertions about preferences, choices or decisions. The content of (1) is linked with observed phenomena only by the total theory; we know which properties of the numbers referred to in sentences like (1) reflect properties of the empirical structure only when the isomorphism has been exactly characterized.

[8] The most systematic and general way of classifying methods of measurement is according to their uniqueness characteristic, which is determined by the group of transformations under which the measurements are invariant. It seems likely that failure to appreciate this point has led to the erroneous view that no kind of measurement appropriate to physics is applicable to psychological phenomena (Cf. [3], p. 118).

It should of course be noticed that to say the preference of an individual can be measured in the sense of an interval scale does not in any way commit one to the assumption that the preferences of different individuals can be quantitatively compared in a meaningful way, as is supposed by some utilitarian theories. Nor does it commit one to the view, as was suggested for pleasure by Plato ([20], 42E) and Sidgwick ([23], pp. 246–247), that there is a natural zero for preference. An analysis which makes a positive commitment on both these points has recently been made by McNaughton in [16]; however, his achievement of interpersonal comparison and a ratio scale depends upon assumptions which are in our opinion unrealistic.

4. The von Neumann-Morgenstern Axiomatization. We have shown in Section 3 the desirability of measuring preference or value in some stronger sense than that given by an RPR, and that achieving such measure does not have to involve us in all the dubious assumptions which have often been thought necessary. In this section we present two alternative axiomatizations of the notion of an individual preference pattern which, while not assuming that preference can be measured in the sense of a ratio scale, do show what it might mean to say it could be measured in the sense of an interval scale. We do not wish to claim that either of these axiomatizations (or others to be mentioned) constitutes a completely adequate explication of any intuitive notion of rational preference; but by considering several alternatives, and pressing the claims of none, we intend to emphasize the variety of possibilities which open up once the problem is approached in a precise way.

Since at least the time of Pareto and Edgeworth, economists, particularly those interested in welfare economics, have investigated with varying degrees of precision some of these alternatives. Philosophers have unfortunately either ignored the positive and illuminating results obtained along these lines, or else dismissed them as having no philosophical applications. One of the main purposes of the present paper is to demonstrate the relevance of this material to problems in value theory. The first axiomatization we consider is in fact only a slight modification of an axiomatization originally given by von Neumann and Morgenstern ([18], pp. 24–29, 617–632) in an economic context.

We begin our discussion of this axiomatization by introducing the single primitive notion which must be added to the three primitive notions used in defining an RPR. (As before, K is a set of alternatives, P is the binary relation of preference which holds between certain elements of K, and E the binary relation of equivalence in preference which holds between certain elements of K.) The new primitive h is a function of three arguments such that if alternatives x and y are in K, and if α is a probability not equal to 0 or 1 (i.e., if α is a real number such that $0 < \alpha < 1$), then $h(x, y, \alpha)$ is the alternative consisting of x with probability α and y with probability $1 - \alpha$.

To illustrate the meaning of the primitive function h we may refer back to the example at the beginning of the previous section. As before, Action 2 (to support the weakened bill) will obtain alternative b. The alternative which will be obtained by choosing Action 1, and which Wright estimates yields a 1/3 chance of obtaining passage of the full bill (a), and a 2/3 chance of defeat (c), we may now represent as $h(a, c, 1/3)$. If b P $h(a, c, 1/3)$, then Wright apparently considers the difference in value between b and c as greater than half the difference in value between a and b; if $h(a, c, 1/3)$ P b, then Wright considers the difference in value between b and c as less than half the difference in value between a and b.

In terms of these primitives we shall now formulate one possible definition of a rational preference pattern. This definition uses the notion of the *open interval* (0, 1) which is the set of all real numbers α such that $0 < \alpha < 1$.

Definition 2. The ordered quadruple $\langle K, \mathrm{P}, \mathrm{E}, h\rangle$ *is a* RATIONAL PREFERENCE PATTERN IN SENSE ONE *if and only if for every* x, y *and* z *in* K *and every* α *and* β *in* (0, 1):

H1. *The ordered triple* $\langle K, \mathrm{P}, \mathrm{E}\rangle$ *is an* RPR (*in the sense of Definition 1*);
H2. $h(x, y, \alpha)$ *is in* K;
H3. *If* $x \mathrm{~E~} y$, *then* $h(x, z, \alpha) \mathrm{~E~} h(y, z, \alpha)$;
H4. *If* $x \mathrm{~P~} y$, *then* $x \mathrm{~P~} h(x, y, \alpha)$ *and* $h(x, y, \alpha) \mathrm{~P~} y$;
H5. *If* $x \mathrm{~P~} y$ *and* $y \mathrm{~P~} z$, *then there is a number* γ *in* (0, 1) *such that* $y \mathrm{~P~} h(x, z, \gamma)$;
H6. *If* $x \mathrm{~P~} y$ *and* $y \mathrm{~P~} z$, *then there is a number* γ *in* (0, 1) *such that* $h(x, z, \gamma) \mathrm{~P~} y$;
H7. $h(x, y, \alpha) = h(y, x, 1 - \alpha)$;
H8. $h(h(x, y, \alpha), y, \beta) \mathrm{~E~} h(x, y, \alpha\beta)$.

The intuitive interpretation of the first axiom has been given in Section 2. The second axiom states that the set K must be taken to contain not only the given alternatives, but also all (finite) probability combinations of them: thus if the given alternatives are ale, gin, and tea, then K must contain, for example, the alternative consisting of ale with probability 1/4, gin with probability 1/4, and tea with probability 1/2—since this can be expressed as $h(h(\text{ale}, \text{gin}, 1/2), \text{tea}, 1/2)$. The third axiom means that if x is equivalent (in preference) to y then the combination of x and any alternative z with probability α is equivalent to the combination of y and z with probability α. The fourth axiom means that if alternative x is preferred to alternative y, then x is preferred to any probability combination of x and y, and any probability combination of x and y is preferred to y. The fifth and sixth axioms mean that if y is between x and z in preference, then there is a probability combination of x and z which is preferred to y, and one to which y is preferred. The meaning of the identity asserted in the seventh axiom is obvious from the intended interpretation of the function h. The last axiom states a rule for combining probabilities.

Now we may consider the plausibility of Definition 2 as an explication of a rational pattern of preferences. The axioms as they stand may strike the reader as rather arbitrary. The *ad hoc* air of this particular list of axioms is partly removed, at least, by observing that a rational preference pattern as here defined can be *proved* measurable in the sense of an interval scale. So far as we know none of the axioms could be omitted and the desired formal properties still be preserved. The formal adequacy of the axiomatization (for interval scale measurement) is summarized by the following theorem:[9]

Theorem 1. If $\langle K, \mathrm{P}, \mathrm{E}, h\rangle$ is a rational preference pattern (in the sense of Definition 2), then: (A) *there exists a function ϕ, which is defined over K and whose values are real numbers, such that for every x and y in K and α in* $(0, 1)$

$$\text{(i)} \qquad x \mathrm{\ P\ } y \text{ if and only if } \phi(x) > \phi(y),$$

$$\text{(ii)} \qquad x \mathrm{\ E\ } y \text{ if and only if } \phi(x) = \phi(y),$$

$$\text{(iii)} \qquad \phi(h(x, y, \alpha)) = \alpha\phi(x) + (1 - \alpha)\phi(y);$$

(B) *if ϕ_1 and ϕ_2 are any two functions satisfying* (A), *then there exist real numbers a and b with $a > 0$ such that for every x in K*

$$\phi_1(x) = a\phi_2(x) + b.$$

The intuitive content of (A) is that it is always possible to assign numbers to alternatives in such a way as to preserve the structure of a rational preference pattern. The content of (B) is that this assignment of numbers has the uniqueness property characteristic of an interval scale, that is, once a zero point and unit have been chosen, the assignment is unique.

We have shown that Definition 2 has certain desirable formal properties; can any argument be offered to indicate that these properties constitute intuitively acceptable conditions of rationality? It would be tedious and perhaps pointless to attempt to defend the axioms one by one. What we shall do instead is to show that a pattern of preferences among alternatives involving uncertainty which did not constitute a rational preference pattern in the sense of Definition 2 would be irrational according to intuitively plausible ideas. It will be seen that the argument is an extension to a more general case of the argument used in Section 2 to prove intransitive preference rankings irrational.

We shall assume that a pattern of preferences among a set K of alternatives is irrational if a rational choice is impossible among any subset L of alternatives in K. A rational choice (relative to L) is, as before, a choice which selects any alternative to which none is preferred (within L). Consider the following example. Let the set K consist of alternatives a, b, c, d, and all finite probability combinations of these four alternatives. Assume that

1. $a \mathrm{\ P\ } b, \qquad b \mathrm{\ P\ } c, \qquad c \mathrm{\ P\ } d$
2. $h(a, c, 1/2) \mathrm{\ E\ } b$
3. $h(b, d, 1/2) \mathrm{\ E\ } c$
4. $h(a, d, 3/5) \mathrm{\ E\ } b$

Clearly 2, 3 and 4 are each consistent with 1, and any three of the assumptions are consistent with each other. Taken together, however, the four assumptions

[9] The proof of this theorem may be obtained by a trivial modification of a proof given in [18], pp. 613–628.

are not consistent in the sense of Definition 2. But we also can show that the pattern of preferences given by assumptions 1–4 is irrational in the sense defined above, for we can find a subset L of alternatives in K among which a rational choice is impossible. Such a subset consists of the two alternatives b and $h(a, d, 5/8)$. First, we see that $h(a, d, 5/8)$ P b. For since a is preferred to d (by assumption 1 and the transitivity of preference), obviously any alternative which gives a better than 3/5 chance of getting a and therefore a less than 2/5 chance of getting d will be preferable to $h(a, d, 3/5)$; and such an alternative is $h(a, d, 5/8)$. Assumption 4 tells us $h(a, d, 3/5)$ is equivalent in value to b; since $h(a, d, 5/8)$ P $h(a, d, 3/5)$, we therefore conclude that $h(a, d, 5/8)$ P b. On the other hand, assumptions 2 and 3 tell us that b is halfway between a and c in value, and c is halfway between b and d; therefore the interval between a and b is one third of the interval between a and d, that is, b E $h(a, d, 2/3)$. Clearly an alternative which gives a less than 2/3 chance of getting a and more than a 1/3 chance of getting d is not as valuable as $h(a, d, 2/3)$; and such an alternative is $h(a, d, 5/8)$. But since b E $h(a, d, 2/3)$ and $h(a, d, 2/3)$ P $h(a, d, 5/8)$, we conclude that b P $h(a, d, 5/8)$. Thus no rational choice is possible between b and $h(a, d, 5/8)$, since each is preferred to the other on the basis of the assumed preferences. This seems an acceptable reason for saying the assumed preferences, taken together, constitute an irrational pattern of preferences.

It is fairly obvious that examples of this kind can be constructed for any pattern of preferences among alternatives involving uncertainty which does not have the properties ascribed to a rational preference pattern (in the sense of Definition 2) by Theorem 1. What we have given is therefore a quite general argument to the effect that Definition 2 does provide important necessary conditions for rationality of preferences among alternatives involving uncertainty.

On the other hand there are certain consequences of Definition 2 which lead to questions regarding its explicative validity. Some of these questions have been widely discussed in the economic literature.[10]

First, Axioms H5 and H6 together make the system what mathematicians call "Archimedean." This means that the definition rules out the possibility that K contains elements (such as, perhaps, receiving a large sum of money, and being eaten by a tiger) one of which is infinitely preferred to the other. Those who believe in a *summum bonum* (or a *summum malum*) in the sense of a good or evil incomparably better or worse than any other alternative might feel that the Archimedean restriction is a serious one.[11]

Another objection can be raised, particularly in connection with H3, H4 and H8; these axioms seem to imply that the element of risk does not in itself alter the value of an alternative. The nature of this assumption can be illustrated in a simple way in the case of H4. Mr. A. dines out with a friend. He would prefer to pick up the check (alternative a) rather than let his friend pay (alternative b); but he would much rather decide the matter by tossing a coin (alternative

[10] See particularly [1], [2] and [15]. These articles contain extensive bibliographical references.

[11] This point was called to our attention by Dr. Leo Simons. See also [2], p. 425.

$h(a, b, 1/2)$). It is clear that this violates H4 which would require that a P $h(a, b, 1/2)$ when a P b. The difficulties involved here are very complex. For further discussion and a partial solution the reader is referred to [7].

A third objection to Definition 2 is that the set K must contain not only the initial alternatives but also all finite probability combinations of them, and therefore an infinite number of elements (this follows from the so-called "closure" axiom, H2). There is a reasonable doubt, at least, what empirical interpretation can be given to the assumption of an infinite number of alternatives.

It would seem desirable therefore to separate assumptions about the number of alternatives from assumptions about the number of probability distributions over them. This separation is accomplished by the following axiomatization, which in addition to the set K of alternatives, the relation P of preference, the relation of E of equivalence of preference, introduces a new primitive T. T is a quaternary relation whose intended interpretation is that if x, y and z are in K and α is a probability in the closed interval $[0, 1]$ (i.e., $0 \leqslant \alpha \leqslant 1$), then $T(x, y, z, \alpha)$ if and only if y is not preferred to x, z is not preferred to y and the alternative consisting of x with probability α and z with probability $1 - \alpha$ is equivalent in preference to alternative y. Informally we may think of $T(x, y, z, \alpha)$ as holding when $h(x, z, \alpha)$ E y.

Definition 3. $\langle K, \mathrm{P}, \mathrm{E}, T \rangle$ *is a* RATIONAL PREFERENCE PATTERN IN SENSE TWO *if and only if for every* x, y, z *and* w *in* K *and every* α *and* β *in* $[0, 1]$:

T1. $\langle K, \mathrm{P}, \mathrm{E} \rangle$ *is an* RPR *(in the sense of Definition 1)*;
T2. *If* x E y *and* y E z, *then* $T(x, y, z, \alpha)$;
T3. *If* x P y *or* z P x, *then not* $T(y, x, z, \alpha)$;
T4. *If* $T(x, y, z, \alpha)$ *and* x E w, *then* $T(w, y, z, \alpha)$;
T5. *If* $T(x, y, z, \alpha)$ *and* y E w, *then* $T(x, w, z, \alpha)$;
T6. *If* $T(x, y, z, \alpha)$ *and* z E w, *then* $T(x, y, w, \alpha)$;
T7. *If* x P z, *then* y E z *if and only if* $T(x, y, z, 0)$;
T8. *If* x P z, *then* x E y *if and only if* $T(x, y, z, 1)$;
T9. *If* x P z, *not* y P x *and not* z P y, *then there is a unique* γ *in* $[0, 1]$ *such that* $T(x, y, z, \gamma)$;
T10. *If* x P y, y P z *and* z P w, *and any two of the following, then the other two*:

$$T(x, y, w, \alpha)$$
$$T(x, z, w, \beta)$$
$$T\left(y, z, w, \frac{\beta}{\alpha}\right)$$
$$T\left(x, y, z, \frac{\alpha - \beta}{1 - \beta}\right).$$

An adequacy theorem analogous to that given for Definition 2 has been proven for Definition 3, thus demonstrating that a rational preference pattern (in the sense of Definition 3) is measurable in the sense of an interval scale. It is easy to show that the statement that K is a finite set (as well as the statement that K is an infinite set) is consistent with the above axioms. In this regard Definition 3 is an improvement on Definition 2.

The adequacy theorem for Definition 3 brings out an interesting formal point. Readers familiar with the economic and statistical literature on utility theory (see for example [4], [10] and [22]) may have wondered why, in place of Definition 2 or 3, we did not use an axiomatization which depends on considering the set of all probability distributions over the set K of alternatives. The proof of the adequacy of Definition 3 shows however that it is necessary to consider only the set of all two-element probability distributions over K. It seemed advisable not to eliminate the relation T in favor of a primitive binary relation defined over the set of all such two-element probability distributions because this would tend to mask the nature of the assumptions made about preference.

5. The Problem of Probability. The two definitions of a rational preference pattern given in the last section have this in common: both tie the concept of value or preference in a fundamental way to the concept of probability. This is because both rely upon alternatives involving probability distributions in order to achieve a stronger measure of value or preference than a simple ordering. The essential interlacing of values and probabilities seems to have both its advantages and drawbacks. Among the advantages, we may list two:

First, axiomatizations such as those given in Definitions 2 and 3 appear to have an extremely direct application to practical situations which are very common, that is, situations in which alternatives must be weighed which involve uncertainties. Policy decisions in business and politics are obvious (but by no means the only) examples.

Second, the axiomatizations which have been given in the previous section suggest relatively simple behavioristic procedures for empirically testing degrees of preference. It is not at all obvious how to determine the relative values of alternatives to a person in any direct way without making unwarranted assumptions (for example, amounts of money or amounts of work can be used to measure degrees of preference only if we assume that degree of preference is a simple mathematical function of hours of work or amounts of money). But by using the function h or the relation T, a simple choice between alternatives in which the probabilities are controlled may be interpreted as evidence for as subtle degrees of preference as we please. To illustrate, Mosteller and Nogee ([17]), in a series of experiments performed at Harvard, offered subjects a choice of betting or not betting 5 cents against a certain amount of money (say 25 cents) at various odds. By adjusting the odds, the experimenters discovered the offer which the subject would accept 50% of the time. This they interpreted as meaning that the subject found the alternative of keeping 5 cents (alternative a) equal in preference to the alternative consisting of a certain chance (α) of losing 5 cents (alternative b) and a certain chance ($1 - \alpha$) of winning 25 cents (alternative c). In terms of Definition 2, we may write this $a \mathrm{E} h(b, c, \alpha)$. By finding the α (i.e., the odds) which results in equality of preference, the relative values of the alternatives to the subject can thus be inferred.

To the advantages just listed there correspond certain limitations on the scope of application or explicative value of definitions which, like Definitions 2 and 3, link value with probability.

First, the characteristics which make these definitions apply directly to situations involving uncertainty make it unclear what, if any, application the definitions have in situations which do not involve uncertainty. As there is no obvious reason why rationality with respect to relative degrees of preference should be restricted to situations involving uncertainty, the definitions of rational preference patterns so far given seem to be limited to a special case. In fact instead of dealing with probabilities and relative degrees of value simultaneously it would seem far more natural to determine relative degrees of value first, and *then* modify these by the probabilities to yield decisions in uncertain situations.

Second, the very factors which make the axiomatizations of Definitions 2 and 3 so amenable to behavioristic interpretations bring in their train attendant difficulties. One of these difficulties is to find a way of separating the value accorded an alternative solely on account of the risks or uncertainties it entails from the values due to the primary alternatives alone. This point has already been discussed in Section 4.

Another, and related, difficulty concerns the interpretation of the notion of probability as it is used in Definitions 2 and 3. So long as a purely normative use of the definitions is intended, there may be no special problem. Suppose we agree that a rational man who holds that $h(a, c, 1/2)$ E b and $h(b, d, 1/2)$ E c must also hold that $h(a, d, 1/3)$ E c. Here the probabilities in the first two equivalences may be taken as what the man believes the probabilities to be; and the probability in the last equivalence what he will believe the probability to be when he judges the alternatives to be equal in preference, provided he is rational. No question about *actual* probabilities needs to be asked. It will be noticed that the natural way of explaining such a normative use of Definition 2 assumes a separation of the judgment of the values of the alternatives and the judgment of the probabilities involved.

If we ask however whether or in what degree someone's actual preferences are rational (in the sense of Definition 2), the following problem arises. Suppose we try (resuming the example from Mosteller and Nogee) to determine empirically a subject's relative preferences for various alternatives by offering him specific choices. Imagine that we repeatedly offer a particular subject, Mr. C., a choice between betting (offer 1) or not betting (offer 2) against these odds: if he draws a spade from a well-shuffled normal pack of cards, he wins 25 cents; if he fails to draw a spade, he loses 5 cents. (As above, neither gaining nor losing money is alternative a, losing 5 cents alternative b, and gaining 25 cents alternative c.) Assume now that Mr. C. accepts the bet as often as he rejects it, and we agree that this means that for Mr. C. alternative a is equal in preference to the proffered odds on b and c. When we come to interpret this result in terms of relative degrees of preference for alternatives a, b, and c, it is tempting to reason that since the actual chances of drawing a spade from a normal, well-shuffled deck are one in four, the experimental findings imply that, for Mr. C., $a \mathrm{E} h(b, c, 3/4)$. But this result, which may be read as saying that the difference in preference between b and a is a third of the difference in preference between a and c, depends entirely upon the unverified assumption that Mr. C. believes, or acts as if he believes, that the chances of drawing a spade out of a normal pack are one in four. We have assumed, in other words, that Mr. C. proportions his *expectation* that he will draw a space to his actual, or mathematical, chances of drawing a spade.

It seems clear that this assumption is not generally justified; it is unreasonable to take for granted, when testing the rationality of a man's preference pattern, that his judgment of probabilities is perfectly rational. Mr. C., whether through ignorance, stupidity, or a conviction that spades are lucky or unlucky for him, may very easily believe or act as if he believed that he has a better (or worse) than one in four chance of drawing a spade; and if so, then this must be considered before we can judge his relative preferences for the alternatives. Mr. C's equal acceptance of offers 1 and 2 *may* mean that he judges the differenc in value between receiving 25 cents and neither gaining nor losing money to be only three times the difference between neither gaining nor losing money and losing 5 cents—provided the psychological probabilities are identical with the mathematical (that is, objective) probabilities; it may equally well mean that for him the difference in preference between c and a is five times the difference between a and b, and he believes he has only one chance in six of drawing a spade. Indeed if we could assume that preference were linear in money, offers of the sort made to Mr. C. would measure, not preference, but psychological probabilities.

Since we do not want merely to assume a measure of preference, and we cannot reasonably assume that psychological probabilities are equal to objective probabilities, results obtained in experiments of the kind made by Mosteller and Nogee are difficult to interpret. Nor is it obvious what other sorts of behavioristic experiments could be performed to measure preferences on the basis of the primitives of Definition 1 and 2 which would be free from this difficulty. However, following a suggestion which F. P. Ramsey made for a somewhat different purpose ([21]), it is possible to design experiments which lead to a separate measurement of psychological probability and preference. The theoretical foundations of this approach are to be found in [7], and the experimental results in [8].[12]

Stanford University

REFERENCES

1. M. ALLAIS, "Le comportment de l'homme rationnel devant le risque: critique des postulats et axiomes de l'école Americaine," *Econometrica*, *21* (1953), 503–546.
2. K. J. ARROW, "Alternative approaches to the theory of choice in risk-taking situations," *Econometrica*, *19* (1951), 404–437.

[12] An outline of the relevant ideas was contained in the version of this paper issued as a technical report of the Stanford Value Theory Project. A continuing program of theoretical and empirical investigations is under way.

3. G. BERGMANN AND K. W. SPENCE, "The logic of psychophysical measurement," reprinted in *Readings in the Philosophy of Science*, edited by H. Feigl and M. Brodbeck, New York, 1953, from *Psychological Review*, *51* (1944).
4. D. BLACKWELL AND M. A. GIRSHICK, *Theory of Games and Statistical Decisions*, New York, 1954.
5. H. G. BOHNERT, "The logical structure of the utility concept," in *Decision Processes*, edited by R. M. Thrall, C. H. Coombs and R. L. Davis, New York, 1954.
6. C. D. BROAD, *Five Types of Ethical Theory*, London, 1944.
7. D. DAVIDSON AND P. SUPPES, "Finitistic rational choice structures," Report No. 3, Stanford Value Theory Project, March, 1955.
8. D. DAVIDSON, S. SIEGEL AND P. SUPPES, "Some experiments and related theory on the measurement of utility and subjective probability," Report No. 4, Stanford Value Theory Project, May, 1955.
9. JOHN DEWEY, *Human Nature and Conduct*, Modern Library edition.
10. M. FRIEDMAN AND L. J. SAVAGE, "The expected-utility hypothesis and the measurement of utility," *Journal of Political Economy*, 60 (1952), 463–474.
11. C. G. HEMPEL, *Fundamentals of Concept Formation in Empirical Science*, vol. II, no. 7, *International Encyclopedia of Unified Science*, Chicago, 1952.
12. O. LANGE, "The determinateness of the utility function," *Review of Economic Studies*, 1 (1934), 218–225.
13. C. I. LEWIS, *An Analysis of Knowledge and Valuation*, LaSalle, Illinois, 1946.
14. A. R. LEYS, *Ethics for Policy Decisions*, New York, 1952.
15. J. MARSCHAK, "Rational behavior, uncertain prospects, and measurable utility," *Econometrica*, *18* (1950), 111–141.
16. R. McNAUGHTON, "A metrical concept of happiness," *Philosophy and Phenomenological Research*, *14* (1953), 172–183.
17. F. MOSTELLER AND P. NOGEE, "An experimental measurement of utility," *Journal of Political Economy*, *59* (1951), 371–404.
18. J. VON NEUMANN AND O. MORGENSTERN, *Theory of Games and Economic Behavior*, 2nd ed., Princeton, 1947.
19. R. B. PERRY, *General Theory of Value*, Cambridge, Mass., 1950.
20. PLATO, *Philebus*.
21. F. P. RAMSEY, *The Foundations of Mathematics and other Logical Essays*, London, 1931.
22. L. J. SAVAGE, *The Foundations of Statistics*, New York, 1954.
23. H. SIDGWICK, *The Methods of Ethics*, 5th ed., London, 1893.
24. P. WHEELWRIGHT, *A Critical Introduction to Ethics*, revised ed., New York, 1949.
25. J. WISDOM, "Review of Waddington's *Science and Ethics*," reprinted in *Philosophy and Psycho-Analysis*, Oxford, 1953.