© Journal of the American Statistical Association
December 1971, Volume 66, Number 336
Theory and Methods Section

Elicitation of Personal Probabilities and Expectations

LEONARD J. SAVAGE*

Proper scoring rules, i.e., devices of a certain class for eliciting a person's probabilities and other expectations, are studied, mainly theoretically but with some speculations about application. The relation of proper scoring rules to other economic devices and to the foundations of the personalistic theory of probability is brought out. The implications of various restrictions, especially symmetry restrictions, on scoring rules is explored, usually with a minimum of regularity hypothesis.

1. INTRODUCTION

1.1 Preface

This article is about a class of devices by means of which an idealized *homo economicus*—and therefore, with some approximation, a real person—can be induced to reveal his opinions as expressed by the probabilities that he associates with events or, more generally, his personal expectations of random quantities. My emphasis here is theoretical, though some experimental considerations will be mentioned. The empirical importance of such studies in many areas is now recognized. It was emphasized for the area of economics in an address by Trygve Haavelmo [28, p. 357]:

> I think most of us feel that if we could use *explicitly* such variables as, e.g., what people *think* prices or incomes are going to be, or variables expressing what people *think* the effects of their actions are going to be, we would be able to establish relations that could be more accurate and have more explanatory value. But because the statistics on such variables are not very far developed, we do not take the formulation of theories in terms of these variables seriously enough. It is my belief that if we can develop more explicit and a priori convincing economic models in terms of these variables, which are realities in the minds of people even if they are not in the current statistical yearbooks, then ways and means can and will eventually be found to obtain actual measurements of such data.

A special instance of the central general principle of this article was recognized long ago by Brier [5], the general principle itself was briefly but colorfully announced by McCarthy [37], and a considerable literature pertaining to it has grown up, some of which will be cited in context and most of which can be found through the references cited, especially the recent and extensive [52] and others that I call "key references."

Bruno de Finetti and I began to write the present article in the spring of 1960, not yet aware of our predecessors and contemporaries. The impetus was de Finetti's, for he had brought us to rediscover McCarthy's [37] insight about convex functions. We expected to make short work of our "little note," but it grew rapidly in many directions and became inordinately delayed. Now we find that the material in the present article is largely mine and that de Finetti has published on diverse aspects of the same subject elsewhere [12, 13, 14, 17]. De Finetti has therefore withdrawn himself from our joint authorship and encouraged me to publish this article alone, though it owes so much to him at every stage, including the final draft.

The article is written for a diverse audience. Consequently, some will find parts of it mathematically too technical, and others will find parts too elementary. If each skips what puzzles or bores him he will, I hope, find the rest reasonably complete for him.

1.2 Summary

The bare essentials of the economic theory of personal probability and expectation are introduced (Section 2). Various difficulties in principle that beset the evaluation of preferences such as those that determine the price at which a man is just willing to sell his car or the probability for him that the car will shortly need a new muffler are discussed (Section 3).

A probability is a price, in a manner of speaking. More accurately, it is a marginal rate of substitution. Such rates can be elicited by a general mode of behavioral interrogation rather like one that has been proposed for ordinary prices (Section 4). These methods admit mathematically special cases that seem to be of particular interest (Section 5).

The rate-eliciting methods, both general and special, are examined with reference to probabilities and expectations (Section 6). The methods of eliciting personal probabilities thus arrived at can also be approached through the ideas of statistical decision theory (Section 7), and this decision-theoretic method provides a relatively

* Leonard J. Savage is Eugene Higgins Professor of Statistics and chairman, Department of Statistics, Yale University, New Haven, Conn. 06520. This article was prepared in connection with research supported by the Army, Navy, Air Force and NASA under contracts administered by the Office of Naval Research. Reproduction in whole or in part is permitted for any purpose of the United States Government. Part of the work was supported by the Michigan Institute of Science and Technology. The author wishes to thank Ward Edwards, Allan Murphy, Richard Savage, Carl-Axel Stael von Holstein, Robert Winkler and an excellent referee for their helping improving the article. He is especially grateful to Bruno de Finetti who provided the impetus for the article.
EDITOR'S NOTE: Professor Savage died suddenly on November 1, 1971, while this article was in press. A memorial article on his work will appear in a forthcoming issue of one of the ASA publications.

new basis due to de Finetti for defining personal probability (Section 8).

The methods are generalized to simultaneous elicitation of several rates, probabilities, or expectations (Section 9). This vectorial discussion concludes with a prominent special case in which the subject's income depends only on that event among disjoint and exhaustive possibilities that actually occurs.

Aspects of possible applications are discussed (Section 10). These include domains of possible application, impediments to application, and criteria for choice among the methods.

2. RATES OF SUBSTITUTION AND PROBABILITY AND EXPECTATION

A man who owns some of a commodity, say wheat, will, under suitable circumstances, be almost indifferent both to buying and selling modest quantities of it at a certain price per bushel, his (marginal) rate of substitution of cash for wheat.

Money payable subject to a contingency, such as the accidential burning of a house or the outcome of a race, can be regarded as a commodity [4]. Such commodities are explicitly dealt in by insurance companies and bookmakers, and we encounter them implicitly wherever we make decisions in the face of uncertainty. There is reason to postulate that an ideally coherent person has a rate of substitution $P(A)$ for money contingent on the event A. When q is not too large, he is indifferent to buying or selling q dollars contingent on A for $qP(A)$ dollars outright, and $P(A)$ is defined as the probability of A for the person [9]. Though the relationship can, because of the nonlinearity of the utility of money, be expected to hold only in the limit for infinitesimal values of q, I shall usually write as though it were exact. This limitation is serious, though a technique for creating utility-free currency and other ways to avoid the effects of nonlinear utility will be mentioned in the next section.

Let U denote the logically certain event and A and B any two logically incompatible events. For an economically coherent person,

$$P(U) = 1,\ P(A) \geq 0,\ P(A \text{ or } B) = P(A) + P(B), \quad (2.1)$$

as shown in [9]. Thus, such a person's P is a (finitely additive) probability measure.

If a finite sequence $\{A_s\}$ of events is a partition (that is, if every pair of them is incompatible but one or another of them obtains), then

$$\sum P(A_s) = 1, \quad (2.2)$$

as follows from (2.1).

To orient the reader in the critical literature on personal probability, the commentary and bibliography of [47] might be useful. Excellent early papers are republished, in English, in [31]. Some more recent ones are [51, 18, 16, 17]. An extensive bibliography with special reference to experimental aspects is [19].

More general than the notion of the probability of an event is that of the expectation of a random quantity. Let $\{A_s\}$ be a partition and $\{v_s\}$ a corresponding sequence of numbers; the two together can be regarded as the random quantity V that takes the value v_s if A_s obtains. If a person's probability of A_s is p_s, what ought he pay for the simultaneous offer of qV dollars, that is, for qv_1 dollars in case A_1 obtains, qv_2 dollars in case A_2 obtains, etc.? Viewing qV as a commodity bundle consisting of qv_1 units of one commodity, qv_2 units of another, etc., the answer (always for moderate q) is clearly

$$qE(V) = q \sum p_s v_s, \quad (2.3)$$

or q times the expectation of V.

A probability $P(A)$ is itself plainly an expectation, namely $E(I_A)$, where I_A is the random quantity that is 1 if A obtains and 0 if A does not obtain. The event A and its associated random quantity I_A can, to considerable advantage, be rigorously regarded as two aspects of one object, simply denoted by A [15]. Thus it is meaningful and convenient to write $P(A) = E(A)$.

A random quantity V need not be defined in terms of a partition nor need it have only a finite number of possible values; it may be simply any (ordinarily) unknown, empirically determinable number. For the unknown payment qV, the person is presumably willing to exchange some definite payment $qE(V)$. The rate of substitution $E(V)$ is the person's expectation of the unknown, or random, quantity V. (In practice, V would be bounded; mathematically, it can be useful to consider some unbounded V and to exclude others.)

3. SOME DIFFICULTIES IN THE EXPERIMENTAL ELICITATION OF PREFERENCES

The difficulties mentioned in this section are mainly tangential to the present article. Some of them have little or no bearing on the particular situation to be studied; others are here treated as secondary for the time being. The first type are mentioned only to emphasize the advantage of a certain method of eliciting prices and rates of substitution, including probabilities and expectations; and the others are mentioned to warn of their existence.

Many conceptual experiments on preference will presumably remain conceptual only, some because their financial cost is prohibitive and others because they imply immoral or impractical interference with people's lives. For example, experiments to determine directly what risk of pauperhood a person will take to avoid spending a hundred dollars or to gain a million dollars are literally fantastic. This does not of course preclude learning something indirectly: from the behavior of buyers of insurance and of gamblers and speculators [23, 24]; or by asking subjects to introspect about hypothetical choice; or by observing changes of economic behavior that follow changes in the policies of governments, firms, and other institutions; or in other ways.

Statistical problems and difficulties arise in empirical studies of preference, as they do in all empirical studies. A more special, but perhaps related difficulty, is that all subjects report, or otherwise reveal, that they do not know their own preferences; they experience wavering and indecision that cannot be identified with mere indifference. See, for example, [47, pp. 21, 59, 168–9; 18, Section 26; 45; and 22 under "Indifference, Intransitive" in the Index].

Another difficulty peculiar to experiments on preference is that once an experimenter satisfies one preference of a subject, he may quite drastically change the subject's pattern of preferences. The thirsty man who now prefers water to wine and wine to whiskey might fail to reveal this in an experiment in which he is first offered his choice between wine and water and then, his thirst quenched with water, is offered his choice between wine and whiskey. A related phenomenon is illustrated by the subject who accepts whiskey (perhaps for later consumption) in preference to wine because he suspects that the experimenter will shortly be offering him an opportunity to obtain water. Such difficulties arise in any attempt to study the gambling (and insuring) preferences of a subject. They are particularly evident and important in experiments to determine the demand price or the offer price of a subject for a specific object or service.

One interesting device for coping with such difficulties was pointed out to me by W. Allen Wallis in 1949 or 1950 and was independently exploited by Allais [2] in experiments conducted in 1952. Let the experimenter put the subject successively in several hypothetical choice situations always with the understanding that, when all the subject's choices have been expressed, one of them will actually be implemented by a chance device that, plain to the subject, has no direct connection with the situations of interest. For instance, a subject asked to rank half a dozen tickets to plays, concerts, and athletic events in the order of his preference will, insofar as he is rational, do so sincerely if he understands that the tickets are to be thoroughly shuffled and that he will be given that one of the top two tickets for which he has indicated preference.

The following application of the same general principle was introduced by Marschak [33, 34]. The experimenter makes a sealed bid for an object in the subject's possession and the subject, before seeing this bid, puts an "asking price" on the object. The sale takes place at the bid price, if and only if the bid is at least as high as the asking price. It is clearly to the advantage of the subject to name his actual offer price, at least if he is not in a position to exclude the possibility of bids in some neighborhood of this price; and even in that exceptional case, there is no advantage in naming any other price. (The procedure can evidently be so modified that the subject is the buyer rather than the seller.)

Actual subjects are of course sometimes blind to their own clear advantage [34; 35, p. 47], failing, for example, to understand that they can only deprive themselves by asking too much. It is no final criticism of such a method to say that subjects do not automatically and instinctively understand it or that, understanding, they have psychological difficulty in doing the rational thing. Such facts do underline the need for education and training prior to, and even during, the application of elicitation devices. Incidentally, such education promises to be of great general benefit to the subject and deserves wide promulgation on its own account.

Marschak's method does not depend on the hypothesis that the utility of money is linear, but throughout most of this article, this hypothesis is relied upon. The hypothesis is presumably a valid approximation if only small transactions are involved. Still better, there would be no approximation at all if payments were made in utiles rather than in money, which may sometimes be roughly feasible.

A certain scheme for effecting payments in utiles is vividly suggested by Smith [50, Section 13]. If, to paraphrase Smith, the currency used consists of tickets in a one-prize lottery, which the subject is known to regard as fair and independent of all the uncertainties that are of direct interest for the experiment, then the subject's utility for this currency is linear, though the transactions may be very substantial. Of course for this scheme to be valid, the utility of the prize to the subject must not depend on the outcome of events that are of interest. For example, a lottery in which the prize is a diamond would obviously not provide a valid utility-free currency for exploring a subject's opinions about the future of the diamond market.

In principle, the utility function of a subject for ordinary money can be determined by certain elicitation experiments that give operational meaning to utility. With this function known, the experimenter would know the utility worth to the subject of any proposed payment (positive or negative) and could arrange to make payments in utiles. Practical limitations on this scheme and on the one in the preceding paragraph are severe, but the ideas are at least sufficiently suggestive to be worth bringing out [57, 60].

The linear approximation seems adequate for many applications. Where it is not quite adequate, much can be said for taking as the next approximation a function for the utility of money of the form

$$(1 - e^{-\lambda x})/\lambda. \tag{3.1}$$

(Any function of the form $b - ae^{-\lambda x}$, with a and λ of like sign, would of course amount to the same thing; the choice of the constants a and b made in (3.1) emphasize that the proper interpretation of $\lambda = 0$ is linear utility, because, for fixed x and small λ, (3.1) is approximately $[1-(1-\lambda x)]/\lambda = x$.) For almost all applications λ would be positive, which is why a minus sign was introduced in the exponent of (3.1). A person who is just willing to toss a fair coin for \$10.00 if he receives a side payment of \$0.10 exhibits a λ of about .0020 per dollar. To show that a person's utility for money is adequately described by (3.1) in some range of practical interest and to determine the value applicable to him may in some contexts provide a practical basis for paying him in utiles.

Taken literally, the exponential utilities described what I would call a perfect miser, a person who does not necessarily rank lotteries, and the like, simply by their expected cash value but whose rankings do not fluctuate with his own wealth at the moment. Within sufficiently narrow limits, any person's utilities can be expected to be practically linear. When these limits are somewhat exceeded so that nonlinearity must be taken into account, there should be new limits within which the great experimental simplification of miserly behavior can still be safely assumed as an approximation. Miserly utility, or utility with constant local risk aversion, seems to have been introduced by Pratt [43].

4. THE ELICITATION OF A RATE OF SUBSTITUTION

Some of the difficulties mentioned in the previous section tend to disappear when the investigation is confined to rates of substitution, of which probabilities are for this paper the prime example. First, the transactions envisaged are then moderate, almost by definition, so that typically the experiments that suggest themselves are not grandiose. Second, the satisfaction of one preference here has ideally no effect on the subject's other preferences, though in practice there can well be a conflict between keeping the transactions small enough to avoid important manifestations of saturation (or of nonlinear utility, as we say in connection with probability or expectation) and yet large enough to justify the subject's close attention. Possible remedies for this conflict, at least in probabilistic cases, were mentioned in Section 3. Finally, by means of devices related to Marschak's method mentioned there, it is possible to present a subject with a single, relatively simple, economic choice in which it is to his interest to reveal any reasonable number of his rates of substitution, as will now be explained for the case of a single rate.

Suppose the experimenter offers, once and for all, to buy some of a commodity at each possible price—more accurately price rate—so much at each rate. The subject will then have an incentive to satisfy the expressed demands of the experimenter at all rates higher than the subject's rate r but not those at lower rates, thereby revealing r.

In mathematical terms, the experimenter offers, for a certain non-negative "schedule of demands" f, to buy

$$\int_{x\leq\rho} f(\rho)d\rho \tag{4.1}$$

units of the commodity at a total cost of

$$\int_{x\leq\rho} \rho f(\rho)d\rho \tag{4.2}$$

for any number x named by the subject. From the subject's own viewpoint, his income from such a transaction, as a function of x, is

$$\begin{aligned} I(x;r) &= -\mathrm{r}\int_{x\leq\rho} f(\rho)d\rho + \int_{x\leq\rho} \rho f(\rho)d\rho \\ &= \int_{x\leq\rho} (\rho - r)f(\rho)d\rho. \end{aligned} \tag{4.3}$$

Plainly, and in correspondence with the verbal argument that preceded this paragraph, the income $I(x;r)$ will be maximized when $x=r$, and for that x alone if $f(\rho)$ is positive near r on both sides. If, for example, $f(\rho)=\rho^{-3}$ for positive ρ, then $I(x;r)=(2x-r)/(2x^2)$, $I(x;r)-I(r;r) = -(x-r)^2/(2rx^2)\leq 0$ for all positive r.

The mathematical analogy between this device for eliciting a rate and Marschak's device (discussed in paragraph 6 of Section 3) for eliciting a price is notable. The possibility that r is negative, in which case the subject regards the commodity as a nuisance, is not necessarily excluded but can often be eliminated from practical consideration. The experimenter might then be content, as in the example above, to define f only for positive prices. Similarly, the experimenter could with impunity attach 0 density to preposterously high prices.

With abbreviations, (4.3) reads

$$I(x;r) = b(x)r + c(x). \tag{4.4}$$

The experimenter is therefore in effect offering to give the subject $c(x)$ units of cash and $b(x)$ units of the commodity for any number x chosen by the subject. The vital feature of the functions b and c, which is automatically assured if b and c are of the form implied by (4.3) with positive f, is that

$$I(x;r) = b(x)r + c(x) \leq b(r)r + c(r) = J(r), \tag{4.5}$$

with equality if and only if $x=r$, where $J(r)$ has been introduced as an abbreviation for $I(r, r)$.

The sort of b and c induced by (4.4) with a positive f satisfies (4.5), but what can be said of the most general solution of (4.5)? In this problem, it is to be understood that x and r have a common convex range, such as all real numbers, the non-negative real numbers, or the real numbers between 0 and 1, inclusive or not. (A reader not familiar with convex sets and convex functions may find [47, Appendix 2] helpful here and elsewhere in the present article.) What (4.5) requires is that, for each fixed x, the function $I(x:r)$, which is linear in r, lies strictly below the function J except at x, where the linear function and J have the common value $I(x;x)=J(x)$. In short, $I(x;r)$ is, for each x, a linear function of support of J at the point x, and only there. This implies that J is a strictly convex function of r. Such a function J often has only one linear function of support at a given x, its tangent at x, but wherever J has a corner, it has more than one support.

Conversely, let J be any strictly convex function of r, well behaved at the endpoints of the range of r if there are any (that is, neither discontinuous nor vertical there), and let $I(x;r)$ be any support of J at x. This I evidently satisfies (4.5). (Examples of bad behavior at the endpoint 0 when the range is the non-negative reals: if $J(0)=1$ and

$J(x)=x^2$ elsewhere, then J is strictly convex but discontinuous and very much without a linear function of support at 0; if $J(x)=-x^{1/2}$, J is strictly convex and continuous but still has no linear function of support at 0.)

Any convex function J has, at each interior x, left and right derivatives $J_L(x)$ and $J_R(x)$. If $x<y$, $J_L(x)\leq J_R(x)\leq J_L(y)\leq J_R(y)$, and the middle inequality is strict if J is strictly convex. The slopes of the supports at x are the numbers between $J_L(x)$ and $J_R(x)$ inclusive. There is just one such slope if and only if this interval degenerates to a single number, which is then the derivative $J'(x)$. Since $b(x)$ is between $J_L(x)$ and $J_R(x)$, the function b is nondecreasing—strictly increasing if J is strictly convex. Consequently, except on an at most denumerable set of points, b is continuous, and except at just those points, b is the derivative of its indefinite integral. The points of continuity of b are the points of differentiability of J, and at those points, $J'(x)=b(x)$.

The function b determines J and c except for an additive constant k; thus

$$J(x) = \int_{y\leq x} b(y)dy + k = b(x)x + c(x) + k. \tag{4.6}$$

If J is twice differentiable at x, then, as (4.6) makes clear, b and c are differentiable there, and therefore

$$b(c) = J'(x) = b(x) + b'(x)x + c'(x). \tag{4.7}$$

Whence

$$b'(x)x = - c'(x). \tag{4.8}$$

If, therefore, J' is absolutely continuous (that is, an indefinite integral of its own derivative where defined), and f denotes J''; then

$$I(x;r) = - r\int_{x\leq\rho} f(\rho)d\rho + \int_{x\leq\rho} \rho f(\rho)d\rho + dr + e, \tag{4.9}$$

where d and e are constants. Thus, under a rather mild regularity hypothesis, (4.3) is the general solution of (4.5), except for a relatively unimportant linear term. This term expresses the possibility of the experimenter's making the subject outright gifts (possibly negative) of d units of the commodity and e units of cash. Such a gift can have no rational influence on the subject's choice (if his utility is linear in the commodity and in cash), though it might play some practical role, such as insuring that an experimental subject will in net be paid, not penalized, for his cooperation.

Though the integral approach to the problem is of some interest and is not of seriously limited generality for one commodity (especially when extended by Stieljes-like integrals), it does not, over all, seem as useful and informative as (4.5) and will not be central to this article.

A convenient form for $I(x;r)$ is often

$$\begin{aligned} I(x;r) &= J(x) + b(x)(r-x) \\ &= J(x) + J'(x)(r-x), \end{aligned} \tag{4.10}$$

where the final line is applicable only if $J'(x)$ exists.

The loss to the subject if he (irrationally) replies with x when his rate is actually r is

$$\begin{aligned} L(x;r) &= I(r;r) - I(x;r) \\ &= J(r) - I(x;r) \\ &= J(r) - J(x) - b(x)(r-x) \\ &= J(r) - J(x) - J'(x)(r-x), \end{aligned} \tag{4.11}$$

where once more the final line is applicable only where the derivative $J'(x)$ exists.

The function L, and equivalently I, has an easily derived and useful monotonicity [6, p. 43; 30, p. 44], according to which it not only pays to choose x equal to r but to keep any unavoidable discrepancy small. Namely, if x is between r and z, then $L(x;r)\leq L(z;r)$, with strict inequality if J is strictly convex and $x\neq z$. Since L is non-negative and b is nondecreasing, the following identity makes this evident.

$$\begin{aligned} &L(z;r) - L(x;r) \\ &= J(x) - J(z) - b(z)(r-z) + J(x)(r-x) \\ &= J(x) - J(z) - b(z)(x-z) + [b(z) - b(x)](x-r) \\ &= L(z;x) + [b(z) - b(x)](x-r). \end{aligned} \tag{4.12}$$

5. SOME SPECIAL FORMS FOR THE ELICITING FUNCTION

5.1 Loss as a Function of Discrepancy; $L(x;r)=H(x-r)$

One suggestive, and possibly desirable, way of limiting the choice of the convex function J is to require the subject's loss $L(x;r)$, if he replies with x when r is his true rate, to be a function only of his discrepancy $x-r$. Let us then investigate the consequences of supposing that $L(x;r)=H(x-r)$ for some function H, non-negative and 0 at 0 but not 0 everywhere. As the remaining paragraphs of this subsection are devoted to showing, this condition is so restrictive that $L(x;r)$ must be of the form $k(x-r)^2$ for some positive constant k. (This result has independently been derived with slightly less generality by Brown [6].)

According to (4.11),

$$H(x-r) = J(r) - J(x) - b(x)(r-x). \tag{5.1}$$

To see the sort of implication latent in (5.1), suppose for the moment that b and J are defined for all sufficiently small $|x|$ and $|r|$; as is rather intuitive and as will later be explicitly shown, this entails no real loss of generality. Under the simplifying assumption, an instance of (5.1) for $|x|$ and $|r|$ small is

$$\begin{aligned} H(x-r) &= H((-r)-(-x)) \\ &= J(-x) - J(-r) - b(-r)(r-x). \end{aligned} \tag{5.2}$$

Therefore

$$\begin{aligned} &J(r) + J(-r) + b(-r)r - b(-r)x \\ &= J(x) + J(-x) - b(x)x + b(x)r. \end{aligned} \tag{5.3}$$

In particular,

$$J(r) + J(-r) + b(-r)r = 2J(0) + b(0)r, \quad (5.4)$$

$$2J(0) - b(0)x = J(x) + J(-x) - b(x)x. \quad (5.5)$$

By means of (5.4) and (5.5), simplify (5.3) thus.

$$\begin{aligned} 2J(0) &+ b(0)r - b(-r)x \\ &= 2J(0) - b(0)x + b(x)r; \end{aligned} \quad (5.6)$$

or

$$[b(x) - b(0)]r = [b(0) - b(-r)]x, \quad (5.7)$$

so $b(x)$ is linear in x, at least for $|x|$ small.

To see that b is linear near every z in the interior of its domain of definition, and therefore linear throughout the interior, let $\hat{x}=x-z$, $\hat{r}=r-z$, $\hat{b}(u)=b(z+u)$, and $\hat{J}(u)=J(z+u)$. Since $(x-r)=(\hat{x}-\hat{r})$, the circumflexed functions and variables also satisfy (5.1), so $\hat{b}$ is linear for $|\hat{x}|$ small, that is, b is linear near z, as asserted, and consequently linear everywhere, with the (temporarily) possible exception of the endpoints of the domain of definition.

Armed with this information and the fact that H is nonnegative but not identically 0, a reader familiar with the theory of the Cauchy-Hamel equation [1, Section 2.1] could easily show (from (5.1) alone) that $H(x-r)=k(x-r)^2$ for some positive k. But it is more elementary to recall that $b(x)=2kx+l$ is the slope of a line of support of J at x. Since this slope is continuous in x, J is differentiable, and according to (4.6),

$$J(x) = kx^2 + lx + m, \quad (5.8)$$

with some positive k, and according to (4.10),

$$\begin{aligned} I(x;r) &= J(x) + J'(x)(r-x) \\ &= kx^2 + lx + m + (r-x)(2kx+l) \\ &= (2kx+l)r - kx^2 + m. \end{aligned} \quad (5.9)$$

And indeed, for any J of the form (5.8), according to (4.11),

$$\begin{aligned} L(x;r) &= J(r) - I(x;r) \\ &= kr^2 + lr + m - \{[2kx+l]r - kx^2 + m\} \\ &= k(x-r)^2, \end{aligned} \quad (5.10)$$

as anticipated.

(The possibility that the linearity of b might fail at the endpoints of the interval of definition has been left open but can easily be removed by means of (5.1) and what has now been proved.)

5.2 Symmetry; $L(x; r)=L(r; x)$

Another suggestive condition is that the loss for replying with x when r applies should be the same as that for replying with r when x applies; that is, $L(x; r)=L(r; x)$. This condition too leads to (5.8), (5.9), and (5.10).

Briefly, the demonstration is as follows.

$$\begin{aligned} 0 &= L(x;r) - L(r;x) \\ &= [b(r) + b(x)](x-r) + 2[J(r) - J(x)]. \end{aligned} \quad (5.11)$$

Consider the two equations that result from (5.11) on replacing (x, r) first by (r, z) and then by (z, x), and add all three equations to conclude that

$$b(x)(z-r) + b(r)(x-z) + b(z)(r-x) = 0. \quad (5.12)$$

Therefore $b(x)$ is linear, and the passage to (5.8), (5.9), and (5.10) can be made as before. Or it can be made thus. When the form $b(x)=2kx+1/2$ is substituted into (5.11), $J(x)$ is seen to be of the form kx^2+lx+m.

Quadratic J's are, then, not only the simplest convex functions from the algebraic point of view; they are characterized also by the condition that $L(x;r)$ is a function of $x-r$ and by the condition that $L(x;r)=L(r;x)$.

5.3 Ratio Discrepancy; $L(x; r)=H(r/x)$

When x ranges over the positive real numbers, the possibility that $L(x; r)$ is of the form $H(r/x)$ might be interesting. This condition too is very restrictive; it implies that J, I, and H are of the compatible forms:

$$J(r) = m + lr - k\log r, \quad (5.13)$$

$$I(x;r) = (m + k - k\log x) + (l - k/x)r, \quad (5.14)$$

$$H(u) = k(u - 1 - \log u), \quad (5.15)$$

with k positive. The demonstration occupies the rest of this subsection.

To begin with,

$$H(r/x) = J(r) - J(x) - J'(x)(r-x), \quad (5.16)$$

except for the at most denumerable set of x where $J'(x)$ is not well defined. Therefore, as is seen on differentiating with respect to r,

$$(1/x)H'(r/x) = J'(r) - J'(x), \quad (5.17)$$

whenever both derivatives on the right exist. Since these two derivatives do exist except on an at most denumerable set, a sequence of conclusions follows one after another:
H is differentiable everywhere; so is J; (5.16) and (5.17) hold without exception; (5.16) can be differentiated on both sides with respect to x;

$$-(r/x^2)H'(r/x) = -(r-x)J''(x); \quad (5.18)$$

J is twice differentiable everywhere; and

$$rJ'(r) - rJ'(x) = x(r-x)J''(x). \quad (5.19)$$

Regarded as a differential equation for J in r, (5.19) easily implies (5.13); (4.10) then implies (5.14); and (4.11) implies (5.15). Since J as defined by (5.13) is strictly convex and results in (5.14) and (5.15), the system is indeed compatible.

5.4 Attempted Generalization; $L(x; r)=H(g(r)-g(x))$

Now consider the seemingly rather general possibility that $L(x; r)$ is of the form $H(g(r)-g(x))$. The special cases $g(x)=-x$ and $g(x)=\log x$ have already been considered. What other functions g are compatible with some strictly convex J? I am content here to examine the question under the simplifying assumption that g is differentiable and has a strictly positive derivative on the domain of J.

The derivation of (5.19) can be recapitulated to conclude that

$$\frac{J'(r)}{g'(r)}=\frac{(r-x)}{g'(x)}J''(x)+\frac{J'(x)}{g'(r)}\cdot \qquad (5.20)$$

Since the right side of (5.20) cannot change with x, and since $J'(x)$ is not constant, $1/g'(r)$ must be linear in r.

If this linear function is constant, it can without loss of generality be taken to be 1, with return to (5.8), (5.9), and (5.10). Incidentally, (5.20) specialized by setting $g'\equiv 1$ provides thus an alternative route to (5.8), (5.9) and (5.10).

If $1/g'(r)$ is linear but not constant, then it is to all intents and purposes of the form $(r-z)$ for some z below the domain of J or of the form $(z-r)$ for some z above the domain of J. In the first case, for example, $g(r)=\log(r-z)+\text{const.}$; and $L(x;r)$ is of the form $G((r-z)/(x-z))$, which is virtually the form that led to (5.13), (5.14) and (5.15). The introduction of g has therefore led to no really new forms of I, J, and L.

For J defined on the unit interval, it might seem interesting to seek $L(x;r)$ of the form $H(r(1-x)/(1-r)x)$. But that would imply the existence of a $g(r)$ of the form $\log r-\log(1-r)$, which has been shown to be impossible.

6. PROBABILISTIC INTERPRETATION

6.1 Application to a Probability as a Rate of Substitution

Suppose that an experimenter, in an effort to elicit the probability p that a subject associates with an event D, invites the subject to choose a number x and promises to pay him $Y(x)$ in case D obtains and $Z(x)$ in case D does not obtain. For what pairs of functions Y and Z will it be to the subject's interest to choose x equal to p?

If p is interpreted as the subject's rate of substitution of dollars for the commodity consisting of dollars that are contingent on D, then the experimenter is in effect offering the subject $Z(x)$ dollars in cash and $Y(x)-Z(x)$ units of the commodity. The worth of such a gift to the subject is

$$\begin{aligned} I(x;p) &= [Y(x)-Z(x)]p+Z(x) \\ &= Y(x)p+Z(x)(1-p), \end{aligned} \qquad (6.1)$$

which is an instance of (4.4) with p, $Y(x)-Z(x)$, and $Z(x)$ playing the roles of r, $b(x)$, and $c(x)$.

Therefore Y and Z accomplish the objective if and only if

$$\begin{aligned} J(p) &= Y(p)p+Z(p)(1-p) \\ &= Y(p)-(Y(p)-Z(p))(1-p) \\ &= Z(p)+(Y(p)-Z(p))p \end{aligned} \qquad (6.2)$$

is a strictly convex function of p, and $I(x;p)$ is in p a linear function of support of J at x. At values of p where J is differentiable, according to (6.1),

$$J'(p)=Y(p)-Z(p), \qquad (6.3)$$

so at such values, according to (6.2) and (6.3),

$$Y(p)=J(p)+(1-p)J'(p) \qquad (6.4)$$

$$Z(p)=J(p)-pJ'(p). \qquad (6.5)$$

The loss entailed by choice of x when p applies is

$$L(x;p)=[Y(p)-Y(x)]p+[Z(p)-Z(x)](1-p), \qquad (6.6)$$

which is ordinarily—that is, if J is differentiable—

$$L(x;p)=J(p)-J(x)-J'(x)(p-x), \qquad (6.7)$$

as in (4.11).

The conditions that J is quadratic, that $L(x;p)$ is a function of $p-x$, and that $L(x;p)\equiv L(p;x)$ are, according to Section 5, all equivalent. In this special case, (6.4) and (6.5) can be put in the suggestive forms

$$Y(p)=m'-k(1-p)^2, \qquad (6.8)$$

$$Z(p)=m-kp^2, \qquad (6.9)$$

which correspond to

$$J(p)=m'p+m(1-p)-kp(1-p). \qquad (6.10)$$

In any real application, p is between 0 and 1, and it might seem natural to confine the range of choice of x to the interval from 0 to 1 inclusive. However, if x is not so confined, a subject who chooses $x<0$ or $x>1$ exposes himself to utterly unnecessary loss for any strictly convex J. In fact, according to the monotonicity pointed out in the final paragraph of Section 4, for any p in the interval $[0, 1]$, $L(x;p)>L(1;p)$ if $x>1$, and $L(x;p)>L(0;p)$ if $x<0$.

6.2 Application to an Expectation as a Rate of Substitution

Since the probability of an event D is the expectation of the same D regarded as an indicator, extension of the method of eliciting probabilities just discussed to the elicitation of the personal expectation r of any random quantity V is to be anticipated. And this is indeed straightforward. If the experimenter offers to pay the subject $b(x)V+c(x)$ for any x chosen by the subject, the worth to the subject of choosing x is

$$I(x;r)=b(x)r+c(x), \qquad (6.11)$$

so the loss for choosing x instead of r is

$$L(x;r)=[b(r)-b(x)]r+[c(r)-c(x)]. \qquad (6.12)$$

For this to be positive if and only if x differs from r means, as shown in Section 4, that $J(r)=b(r)r+c(r)$ is strictly convex and $b(x)r+c(x)$ is in r a linear function of support of J at x.

If, for example, the subject feels certain that $V\leq\alpha$ for some constant α, his $r=E(V)$ cannot exceed α. And in fact it is to his advantage to choose α rather than any x

larger than α, as can be verified thus. The difference in worth between these two choices is

$$[b(\alpha) - b(x)]V + [c(\alpha) - c(x)] = L(x; V) - L(\alpha; V). \quad (6.13)$$

In view of the monotonicity mentioned at the end of Section 4, (6.13) is therefore negative if $x > \alpha$. Thus whatever the actual value of V may be—not greater than α—the subject will receive more for choosing α than for choosing any larger number.

6.3 The Most General Eliciter of an Expectation

Is $b(x)v+c(x)$ the only possible form for a function $S(x; v)$ for which

$$E(S(x; V)) < E(S(r; V)) \quad (6.14)$$

for all x different from $r=E(V)$? The answer cannot quite be "yes," because to any such S it is clearly possible to add $f(V)$ with f any function for which $E(f(V))$ is finite for the class of distributions envisaged for V. But little if any further extension is possible if (6.14) is to hold for a reasonably large class of distributions for V, as the next four paragraphs demonstrate. Even this slight extension is nugatory in case V is the indicator of an event, that is, in case $E(V)$ is a probability, because when V takes only two values, any function f is linear on that pair of values.

To begin with, let $S(x; v)$ be defined for all x and v in an interval, and let (6.14) hold for every V that is subject to a 2-point distribution in that interval. Specialized to a V that takes the values v and v' with probabilities p and $\bar{p}=1-p$, (6.14) takes the form:

$$pS(x; v) + \bar{p}S(x; v') < pS(pv + \bar{p}v'; v) + \bar{p}S(pv + \bar{p}v'; v'), \quad (6.15)$$

unless $x=pv+\bar{p}v'$.

For x interior to the interval of definition and S differentiable in x,

$$p\frac{\partial}{\partial x}S(x; v) + \bar{p}\frac{\partial}{\partial x}S(x; v') = 0 \quad (6.16)$$

if $x=pv+\bar{p}v'$. That is, if $v \le x \le v'$ and the derivatives exist at x, then

$$(v' - x)\frac{\partial}{\partial x}S(x; v) = (v - x)\frac{\partial}{\partial x}S(x; v'). \quad (6.17)$$

Therefore, if S is everywhere differentiable with respect to x (more generally, if S is absolutely continuous in x for each v), then $S(x; v)$ is of the form $b(x)v+c(x)+f(v)$ as anticipated. This conclusion obtains even without any differentiability assumption, as will be shown in the next paragraph for those who share an interest in such points.

According to (6.15), the function

$$(v' - x)S(x; v) + (x - v)S(x; v') \quad (6.18)$$

is convex in x and has

$$S(x; v') - S(x; v) \quad (6.19)$$

among its slopes of support at x. Therefore the functions (6.18) and (6.19) are of bounded variation in x, whence so is $S(x; v)$ for each v. Let S_0 and S_1 be the singular and the absolutely continuous parts of S with respect to x, rendered unique by the convention that $S_0(x_0; v)=0$ for some x_0 and for arbitrary v. Since (6.18), being convex, is absolutely continuous, $S_0(x; v)=(v-x)S_0(x; v')/(v'-x)$ and is therefore of the form $b_0(x)v+c_0(x)$. (Now, the argument employed when S was assumed to be absolutely continuous applies almost unmodified to S_1, the absolutely continuous part of S.)

An important and widely known example is

$$S(x; v) = -(x - v)^2 = -x^2 + 2xv - v^2.$$

For it is central to the theory of least squares that this S satisfies (6.14) according to the following familiar calculation.

$$E[(x - V)^2] = E\{[(x - E(V)) - (E(V) - V)]^2\} = (x - E(V))^2 + E[(E(V) - V)^2]. \quad (6.20)$$

7. DECISION THEORY

A different approach to the problem of determining those functions Y and Z that encourage a subject to reveal his true p is implicit in statistical decision theory, as will now be explained.

Imagine a person in an economic situation in which he is free to choose one of several acts and knows that if he chooses the act a he will receive a payment of $Y(a)$ dollars in case the event D obtains and $Z(a)$ dollars in case D does not obtain. If the person's probability for D is p, the worth of the act a to him is

$$I(a; p) = Y(a)p + Z(a)(1 - p), \quad (7.1)$$

a linear function of p. The worth to him of the situation in which he is free to choose among a finite set of acts a is therefore the function

$$J(p) = \max_a I(a; p). \quad (7.2)$$

Indeed, any act a for which the maximum is attained is worth $J(p)$ to him and no available act is worth more. The function $J(p)$ is evidently convex.

When an infinite set of acts is envisaged, little changes, especially if the set is such that the maximum required by (7.2) is attained. With an unrestricted infinite set of acts, J can be an arbitrary continuous, convex function. (If the end points $p=0$ and $p=1$ are to be included and if the maximum is to be attained, J cannot be permitted to be vertical at those points.) If J is strictly convex, a person choosing a support of J will reveal his value of p; for no linear function supports a strictly convex function at more than one point.

The original problem of characterizing functions Y and Z that will elicit the person's true p can now be envisaged thus. For each number x, say in the interval from 0 to 1, an act is made available to the subject that will pay him $Y(x)$ dollars in case D obtains and $Z(x)$ dollars otherwise.

For what functions Y and Z will it be optimal for the person to choose x equal to his p, no matter what p may be? As the preceding paragraph shows, it is necessary and sufficient that there be some strictly convex J for which $Y(x)p+Z(x)(1-p)$ is a line of support at x; and this was the main conclusion of Section 6.

8. AN ALTERNATIVE DEFINITION OF PROBABILITY

The preceding section suggests an alternative way to define personal probability and to argue that personal probability exists. This avenue has been pointed out and explored by de Finetti, for example, in [13, 17].

Imagine in fact that the very notion of the probability p attached by a person to the event D has not yet been defined, and consider a person confronted with the choice among all those acts a for which the linear functions $I(a; p)$ nowhere exceed a specified convex function $J(p)$ (without vertical endpoints). Suppose that there is one such act a to which he prefers no other. It would be unreasonable for this a not to be one for which $I(a; p)$ is a line of support of $J(p)$ at some point p_0. For otherwise there would be an a' for which $I(a'; p)$ is a line of support of J and is parallel to $I(a; p)$, in which case $Y(a') > Y(a)$ and $Z(a') > Z(a)$, so that a' is clearly superior to a.

If J is strictly convex, the unique p_0 for which $I(a; p)$ is a line of support of J can be defined as the person's personal probability of D. But this raises three important questions: Could a different function J^* lead to a different p_0? Does this definition lead back to the idea of probability as a rate of substitution? Is probability as thus defined a probability measure?

To progress toward answering the first two questions, suppose the person participates simultaneously in two decision problems of the kind under discussion, one determined by J_1 and one determined by J_2, where J_1 and J_2 are convex but not necessarily strictly convex. Assume that he acts in each of the two component problems as he would if he were faced with that problem alone.

(This assumption is not altogether unobjectionable; for it may imply that the person's utility function is linear in money. But such linearity assumptions are made almost throughout the present paper and are presumably tolerable if only moderate sums of money are involved. In the purely mathematical formulation of the decision problems no precautions have been, or need be, taken to keep these sums moderate, but it is fairly clear how such precautions could be taken in applying the theory or how the devices mentioned at the end of Section 3 might be invoked. I do not attempt great caution about this point, because my object in this section is only to touch briefly on an approach to personal probability that may be more suggestive and more practical in some respects than such formally more rigorous approaches as those reported on in [47, Ch. 3].

The linear function of support $I_1(a_1; p)$ chosen for J_1 and the linear function of support $I_2(a_2;p)$ chosen for J_2 should support J_1 and J_2 respectively at some common point p^*, as will be argued. The person, in choosing a_1 and a_2, has in effect chosen an act a for which $I(a; p) = I_1(a_1; p)+I_2(a_2; p)$. Some linear function of support $I^*(p)$ of $J=J_1+J_2$ at some point p^* is either the same as, or everywhere higher than, $I(a; p)$. Since, as is not hard to see, I^* can be represented as the sum of linear functions of support of J_1 and J_2 at p^*, I^* represents an act (that is, a choice of a_1 and a_2) that was available to the person. Therefore his choice of a_1 and a_2 is discredited, or seen to be inadmissible, unless $I^*(p)=I(a; p)$ for all p, which is impossible unless $I_1(a_1; p^*)=J_1(p^*)$ and $I_2(a_2; p^*) = J_2(p^*)$ as asserted. This conclusion will yield affirmative answers to the first two questions.

First, as is now evident, two strictly convex functions cannot lead to different values of p if the linear-utility assumption holds and if the person is coherent.

Second, suppose p^* has been determined by means of a strictly convex function J_1, and consider also the decision problem defined by the broken-line convex function J_r with $J_r(p) = \max(rq, pq)$, where r is a positive fraction and q is a positive number. According to what was proved in the paragraph before last, the person will prefer q dollars contingent on D to rq dollars outright if $r > p^*$ and vice versa if $r < p^*$, so p^* is indeed a rate of substitution.

A certain approach to the third question, whether probability as here defined is a probability measure, is best postponed to the next section. But it can be argued now that since personal probabilities defined as rates of substitution constitute a probability measure, the same must be true for the equivalent new definition.

De Finetti [13] has shown how the approach of this section applied to conditional probability.

9. THE SIMULTANEOUS ELICITATION OF SEVERAL RATES

9.1 Vector Rates

When several rates r_s are concerned, each can be independently elicited by means of a function $I_s(x_s; r_s)$ as in Section 4, but there are other possibilities.

Generalizing from Section 4, let $r=\{r_s\}$ be a finite sequence of rates, which can advantageously be thought of as a vector. (Infinite sequences or even functions on an arbitrary domain might also have a useful interpretation, for instance, as distribution functions or densities. Vectors without any interpretation as functions could also be handled and might have applications. Extensions of this sort are presented by Hendrickson [30].) Let $x=\{x_s\}$ be a sequence of possible numerical responses of a subject, a response vector. The experimenter's aim is to provide the subject with an incentive to choose $x=r$ by offering to give $c(x)$ units of cash and $b_s(x)$ units of the sth commodity for whatever x is chosen by the subject.

The experimenter may already know something about r. For example, in the important case of probabilities, he knows that each probability is between 0 and 1; and when the r_s are probabilities of the elements of a partition, he knows that they add up to 1. In other cases, it will often be known that each $r_s > 0$. For still one more example, the rate for a certain high quality commodity may be plainly higher than that of certain other commodities.

It may be to the experimenter's advantage to exclude all, or at least some, unreasonable values of x. In particular, he might confine acceptable responses x to some convex set K—a scheme that gives adequate flexibility for the examples just mentioned. For the moment, assume that K has at least one extrinsic interior point z, that is, a point for which $z+\Delta r=\{z_s+\Delta r_s\}$ is in K for all sufficiently small vectors Δr. This temporary simplification excludes at least one important example, that of probabilities p_s so constrained that $\Sigma p_s=1$.

With the exception of the expression of I by integrals, the whole of Section 4 will now easily be paraphrased with sets, or bundles, or commodities (or, more abstractly, vector commodities) playing the role that a single commodity played there.

Though the experimenter does not initially know the vector r, he can in effect offer the subject the income $I(x; r)$ for choosing the vector x, where

$$\begin{aligned} I(x; r) &= \Sigma_s\, b_s(x) r_s + c(x) \leq \Sigma\, b_s(r) r_s + c(r) \\ &= I(r; r) = J(r), \end{aligned} \tag{9.1}$$

with equality if and only if $x=r$.

According to (9.1), $I(x: r)$ is, for each x, a strict linear function of support at x for the function J. Therefore, J is strictly convex, and virtually all strictly convex functions do lead to at least one I. If K has no boundary points—and it is seldom if ever important to include boundary points in K—then every strictly convex function will serve. If K does have boundary points, then discontinuities at the boundary points and milder sorts of misbehavior incompatible with linear functions of support at the boundary points are to be excluded.

If J is differentiable at the nonboundary point x of K or, equivalently, if the $b_s(x)$ are continuous at x, then

$$b_s(x) = \frac{\partial}{\partial x_s} J(x), \tag{9.2}$$

and

$$c(x) = J(x) - \Sigma x_s \frac{\partial}{\partial x_s} J(x). \tag{9.3}$$

If the $b_s(x)$ are continuous for all such x, they obviously determine J through (9.2), except for an additive constant. (But continuity is not needed for the conclusion that J is determined except for an additive constant by the b_s.)

The loss incurred by the subject on choosing x when r applies is

$$\begin{aligned} L(x; r) &= [J(r) - J(x)] - \Sigma\, b_s(x)(r_s - x_s) \\ &= [J(r) - J(x)] - \Sigma\, (r_s - x_s) \frac{\partial}{\partial x_s} J(x), \end{aligned} \tag{9.4}$$

the last line applying only where J is differentiable

The condition that $J(x)$ be quadratic, that is, of the form

$$J(x) = \sum_{s,t} k_{s,t} x_s x_t + \Sigma_s l_s x_s + m \tag{9.5}$$

is equivalent to the condition that $L(x; r)$ depends only on $x-r=\{x_s-r_s\}$ and to the condition that $L(x; r) = L(r; x)$.

Much the same can be said even without assuming that K has extrinsic interior points, with slight differences because it may no longer be possible to change one x_s without changing other components of x. Equation (9.1) and the paragraph following it remain in force, but it now may be possible to change the individual $b_s(x)$ without really affecting I. The meaning of (9.1) and (9.4) is therefore better conveyed by

$$I(x; r) = b(x)(r) + c(x) \leq b(r)(r) + c(r) = J(r), \tag{9.6}$$

$$L(x; r) = J(r) - J(x) - b(x)(r - x), \tag{9.7}$$

where for each x, $b(x)(r)$ is linear and homogeneous in the vector r and $I(x; r)$ is, as always, a linear function of support of J at x.

9.2 Vector Expectations

The case of a sequence $V=\{V_s\}$ of random numbers considered as commodities is important. A little more generally, V can be a random vector whose values lie in a finite dimensional convex set K. The random payment to the subject who chooses the vector x is

$$b(x)(V) + c(x), \tag{9.8}$$

which in V is linear and supports a strictly convex J at x. Where J is differentiable, (9.8) can be suggestively written as

$$J(x) + J'(x)(V - x). \tag{9.9}$$

As in Subsection 6.3, it can be asked whether there are functions $S(x; v)$ that elicit the expected value of a vector V other than S of the form $I(x; v)+f(v)$, where I is as in (9.6). The negative answer given in Subsection 6.3 is not hard to generalize, at least under generous regularity hypotheses.

The interesting and familiar instance of quadratic S mentioned in Subsection 6.3 has a hardly less familiar extension to the present, multidimensional, situation. For if S is a homogeneous, strictly negative-definite, quadratic function of $x-v$, then

$$\begin{aligned} E[S(x - V)] &= S(x - E(V)) + E[S(V - E(V)] \\ &> E[S(V - E(V)] \end{aligned} \tag{9.10}$$

if $x \neq E(V)$.

A subject who feels sure that V is in a specified closed convex subset K^* of K will have $E(V)$ in K^* and therefore ought not choose an x not in K^*. But, under suitable regularity hypotheses, such an x is also discredited in the deeper sense that there is a v in K^* for which $I(v; r)$

$>I(x; r)$ for every r in K^*. This is demonstrated in the next paragraph, which is followed by one showing that some regularity hypothesis is indispensable.

Suppose that J is differentiable in K^* and that the infimum of $L(x; r)$ as a function of r in K^* is attained, say at $r=v$. Then, for all r in K^*,

$$\begin{aligned} L(x; r) &= J(r) - J(x) - b(x)(r - x) \\ &\geq J(v) - J(x) - b(x)(v - x) = L(x; v) \end{aligned} \tag{9.11}$$

$$\begin{aligned} 0 &\leq L(x; r) - L(x; v) \\ &= [J'(v) - b(x)](r - v) + o(r - v). \end{aligned} \tag{9.12}$$

Therefore, $[J'(v)-b(x)](r-v)\geq 0$ for all r in K^* sufficiently close to v; but if such a linear inequality holds for r in K^* close to v, it holds for all r in K^*. Put the facts together, thus.

$$\begin{aligned} L(x; r) &- L(v, r) \\ &= J(v) - J(x) - b(x)(r - x) + J'(v)(r - v) \\ &= [J'(v) - b(x)](r - v) + L(x; v) \\ &\geq 0, \end{aligned} \tag{9.13}$$

as asserted.

(In one dimension, as has been seen in Subsection 6.2, no regularity hypothesis is required for this conclusion, but the following counterexample shows that some such hypothesis is needed in two and more dimensions. Let $J(x, y)=|x|+|y|+\epsilon(x^2+y^2)$ for some small positive ϵ. This J is strictly convex over the whole plane, and I can be consistently defined thus.

$$I(x, y; r, s)=J(x, y)+(\text{sgn } x+2\epsilon x)(r-x)+(\text{sgn } y+2\epsilon y)(s-y).$$

Let K be the convex set

$$\{x, y: 1 \leq 2x + y; x \leq 2, y \leq 3\}.$$

Then $I(0, 0; r, s)=0$ for all (r, s), but, for each (x, y) in K,

$$\inf_{(r,s)\in K} I(x, y; r, s) \leq -1 + 0(\epsilon).)$$

The significance of the general conclusion in case of regularity is particularly vivid if J is quadratic. For then, in terms of the Euclidean distance associated with the quadratic function J, (9.13) says that the point v in K^* closest to x is closer to each point of K^* than x is. This is geometrically rather evident and is easy to prove directly or as an instance of the general argument.

To apply the argument about K^*, if the person is sure that a linear equality or inequality is satisfied by several random variables $V_1, \cdots, V_n$, he will expose himself to needless loss if the $x_1, \cdots, x_n$ with which he responds to any regular J do not satisfy this equality or inequality. In particular, if some V_s is sure to be 1, the person should choose $x_s=1$; if V_s is the indicator of an event, he should choose x_s between 0 and 1; if V_s is the indicator of the union of two disjoint events of which V_j and V_k are the indicators, he should so choose that $x_s=x_j+x_k$. As these remarks show, the program for defining probabilities in Section 8 does lead to a probability measure for any person who does not blunder. (More completely, the remarks do lead directly to the anticipated conclusion when the program is applied to any J that is regular. But, as was shown in Section 8, all strictly convex J elicit the same probabilities.)

9.3 Probability Distributions as Vector Expectations

Consider now the case in which the V_s $(s=1, \cdots, n)$ are the indicators of the elements of a partition $\{D_s\}$. All previous literature on rate elicitation seems to be confined to this important case. (A few references on the method not mentioned elsewhere in this article are [7, 42, 54].)

Since it is (especially patently for regular J) wasteful in the present case for the person to choose any $x=\{x_s\}$ other than a probability distribution (that is, an x for which $x_s\geq 0$ and $\Sigma x_t=1$), K can for many applications be taken to be the simplex of all probability distributions p. If $J(p)$ is a differentiable function on K, the person who chooses x will receive $I(x; V)$, which can, in a somewhat figurative but not unnatural notation, be written,

$$\begin{aligned} I(x; V) &= J(x) + J'(x)(V - x) \\ &= J(x)\, \Sigma_t\, J_t'(x)(V_t - x_t); \end{aligned} \tag{9.14}$$

this is figurative because x cannot be varied one coordinate at a time and remain in the simplex. In case D_s is the event in the partition that actually obtains, this is,

$$I(x; V) = J(x) - \Sigma_t J_t'(x)x_t + J_s'(x). \tag{9.15}$$

The numbers $J_s'(x)$ here are not necessarily derivatives and are determined only up to an additive constant. (They can of course be so chosen that $\Sigma J_t'(x)=0$, but the symmetry of that particular choice ought not to be invested with much importance.) If J is originally defined not only on the simplex K but, say, for all positive n-tuples of numbers, then the $J_s'(x)$ can be taken to be $\partial J(x)/\partial x_s$, which would not be meaningful on K alone.

One suggestive choice for J, whether on K or on a larger set, is $J(x)=\Sigma x_t^2$. In this case, (9.15) on K becomes

$$\begin{aligned} I(x; V) &= -\, \Sigma_t\, x_t^2 + 2x_s \\ &= -\, (1 - x_s)^2 - \textstyle\sum_{t\neq s} x_t^2 + 1 \\ &= -\, \Sigma_t\, (V_t - x_t)^2 + 1, \end{aligned} \tag{9.16}$$

that is, 1 minus the square of the usual Euclidean distance from x to the vector $(0, 0, \cdots, 1, \cdots, 0, 0)$ representing the D_s that actually obtains. The final constant 1 is of course not very important in (9.16). Specialized to a single event D (with indicator V) and its complement, (9.16) without its 1 becomes

$$\begin{aligned} I(x; V) &= -\, [V - x]^2 - [(1 - V) - (1 - x)]^2 \\ &= -\, 2(V - x)^2 \\ &= \begin{cases} -2(1 - x)^2 & \text{if } D \text{ obtains} \\ -2x^2 & \text{if } D \text{ does not obtain.} \end{cases} \end{aligned} \tag{9.17}$$

If each of the events D_s is treated separately according to (9.17), the net effect is practically the same as that of using $J=\Sigma x_t^2$; it is exactly the same as using $2J-2$.

The most general quadratic function of n variables is

$$J(x) = \sum_{s,t} k_{st}x_s x_t + \sum_s l_s x_s + m. \qquad (9.18)$$

To this J corresponds

$$I(x; V) = J(V) - \sum_{s,t} k_{st}(V_s - x_s)(V_t - x_t). \qquad (9.19)$$

If further, J is symmetric in the x_s, (9.18) specializes to

$$J(x) = k' \sum_s x_s^2 + k''(\sum_s x_s)^2 + l \sum_s x_s + m, \qquad (9.20)$$

which is convex if and only if $k'>0$ and $k'+nk''>0$. For this J,

$$\begin{aligned} I(x; V) &= J(V) - k' \sum_s (V_s - x_s)^2 \\ &\quad - k''[\sum_s (V_s - x_s)]^2. \end{aligned} \qquad (9.21)$$

If the V_s are the indicators of a partition, $\Sigma V_s^2 = \Sigma V_s = 1$; so $J(V) = k'+k''+l+m$, which simplifies (9.21). If further $\Sigma x_s = 1$ as it "ought to," then (9.21) simplifies to

$$(k' + k'' + l + m) - k' \sum_s (V_s - x_s)^2, \qquad (9.22)$$

which is not essentially different from (9.16).

9.4 Separated Income for Distributions

It is somewhat attractive to seek functions J on the simplex K such that when D_s obtains the income of the subject will depend on x_s alone. The pertinent facts have been known for some time, a review of them here without any supplementary hypotheses may be useful. (An early reference is [37], which attributes the main fact to Andrew Gleason.)

If $n=2$, the condition is plainly vacuous, but for $n>2$ it is very severe—strictly speaking, not quite attainable. What is wanted is n functions f_s on the interval from 0 to 1 (inclusive if possible) such that, for any distribution p,

$$\sum_t f_t(x_t)p_t < \sum_t f_t(p_t)p_t \qquad (9.23)$$

for every distribution x different from p. As will be shown, after preliminary discussion, if this is to hold for all p with each p_s strictly positive, then

$$f_s(z) = k \log z + l_s, \qquad (9.24)$$

for some $k>0$. Whence

$$J(p) = k \sum_s p_s \log p_s + \sum l_s p_s; \qquad (9.25)$$

$$L(x; p) = k \sum_s p_s \log \frac{p_s}{x_s} \cdot \qquad (9.26)$$

Of course (9.23) and (9.24) are not compatible with any assignment of a finite value for $f_s(0)$.

The loss (9.26) is a constant multiple of what is sometimes called the information of the distribution p with respect to the distribution x. The scheme of elicitation implied by (9.25) seems to have been suggested first by Good [26, p. 112], who confined himself to two-fold partitions, for which the uniqueness theorem is not relevant. An interesting application and discussion are given by Mosteller and Wallace [38, Section 4.9].

It is not hard to prove (9.24) once it is known that each f_s is sufficiently regular. The next three paragraphs are devoted to proving that regularity.

Apply (9.23), for $n>2$, to distributions x and p of the special form $x=\{yw, \bar{y}w, \bar{w}/(n-2), \cdots, \bar{w}/(n-2)\}$ and $p=\{qw, \bar{q}w, \bar{w}/(n-2), \cdots, \bar{w}/(n-2)\}$, where $\bar{y}=1-y$, $\bar{q}=1-q$, $\bar{w}=1-w$, and $0<y, q, w<1$.

$$f_1(yw)q + f_2(\bar{y}w)\bar{q} < f_1(qw)q + f_2(\bar{q}w)\bar{q} \qquad (9.27)$$

if $y \neq q$. The left side of (9.27) is therefore, in q, a strict linear function of support at y of g_w, where

$$g_w(y) = f_1(yw)y + f_2(\bar{y}w)\bar{y}. \qquad (9.28)$$

Therefore, g_w is strictly convex in (0, 1). So g_w is continuous in y and, except possibly on a denumerable set, differentiable in y.

Since $f_1(yw)-f_2(\bar{y}w)$ is a slope of support of g_w at y, it is locally of bounded variation, and the same can therefore be said of $f_1(yw)=g_w(y)+\bar{y}[f_1(yw)-f_2(\bar{y}w)]$ as a function of y. Therefore each f_s has at most a denumerable number of discontinuities and is differentiable almost everywhere.

Returning to (9.28) and recalling that g_w is continuous in y, it can be seen that if f_1 is discontinuous at z, then f_2 is discontinuous at $w-z$ for all w between z and 1; so f_1, and therefore all f_s, must be continuous everywhere. Therefore, g_w is differentiable at all y. Arguing as before, if f_1 is not differentiable at z, then f_2 is not differentiable at $w-z$. At last, we know that each f_s is differentiable at all z in (0, 1). Let f_s' be the derivative of f_s.

The rest is straightforward:

$$\begin{aligned} \frac{d}{dy} g_w(y) &= f_1(yw) - f_2(\bar{y}w) \\ &= \frac{d}{dy}\{f_1(yw)y + f_2(\bar{y}w)\bar{y}\} \\ &= w\{f_1'(yw)y - f_2'(\bar{y}w)\bar{y}\} \\ &\quad + f_1(yw) - f_2(\bar{y}w) \end{aligned} \qquad (9.29)$$

$$f_1'(yw)yw = f_2'(\bar{y}w)\bar{y}w. \qquad (9.30)$$

This means that $f_s'(z)z=k$ independently of s or z, which establishes (9.24).

10. PRACTICAL SIDELIGHTS

10.1 Scope of this Section

This article, which is largely about the mathematical aspects of the procedures for eliciting personal probabilities that are now often called proper scoring rules (and also admissible probability measurement procedures) would be misleading and incomplete without some discussion of their actual and potential applications. My preparation for that is inadequate; for I have done no practical work in the area nor even followed the practically oriented literature with energy and care. But it seems incumbent on me to mention, to the best of my knowledge, the sorts of applications that have been envisaged, some difficulties that threaten them, and criteria that might help in selecting among the plethora of proper scoring rules. The subject is a ramified one, and even a cursory survey demands considerable space.

One serious omission is discussion of the experiments that have been done on proper scoring rules, for which some key references are [12; 52, Sec. 3.4 and Ch. 10–11; 55; 56].

The applications thus far envisaged have tended to emphasize the elicitation of probabilities over that of other expectations. For this reason and for vividness, I focus here on probabilities, but extension of the ideas to other expectations will often be obvious and ought to be kept in mind.

10.2 The Uses of Opinions

Strictly proper scoring rules enable us, in principle, to discover people's opinions, so possible fields of application are brought to mind by asking why and when we are interested in opinions.

Often, we want to make use of the opinion of a person whom we regard as an expert. Does the weatherman think that it will rain, the doctor that we shall soon get well, the lawyer that it would be better to settle out of court, or the geologist that there might be lots of oil at the bottom of a deep hole? Most of the following subsections are concerned explicitly with the utilization of experts.

And often, we want to know a person's opinions in order to judge how well informed he is. Every academic examination can be viewed in that light. It is, therefore, interesting to explore the possible usefulness of proper scoring rules in academic examinations, and the final subsection returns to academic testing. This domain of application shades into that of trying to determine which among possible experts are most valuable for a given task by means of the relation between their past opinions and reality.

Since public-opinion polls are ostensibly concerned with finding out the opinions of the public, we might at first expect proper scoring rules to have important applications there, in harmony with the remarks of Haavelmo quoted in Section 1. This has, however, apparently never been suggested, partly perhaps because public-opinion polls are seldom so much concerned with the opinions of the public as with the preferences of the public. But investigators do occasionally want to know how firmly the public believes some as yet unresolved matter of fact, and in a few of these cases, proper scoring rules might have some role. Literally paying the participants in a poll on the basis of the accuracy of their predictions would of course ordinarily be infeasible, especially in the common case of casual, one-time participants. Yet scoring rules might be used in training panelists to assess their own probabilities, perhaps partly in paid practice sessions about immediately verifiable predictions. This would seem less farfetched should scoring rules come to be widely applied in the schools.

One domain of potential application of proper scoring rules is suggested not so much by the problem of obtaining the opinions of others as by the difficulty of obtaining our own. For it is by no means easy to elicit your own probabilities. Vagueness is a major obstacle, and your first reactions are often greatly modified when you reflect upon their implications. Those who have experimented on themselves and on others generally feel that frequent practice with proper scoring rules and with other probability elicitors helps a person to combat vagueness and to arrive more promptly and accurately at his personal probability. This proposition might be difficult to investigate experimentally and even seems difficult to state with precision, but its promise of benefit is great. (See, for example, [12, 59].)

10.3 Yes, No, or Maybe Is Not Enough

Traditionally, experts, except for turf experts, have not communicated their opinions in probabilities. The doctor says, "That child will soon be well and then he had better have his tonsils out." The geologist says, "That looks like a good place to drill." And, until recently, most of us were content when the weatherman said simply, "Rain tonight."

The importance of a system of communication in which experts express themselves in terms of genuine personal probabilities and in which those who utilize the opinions of experts—that is, all of us—are trained in understanding and using such probabilities was energetically underlined by Grayson [27]. In recent years, meteorologists have been announcing forecasts in terms of probability on some radio and television services, but not all who offer these probability forecasts think in terms of personal probability, and some seem to be very vague indeed about what they mean by probability in a forecast. The earliest known reference to proper scoring rules is by the meteorological statistician, Brier [5], and much of the current literature on proper scoring rules is inspired by meteorology, as in [21, 39, 40, 41, 46, 53, 58, 59, 60] and works cited in them.

Applying a proper scoring rule to obtain better opinions from an expert might mean, at one extreme, merely using the rule to keep score as a training device to give the expert a mild incentive to understand what probability means and to ask himself whether it is really his personal probabilities that he is reporting. At the opposite extreme, the scoring rule might be implemented by substantial cash payments. For example, an oil geologist willing to buy a \$100 interest in a million dollar drilling investment is saying very clearly that to him, the expected revenue of the well is greater than the expected cost. More accurately, that is what he is saying if his expression of opinion is not affected by nonlinearities in his utility, possibly reflecting the nonlinearity of income after taxes and possibly associated with trying to explain to his wife how he lost $100 gambling or with a very small probability of an enormous revenue.

This same example, which suggests so vividly how any proper scoring rule might be applied to elicit from a geologist his opinions about the yield of a well that is to be drilled also serves to illustrate what seems to be an insurmountable obstacle to the application of arbitrary scoring rules, and even of any proper scoring rules to all, to many situations in which opinion is sought as a basis for decision. If whether the well is to be drilled does not depend on this geologist's opinion, then any proper scoring rule concerning its expected yield can be implemented, in particular by offering to sell shares in the well at various prices. But if whether to drill the well depends in part on the opinion of the geologist, then some of the events about which his opinion is wanted may never be tested precisely because of his advice, so the scoring-rule contract cannot be offered literally. The phenomenon is ubiquitous. We can never know what would have happened had surgery been ventured, had a certain product been marketed, or had a certain student been admitted. Business sharing, to be discussed in Subsection 10.12, seems to offer some possibility of circumventing this widespread obstacle to the literal application of proper scoring rules.

10.4 Big Money

The interrogator has an interest in making the possible fluctuations in the wealth of the expert large. For this motivates the expert to reflect hard and well before answering and yet need not add systematically to his expected fee. The larger the fluctuations are, however, the more the expert is motivated to report not his real probabilities but numbers that reflect in part the nonlinearity of his utility. The theory of these distortions has been somewhat explored by Winkler and Murphy [60], and some possible ways to avoid them or compensate for them were reported on in Section 3. My own hope and expectation is that the skillful use of small, or even purely symbolic, scoring-rule payments (in addition to the usual compensations) will enable experts to know and to communicate their opinions much more accurately than has been usual.

The only practical experience with elicitation involving substantial cash prizes and losses thus far apparently consists of observations on gambling behavior that Ward Edwards has made, and expects soon to publish. The use of substantial, or even of token, scoring-rule payments in business would often be a considerable break with tradition likely to involve legal and other administrative problems.

10.5 Tiny Probabilities and the Expert

Probabilities corresponding to odds of one in a million or even less, such as the probabilities of specific disasters are sometimes important [32]. Any scheme to give an expert a serious cash incentive to reveal his personal probabilities for very improbable events would seem to court insuperable difficulties with the nonlinearity of utility. And, even if we do not despair of obtaining sincere opinions about such matters by engaging sincerely inclined experts in the right kind of make-believe, the difficulty of such make-believe, and the special training required for it, are in danger of being greatly underrated.

Magnification is sometimes possible. For example, though it would not be practical to engage me in a meaningful bet that the birth awaited by the recently married Smiths will be quadruplets, my expectation for the number of sets of quadruplets among the next million American births (about two, based on a little reading) could be elicited and might be of some use. But this relief is largely illusory. For a real expert on multiple births would be expected to take into account such data as that Mrs. Smith is herself a twin and a very young bride, so there is no practical possibility of counting the quadruplet births in a large number of cases similar in the respects considered pertinent.

10.6 Employing Expert Opinion

In just what ways can you expect to profit from the opinions of experts in serious matters? According to a very broad model, you have an important decision to make in the light of all sorts of data at your disposal, and this data may include the behavior of experts. You could, in principle, explore all sorts of ways of interrogating an expert—not confining yourself to eliciting his opinions—and study empirically how his responses can profit your business. Conceivably, pain in the weatherman's great toe would better help you plan picnics than would his opinion about the weather. Yet I presume that ordinarily little of importance would be lost if you could obtain only the opinion of the expert, that is, his personal probabilities. What should be done with such an opinion? The simplest thing, and sometimes the appropriate thing, would be to make the expert's opinion your own.

This is by no means mandatory. For example, an expert who always ascribes very small probability to what actually occurs and to nothing else would be as useful as one who is omniscient, but you would of course not make the opinions of this perfect fool your own. Again, you might discover with experience that your expert is optimistic or pessimistic in some respect and therefore temper his judgments. Should he suspect you of this, however, you and he may well be on the escalator to perdition.

10.7 Divergent Opinion

We often have access to more than one expert, and what to do when doctors disagree has always been, and will always be, a quandary. One important thing to do, but far outside the scope of this paper, is to encourage the right kind of communication between the experts. Exploration of how to do this is, for example, one of the aims of the Delphi technique [3; 8; 29, Part

II]. In general, good communication is what makes the experts share factual information and help each other think their opinions through thoroughly, and bad communication is what encourages various vices such as exaggeration and excessive deference. Sooner or later, despite all techniques of communication, divergent expert opinions will have to be faced. Perhaps you will make some composite of one or more expert opinions and your own opinion. An extreme way to do that would be to decide, on the basis of past experience or otherwise, that a particular one of the experts is the only one worth listening to and to make his opinion your own. A more general procedure would be to average the opinions, that is, to average the probability distributions associated with the experts (possibly including yourself), giving each the weight you think appropriate Thus, rather than simply choose one expert among several, you can choose among the infinite number of synthetic experts that constitute the convex closure of the several. (A few key references bearing on the subject of this paragraph are [10; 11; 47, Ch. 10; 52 p. 65ff; 56].)

10.8 The Expert as an Instrument

When is one expert, real or synthetic, to be preferred to another? An "expert" in this context is a mechanism, possibly with human components, generating numbers that you contemplate using instead of your own personal probabilities (or, more generally, expectations) in certain contexts One crude, practical answer sometimes available and appropriate, is this: Employ, until you have further experience, that expert whose past opinions, applied to your affairs, would have yielded you the largest average income.

No rule of this sort can claim absolute or objective validity, and this one has been couched especially roughly for the sake of simplicity. For example, actual past experience with the experts may be extensive, moderate, meager, or absolutely lacking. When past experience is extensive, but not too extensive, the rule often has much to recommend it; when direct past experience with the experts is meager, the rule is silly; and when such experience is altogether lacking, the rule results in a tie and is therefore empty. Actually, if you have little or no past experience with the experts, you will have to ponder them in terms of whatever information it was that brought you to regard them as promising in the first place: this well finder is regarded by the whole neighborhood as infallible with the hazel fork; that one is a professor of geology and the author of an important treatise on subsurface hydrology but has never before tried to help anyone locate a well. In such a context, the subjective aspect of your decision is thrown into prominence, but no matter how much direct past experience you may have with the experts, the ultimate subjectivity of your choice among them never disappears, though its effects may become less agonizing—according to the personalistic Bayesian theory of statistics, as in [48].

When your past experience with the experts is very extensive indeed, it may become profitable for you to refine the original rough rule by dividing up circumstances into categories and confiding in different experts for different categories of decisions. Such discussion could continue indefinitely; for the situations are innumerable and tend to parallel the whole field of decision-theoretic statistics.

An interesting process for coordinating the efforts of experts in different fields known as PIP (for probability information processing) has been vigorously pursued [20].

10.9 Slippery Utilities

A different kind of complication in applying, or adapting, the rough rule is that the notion of average income may not be readily applicable to your affairs. For one thing, it may be important to measure your own income in utility rather than in cash. If this involves only determining your utility for cash, you may be able to do that reasonably well with moderate effort. If, however, the consequences of your act are not easily converted into cash but involve values difficult to weigh against each other such as beauty, justice, and health, your dilemma may be especially severe. Raiffa [44] has recently published a book largely on these subtle problems of pondering the imponderable and evaluating the invaluable that reviews, and contributes to, a considerable literature.

Another difficulty in measuring utility is seen in this example: You are the person responsible for choosing which of several televised meteorological forecasters shall serve your city. There is, I assume for simplicity, abundant evidence of past performance, and the members of the community who use the forecasts will behave in accordance with the probabilities announced in them. Since yours is a public trust, you would like to choose the forecaster that in the past would have maximized the mean income of members of the community.

Subtle welfare-economic decisions about the relative importance of bent-pin anglers and barn painters could complicate your problem, but, even more important, relatively little is really known about the uses to which public weather information is put and what its economic consequences are. Thought has been given to the problem of the economic value of meteorological forecasts, both for the general public and for special purposes, but difficult, important empirical aspects of the question remain to be explored. (Key references are [36, 40].)

10.10 Which Scoring Rule for the Trained Respondent?

There are as many proper scoring rules for a trichotomy, for example, as there are convex functions over the baricentric triangle, or two dimensional simplex. It would therefore seem important to study in what respects one scoring rule is better than another. But this question has thus far proved surprisingly unproductive. Its elusiveness is brought out by the consideration than an ideal subject responds to all proper scoring rules, including those involving extremely small payments, in exactly the same way. Therefore, any criteria for distinguishing among scoring rules must arise out of departures of actual subjects from the ideal.

Since we all do depart markedly from the ideal, it might seem that one proper scoring rule would be much more effective with a real person than another, and this presumably is often so. But if a person is reasonably sophisticated, though far from ideal, the form of a proper scoring rule for eliciting his probabilities, for, say, a trichotomy should—provided its amplitude is sufficient to command attention—have little or no effect on his response. To see this, put yourself in his place. You are offered a contract that will result in certain cash payments to you depending on your choice of three numbers p, q, and r and on whether a certain game ends in win, lose, or tie. If you know what personal probabilities are and understand that the contract is so drawn that it is to your interest to report your personal probabilities, then the details of the contract seem unimportant; for no matter which proper scoring rule it corresponds to, you should ask yourself what your personal probabilities for the three events are and report them.

Yet, the terms of the contract might make a modest difference to you. Suppose, for example, that very little money is to change hands in case of a tie, no matter what your response is. In this case, you have little incentive to ask yourself carefully the probability of a tie and are thus left free to focus on the relative probability of a win given that there is not a tie. In this case, your questioner will be well served if he is mainly interested in that conditional probability, and he will be badly served if he particularly wants to know the probability that you attach to a tie.

Thus, at least a vague criterion applicable even to sophisticated but human respondents emerges. Insofar as responding is hard work, the scoring rule should encourage the respondent to work hardest at what the questioner most wants to know. If this is to be effective, it must not merely be mathematically true but also plain to the respondent that he will be rewarded most for working on the right aspects of his opinion. This is in part a psychological question of human communication, subject to much speculation and experimentation. To illustrate, you are faintly curious to know whether the respondent thinks that a tie is likely and desperate to know whether he believes that the home team will win. You might be best served by entering into two palpably separate contracts with the respondent, a small one hinging only on whether there is a tie and resulting in a rough casual elicitation of the respondent's r, and a larger one, involving no payment in case of a tie, resulting in a well considered evaluation of the ratio $p:q$ for the respondent. Of course the two contracts together amount to a single scoring rule, though presenting them separately might work better psychologically.

The appropriate incentive for you to offer a respondent for his opinion depends not only on the importance for you of obtaining that opinion with a specified accuracy but also on the difficulty for the respondent in obtaining it from himself. This makes the choice of a scoring rule designed to evoke the right degree of effort from the respondent on the various components of his task particularly subtle.

10.11 There are no Bargain Scoring Rules

Since a scoring rule is a scheme of payments, it might seem natural to choose the rule that promises to obtain the required information as cheaply as possible or perhaps the one that obtains the most for the money. There may be something to this line of thinking, but not in first approximation, and I have been unable to make progress with it.

The respondent will presumably work for what he regards as an acceptable wage. In the presence of a scoring rule, he will perceive his wage as random with an expectation that can be adjusted by adding a constant to the scoring rule. So, within the linearity approximation, the respondent does not charge extra for submitting to the scoring rule, and it therefore seems roughly reasonable to reckon that the questioner is not charged for it.

The higher the amplitude of the scoring rule, the more incentive it gives the respondent to reply with care. On this account, the respondent might in principle come to insist on a higher mean wage for facing a highly fluctuating scoring rule, and this could tend to deter the questioner from using high amplitudes. But the important practical limitation on amplitude would seem to be the need to avoid the distortion of response induced by the nonlinearity of the respondent's utility.

10.12 Business Sharing

In common sense, we feel without any overt reference to economic models that some responses are not so wrong as others and ought not to be so heavily penalized. If a respondent is pretty sure that the home team will win, and there is in fact a tie, then he is perhaps not so wrong as if it had lost. Scoring rules reflecting this idea have been sought and easily found. See, for example, [21, 53]. One interesting way to adjust the rewards and penalties of the respondent to the interests of the interrogator, which was brought out in a dramatic and more radical form in McCarthy's [37] pioneering note on proper scoring rules, is to give the respondent a fractional interest in the business involved. To illustrate with an overidealized example, an oil prospecting company could give its geologist a small fraction of all profits and losses with the understanding that all decisions in the business would be made using the geologist's personal probabilities about geological uncertainties.

(It would be interesting to consider with some care the respects in which such an example is realistic and unrealistic, but I can only go a step or two in that direction here. Stock in the company would seem to give the geologist an interest in reporting his probabilities honestly if he could be assured that they would be adopted as the personal probabilities of the management for the events concerned. But this incentive may not be fully in harmony with the expert's incentives to appear worthy, as opposed to simply being worthy, of retention and promotion—a complication that affects not only business sharing but all applications of proper scoring rules to a professional expert. Business sharing does not present the expert with an explicit scoring rule in any business complicated enough to provide a more than mechanical role for the managers, in particular in any business in which there are other uncertainties than those about which the expert is consulted, but an implicit rule is as effective in principle as an explicit one.)

Long ago, Gauss [25, Sec. 6] proposed that economic losses (such as those in a game of chance) provide a good model for the incentive to estimate accurately even in the most academic contexts. Decisively to uphold or to overthrow this suggestion does not seem possible. Personally, it appeals to me. Correspondingly, when we say that a tie should not be regarded as so distant from a win as a loss would be, I am inclined to think that that is because we have in mind various uses for sport forecasts in which the penality is less for one kind of error than another. Of course, the penalty need not be a monetary one; it might involve, for example, loss of social prestige. Fortunately, the elusive question of whether all that is good and bad about a forecast can ultimately be referred to profit and loss in economic decisions, sufficiently, widely interpreted, need not be resolved in order to show the interest and utility of viewing proper scoring rules as a share in a real or a fictional business. For the business model is certainly a mathematically general model for all proper scoring rules and a fertile point of view for the generation of proper scoring rules that penalize some errors more than others.

The technical point that every proper scoring rule can be viewed as a share in a business and that every such share leads to an at least weakly proper scoring rule should be appreciated. Section 7 makes these points clear. Every strictly proper scoring rule amounts to the possibility of choosing among acts, only one of which is appropriate to each system of personal probabilities. Conversely, a person knowing that an act in a specified economic situation is to be chosen for him in accordance with his announced system of personal probabilities will have no incentive to announce a false one. However, if the convex function arising from the family of acts has flat places—technically, is not strictly convex—then the person has no positive incentive to distinguish among certain systems of probabilities, and the scoring rule is only weakly proper.

Insofar as business sharing is a practical method of elicitation, it makes possible the use of a proper scoring rule even in those situations stressed at the end of Subsection 10.3 in which arbitrary proper scoring rules cannot be implemented because some of the conditional probabilities to be elicted will not be tested, depending on the opinions expressed by the expert.

The parallelism between implementing a scoring rule by business sharing and the rough rule for rating experts according to how their advice would have affected your business in the past (discussed in Subsection 10.8) is evident, but the two things must not be confused. In particular, the rating rule can be used regardless of what scoring rule if any is used to elicit the opinions.

10.13 Some Armchair Psychology

Consider now a subject quite untrained in personal probability and the theory of scoring rules whose only incentives are provided by the scoring itself. It is questionable whether, in any serious application, this ought to be allowed to happen. Though any strictly proper scoring rule is a sufficient guide for an ideally intelligent Robinson Crusoe no matter how uninstructed, can we expect real people to respond well with no other coaching than is provided by the scoring rule itself or even by extended experience with the scoring rule? And even if investigation should yield a somewhat affirmative reply, is there any point in withholding instruction? Whatever the answers, it does seem stimulating to speculate on what kinds of scoring rules presented in what way would most nearly operate on naive subjects as all strictly proper scoring rules are supposed to operate on sophisticated ones, and this should lead to ideas of practical value for subjects of intermediate sophistication.

In the first place, the subject must understand the scoring rule. If it makes explicit reference to logarithms, or even to squares, most ordinary people will not understand it at all; and even those with mathematical training may not be nearly apt enough at calculation to use the rule effectively. This is an important reason to present the rule through some vivid tabular or graphic device, which could, for example, take the elaborate form of conversational-mode digital computation, or more simply of some slide-rule device such as those of the Shuford-Massengill Corporation (Lexington, Mass.), or perhaps tabulation of the scoring rule itself, possibly very boldly rounded [14].

Perhaps it is helpful to a subject responding about a partition of possible events if the economic consequence of his response is a function only of the element of the partition that happens to obtain. This condition imposes no constraint at all for two-fold partitions, but for n-fold partitions with $n>2$, it leads to the logarithmic scoring rules of Subsection 9.4. The possible advantages of the simplicity might often be outweighed by the inappropriateness of a symmetric scoring rule in asymmetric situations or of a scoring rule that lays emphasis on the correct elicitation of small probabilities. I have sometimes heard the possibility that a subject responding to a logarithmic scoring rule could be subjected to an infinite (or at any rate, unlimited) penalty raised as an overwhelming objection. This possibility does of course imply that the method cannot be applied literally, but approximate applications, in which the subject is not allowed to name probabilities less than, say, 10^{-3} suggest themselves. And, as mentioned in Subsection 10.5, obtaining probabilities very close to 0 by means of direct incentives does not seem practical by any scoring rule.

10.14 Proper Scoring Rules in School

Proper scoring rules hold forth promise as more sophisticated ways of administering multiple-choice tests in certain educational situations [14, 49]. The student is invited not merely to choose one item (or possibly none) but to show in some way how his opinion is distributed over the items, subject to a proper scoring rule or a rough facsimile thereof.

Though requiring more student time per item, these methods should result in more discrimination per item than ordinary multiple-choice tests, with a possible net gain. Also they seem to open a wealth of opportunities for the educational experimenter.

Above all, the educational advantage of training people—possibly beginning in early childhood—to assay the strengths of their own opinions and to meet risk with judgment seems inestimable. The usual tests and the language habits of our culture tend to promote confusion between certainty and belief. They encourage both the vice of acting and speaking as though we were certain when we are only fairly sure and that of acting and speaking as though the opinions we do have were worthless when they are not very strong.

Effects of nonlinearity in educational testing deserve some thought, but presumably nonlinearity is not a severe threat when a test consists of a large number of items. One source of nonlinearity that has been pointed out to me is this. A student competing with others for a single prize is motivated to respond so as to maximize the probability that his score will be the highest of all. This need not be consistent with maximizing his expected score, and presumably situations could be devised in which the difference would be important.

REFERENCES

[1] Aczel, J., *Lectures on Functional Equations and Their Applications*, New York: Academic Press, 1966.

[2] Allais, Maurice, "Le Comportement de l'Homme Rationnel Devant le Risque: Critique des Postulats et Axioms de l'École Americaine," *Econometrica*, 21 (October 1953), 503–546.

[3] Ament, Robert H., "Comparison of Delphi Forecasting Studies in 1964 and 1969," *Futures*, 2 (March 1970), 35–44.

[4] Bayes, Thomas, "Essay Towards Solving a Problem in the Doctrine of Chances," *The Philosophical Transactions*, 53 (1763), 370–418. Reprinted in W. Edwards Deming, ed., *Facsimiles of Two Papers by Bayes*, Washington, D.C.: The Graduate School, The Department of Agriculture, 1940, and in "Thomas Bayes's Essay towards Solving a Problem in the Doctrine of Chances," *Biometrika*, 45 (December 1958), 293–315.

[5] Brier, G. W., "Verification of Forecasts Expressed in Terms of Probability," *Monthly Weather Review*, 78 (January 1950), 1–3.

[6] Brown, T. A., "Probabilistic Forecasts and Reproducing Scoring Systems," Santa Monica, Calif.: RAND Corporation, RM-6299-ARPA, June 1970, 65 pp.

[7] Buehler, Robert J., "Measuring Information and Uncertainty," in *Proceedings of the Symposium on Foundations of Statistical Inference*, Toronto: Holt, Rinehart, and Winston, 1971 (in press).

[8] Dalkey, N. C., "An Experimental Study of Group Opinion: The Delphi Method," *Futures*, 1, 5 (September 1969), 408–26.

[9] de Finetti, Bruno, "La prévision: Ses Lois Logiques, Ses Sources Subjectives," *Annales de l'Institut Henri Poincaré*, 7 (1937), 1–68. English translation in (Kyburg and Smokler, 1964).

[10] ———, "La Notion de Distribution d'Opinion comme Base d'un Essai d'Interpretation de la Statistique," *Publications de l'Institut de Statistique de l'Université de Paris*, 1, 2 (1952), 1–19.

[11] ———, "Media di Decisioni e Media di Opinioni," *Bulletin de l'Institut International de Statisque*, 28th session, 34, 2 (1954), 144–57.

[12] ———, "Does It Make Sense to Speak of 'Good Probability Appraiser'?" in *The Scientist Speculates: An Anthology of Partly-Baked Ideas*, New York: Basic Books, 1962, 357–64.

[13] ———, "Probabilità Composte e Teoria Delle Decision," *Rendiconti de Matematica*, 23 (1964), 128–34.

[14] ———, "Methods for Discriminating Levels of Partial Knowledge Concerning a Test Item," *The British Journal of Mathematical and Statistical Psychology*, 18 (May 1965), 87–123.

[15] ———, "Quelques Conventions Qui Semble d'Être Utile," *Revue Roumaine de Mathematique Pures et Appliquées*, 12, 9 (1967), 1227–33.

[16] ———, "Probability: Interpretations," Vol. 12, 496–505, in the *International Encyclopedia of the Social Sciences*, New York: The Macmillan Company, 1968.

[17] ———, "Logical Foundations and Measurement of Subjective Probability," *Acta Psychologica*, 34 (December 1970), 129–45.

[18] ———, and Leonard J. Savage, "Sul Modo di Scegliere le Probabilità Iniziali," *Sui fondamenti della statistica, Biblioteca del Metron*, Series C, 1 (1962), 81–154 (English summary, 148–151).

[19] Edwards, Ward, "A Bibliography of Research on Behavioral Decision Processes to 1968," *Memorandum Report No. 7*, Human Performance Center, University of Michigan, January 1969, 93.

[20] ———, L. D. Phillips, W. L. Hays, and B. C. Goodman, "Probability Information Processing Systems: Design and Evaluation," *IEEE Transactions on Systems Science and Cybernetics*, SSC-4 (1968), 248–65.

[21] Epstein, Edward S., "A Scoring System for Probability Forecasts of Ranked Categories," *Journal of Applied Meteorology*, 8 (December 1969), 985–7.

[22] Fishburn, Peter C., *Utility Theory for Decision Making*, New York: John Wiley and Sons, Inc., 1970.

[23] Friedman, Milton, "Choice, Chance, and Personal Distribution of Income," *Journal of Political Economy*, 61, 4 (August 1953), 277–90.

[24] ———, and Leonard J. Savage, "The Utility Analysis of Choices Involving Risk," *Journal of Political Economy*, 56 (August 1948), 279–304. Reprinted in *Readings in Price Theory*, Chicago: Richard D. Irwin, 1952, and in *Landmarks in Political Economy*, edited by Earl J. Hamilton, *et al.*, The University of Chicago Press, 1962.

[25] Gauss, Carl Friedrich, "Theoria Combinationis Observationum Erroribus Minimis Obnoxiae," *Commentationes Societatis Regiae Scientarum Gottingensis Recentiores*, 5 (1821), 33–90. In German translation, A. Börsch and P. Simon, *Abhandlungen zur Methode der kleinsten Quadrate*, Berlin: Westdruckerei Joachim Frickert, 1887 (reprinted Wurzburg: Physica-Verlag, 1964). In French translation, J.

Bertrand, *Methode des Moindres Carrés*, Paris: Mallet-Bachelier, 1855.
[26] Good, I. J., "Rational Decisions," *Journal of the Royal Statistical Society, Ser. B*, 14 (1952), 107–14.
[27] Grayson, Charles Jackson, *Decisions under Uncertainty: Drilling Decisions by Oil and Gas Operators*, Boston: Harvard University, Division of Research, Graduate School of Business Administration, 1960.
[28] Haavelmo, Trygve, "The Role of the Econometrician in the Advancement of Economic Theory," *Econometrica*, 26 (July 1958), 351–7 (Presidential address, Econometric Society, Philadelphia, Dec. 29, 1957).
[29] Helmer, Olaf, *Social Technology*, New York: Basic Books, 1966.
[30] Hendrickson, Arlo D., "Payoffs to Probability Forecasters," *Technical Report No. 145*, University of Minnesota, September 1970.
[31] Kyburg, Henry E., Jr., and Smokler, Howard E., *Studies in Subjective Probability*, New York: John Wiley and Sons, 1964.
[32] Maloney, Clifford J., "A Probability Approach to Catastrophic Threat," *ARO-D Report 70-2*, U. S. Army Research Office—Durham, Durham, North Carolina, 1970?.
[33] Marschak, Jacob, "Management Research and Behavioral Science," *Working Paper No. 2*, Western Management Science Institute, University of California, Los Angeles, 1961.
[34] ———, "Actual Versus Consistent Decision Behavior," *Behavioral Science*, 9 (April 1964), 103–10.
[35] ———, "Decision Making: Economic Aspects," Vol. 4, 42–5, in the *International Encyclopedia of the Social Sciences*, New York: the Macmillan Company, 1968.
[36] Maunder, W. J., *The Value of the Weather*, London: Methuen, 1970.
[37] McCarthy, John, "Measures of the Value of Information," *Proceedings of the National Academy of Sciences*, (1956), 654–5
[38] Mosteller, Frederick, and Wallace, David L., *Inference and Disputed Authorship: The Federalist*, Reading, Mass.: Addison-Wesley, 1964.
[39] Murphy, Allan H., "The Ranked Probability Score and the Probability Score: A Comparison," *Monthly Weather Review*, 98 (December 1970), 917–24.
[40] ———, and Allen, R. A., "Probabilistic Prediction in Meteorology: A Bibliography," Silver Spring, Md., Environmental Science Services Administration, Technical Memorandum No. WBTM TDL 35, 1970.
[41] ———, and Epstein, Edward S., "Verification of Probabilistic Predictions: A Brief Review," *Journal of Applied Meteorology*, 6 (October 1967), 748–55.
[42] ———, and Winkler, Robert L., "Scoring Rules in Probability Assessment and Evaluation," *Acta Psychologica*, 34 (December 1970), 273–86.
[43] Pratt, John W., "Risk Aversion in the Small and in the Large," *Econometrica*, 32, 1–2 (January–April 1964) 122–36.
[44] Raiffa, Howard, *Decision Analysis: Introductory Lectures on Choices under Uncertainty*. Reading, Massachusetts: Addison-Wesley, 1968.
[45] Roberts, Harry V., "Risk, Ambiguity, and the Savage Axioms: Comment," *Quarterly Journal of Economics*, 77 (1963) 327–42.
[46] ———, "On the Meaning of the Probability of Rain," in *Proceedings of the First National Conference on Statistical Meteorology*, Boston: American Meteorological Society (1968) 133–41.
[47] Savage, Leonard J., *The Foundations of Statistics*. New York: John Wiley & Sons, Inc; and London: Chapman and Hall, 1954.
[48] ———, "The Foundations of Statistics Reconsidered," in *Proceedings of the Fourth Berkeley Symposium*, Berkeley, University of California Press (1961) 575–86. Reprinted in (Kyburg and Smokler, 1964).
[49] Shuford, Emir H., Jr., Albert Arthur, and H. Edward Massengill, "Admissible Probability Measurement Procedures," *Psychometrika*, 31 (June 1966), 125–45.
[50] Smith, Cedric A. B., "Consistency in Statistical Inference and Decision," *Journal of the Royal Statistical Society*, Ser. B, 23 (1961), 1–25.
[51] ———, "Personal Probability and Statistical Analysis," *Journal of the Royal Statistical Society*, Ser. A, 128 (1965), 469–89.
[52] Staël von Holstein, Carl-Axel S., *Assessment and Evaluation of Subjective Probability Distributions*, Stockholm: The Economic Research Institute at the Stockholm School of Economics, 1970.
[53] ———, "A Family of Strictly Proper Scoring Rules Which Are Sensitive to Distance," *Journal of Applied Meteorology*, 9 (1970), 360–4.
[54] Toda, Masanao, "Measurement of Subjective Probability Distribution," *Report No. 3*, Division of Mathematical Psychology, Institute for Research, State College, Pennsylvania, 1963.
[55] Winkler, Robert L., "The Assessment of Prior Distributions in Bayesian Analysis," *Journal of the American Statistical Association*, 62 (September 1967), 776–800.
[56] ———, "The Consensus of Subjective Probability Distributions," *Management Science*, 15, 2 (October 1968), B-61-B-75.
[57] ———, "Scoring Rules and the Evaluation of Probability Assessors," *Journal of the American Statistical Association*, 64 (September 1969), 1073–78.
[58] ———, and Allan H. Murphy, "Evaluation of Subjective Precipitation Probability Forecasts," in *Proceedings of the First National Conference on Statistical Meteorology*, Boston: American Meteorological Society, 1968, 148–57.
[59] ———, and Allan H. Murphy, " 'Good' Probability Assessors," *Journal of Applied Meteorology*, 7 (October 1968), 751–8.
[60] ———, and Allan H. Murphy, "Nonlinear Utility and the Probability Score," *Journal of Applied Meteorology*, 9 (February 1970), 143–8.