A MODEL FOR THE EXPERIMENTAL MEASUREMENT OF THE UTILITY OF GAMBLING

by Halsey L. Royden, Patrick Suppes and Karol Walsh
Stanford University

Reprinted from BEHAVIORAL SCIENCE
Vol. 4, No. 1, January, 1959
Printed in U.S.A.

Reprinted from BEHAVIORAL SCIENCE
Vol. 4, No. 1, January, 1959
Printed in U.S.A.

It seems obvious that a gambler is motivated to take risks by expectation of gain. In the development of a theory of gambling behavior the need has been recognized of determining the gambler's "utility function" with respect to money, that is, the relative worth to him of various amounts of money. But more than money may be involved in the gambler's expected gain. Gambling itself may have a "utility" for him. Here a theory of gambling decisions takes into account both utilities.

A MODEL FOR THE EXPERIMENTAL MEASUREMENT OF THE UTILITY OF GAMBLING[1]

by Halsey L. Royden, Patrick Suppes and Karol Walsh

Stanford University

THE MODEL

THE experiment described in this paper was designed to measure utility of gambling to an individual and to predict, on the basis of this measured utility, the individual's future choices. The vexing problem of the existence of a specific utility of gambling has been much discussed in the literature of decision making. Von Neumann and Morgenstern (5, p. 28) have this to say: "We have practically defined numerical utility as being that thing for which the calculus of mathematical expectations is legitimate. Since [our axioms] secure that that necessary construction can be carried out, concepts like a 'specific utility of gambling' cannot be formulated free of contradiction on this level." And they add in a footnote to this last sentence: "This may seem to be a paradoxical assertion. But anybody who has seriously tried to axiomatize that elusive concept, will probably concur in it." As far as we know the work reported in this paper constitutes the first systematic attempt to measure experimentally the utility of gambling. The model at the basis of our experiment does not yield a complete axiomatization of this "elusive" concept, but it is a mathematically definite model for monetary outcomes.[2]

The model can best be explained by presenting a detailed illustration of how it is

[1] This research was supported by Group Psychology Branch, Office of Naval Research.

[2] In 1906 Irving Fisher (*Income and Capital*) suggested that a person's choice between alternative investments depends not only on the mean return $\left(\frac{x+y}{2}\right.$ in the cases studied in this present article$\left.\right)$ but also on other properties, such as dispersion, of the probability distribution of returns. It has become customary among economists to describe as a "risk-lover" a person who, of two investments with the same mean return, prefers the one with higher dispersion. The experiments described in the article throw light on the occurrence and extent of such behavior. As stated later in the article, such behavior, and the consequent existence of a "utility of gambling" in the authors' sense, does not, in fact, contradict the Neumann-Morgenstern hypothesis that a person is characterized by a "utility function" whose expected value he maximizes. A "specific utility of gambling . . . formulated free of contradiction on this level" (to use the above quotation from Neumann-Morgenstern) would have to be defined differently. For example, the Neumann-Morgenstern hypothesis is, in fact, refuted by a person who plays Russian roulette or in some other way shows that he prefers a high to a low probability of an undesirable outcome. In economic literature, such behavior is sometimes called "love of danger," and distinguished from "love of risk."

tested in the experiment. Offers in the form of a game matrix

A	B
x	z
y	z

are given S, where x and y are different small amounts of money, and z is some other small amount of money approximately half-way between x and y. S is asked to select either option A or option B, and is told that if he chooses option A, he will be allowed to throw a fair die to decide which value he will actually win; if he chooses option B, his win is certain, the amount z. The amount z is varied during the experiment until S's point of indifference between the "sure-thing option" (z, z) and the "gamble" (x, y) is reached, that is, until the value z' is found such that S, when presented at random the offer

A	B
x	z'
y	z'

shifts back and forth in his choice of A or B, and no longer consistently picks one or the other. The value z' is called S's indifference point. This indifference point, z', minus the mean of the "true" option (x, y) is the utility of gambling value associated with the pair of values x and y. In symbols,

$$z' - \frac{x+y}{2} = \varphi'(x, y)$$

where φ is the utility of gambling function.

Since each option (x, y) is uniquely determined by its mean $m = \frac{x+y}{2}$ and its absolute difference $d = |x - y|$, we replace φ' by the function φ defined for the means and differences, that is,

$$\varphi\left(|x - y|, \frac{x+y}{2}\right) = \varphi'(x, y).$$

In view of the fact that the fair die used is tested to insure that it has subjective probability ½, the difference $|x - y|$ is simply twice the standard deviation of the offer (x, y) for

$$\sigma^2(x, y) = \left(x - \frac{x+y}{2}\right)^2 \frac{1}{2} + \left(y - \frac{x+y}{2}\right)^2 \frac{1}{2},$$

whence

$$\sigma(x, y) = \tfrac{1}{2}\,|x - y|\,.$$

The experimental problem is then, on the basis of limited observations, to construct each S's utility of gambling surface. As would be expected, our procedure is to draw a contour map of φ-values, where means are plotted on the abscissa and differences on the ordinate.

The standard hypothesis of rational behavior that individuals maximize expected utility is now replaced, for the special case of monetary outcomes, by the hypothesis that they maximize the *sum* of the expected monetary value and the utility of gambling of the option. If $\succeq$ represents weak preference among options, we may symbolize this hypothesis by:

H. *For each S, there is a utility of gambling function φ such that for all small amounts of money x, y, u and v if $x \neq y$ and $u \neq v$ then*

$$(x, y) \succeq (u, v) \quad \text{if and only if} \quad \frac{x+y}{2} + \varphi'(x, y) \geq \frac{u+v}{2} + \varphi'(u, v).$$

There are two remarks of some theoretical interest on the relationship between Hypothesis H and the maximization of expected utility hypothesis (MEU). The first is to note the somewhat surprising fact that MEU implies H. For the observation of this fact, we are much indebted to Professor Marschak. His argument runs as follows. Let the indifference point z' of any option (x, y) be designated by i_{xy}, and let $\approx$ be indifference between options. Then

(1) $$i_{xy} \approx (x, y),$$

and thus assuming that always more money is preferred to less:

(2) $$(x, y) \succeq (u, v) \quad \text{if and only if} \quad i_{xy} \geq i_{uv}\,.$$

Given the utility function u on the basis of MEU, then from (1), we have

$$u(i_{xy}) = \frac{u(x) + u(y)}{2}$$

and thus

$$i_{xy} = u^{-1}\left(\frac{u(x) + u(y)}{2}\right).$$

Consequently from our earlier definition of φ', we have

$$\varphi'(x, y) = i_{xy} - \frac{x + y}{2}$$

(3)

$$= u^{-1}\left(\frac{u(x) + u(y)}{2}\right) - \frac{x + y}{2},$$

and also from (3) we have

$$i_{xy} = \varphi'(x, y) + \frac{x + y}{2}, \quad (4)$$

but substituting (4) into (2) yields exactly H, with φ now defined in terms of u.

On the other hand, the second remark is that H does not imply MEU. To begin with, we notice that the existence of such examples seems intuitively obvious since the utility function u is a function of one variable and φ' is a function of two variables on which only mild restraints are placed. To facilitate further discussion, we define:

$$\psi(x, y) = \frac{x + y}{2} + \varphi'(x, y).$$

It is compatible with Hypothesis H to impose the following conditions on ψ:

(i) $\psi(x, y) = \psi(y, x)$
(ii) if $x < y$ then $x \leq \psi(x, y) \leq y$
(iii) if $x < x'$ and $y < y'$ then $\psi(x, y) < \psi(x', y')$.

We call these three conditions together Hypothesis J. Our experiment was so constructed that J was always satisfied. A much more restrictive additivity condition is the assumption that S's preference is not changed by increasing all amounts of money by a fixed sum. This is essentially equivalent to saying S's preference is independent of a prepayment unrelated to his bet, or that his choices are not influenced by the amount of his capital. This assumption, which we call Hypothesis A, may be expressed formally by asserting the existence of a function f such that

(iv) $f(x - y) = \varphi'(x, y)$
(v) $-\frac{1}{2}t < f(t) < \frac{1}{2}t$
(vi) $f(0) = 0$
(vii) $f(-t) = f(t)$.

Now it follows from (3) above that if Hypothesis MEU is satisfied, then

$$\psi(x, y) = u^{-1}[\tfrac{1}{2}(u(x) + u(y))]. \quad (8)$$

In terms of these hypotheses and suitable assumptions on the smoothness of the utility function (existence of second derivatives is sufficient but not necessary), it may be shown that Hypothesis A and MEU imply Hypotheses H and J, and, more importantly, imply that

$$u(x) = ax + b$$

$$\psi(x, y) = \frac{x + y}{2}.$$

In view of these results, in order to exhibit functions ψ (or φ) which are inconsistent with the existence of a utility function, we need only select a ψ which satisfies Hypothesis A, but which is such that in general

$$\psi(x, y) \neq \frac{x + y}{2}.$$

We give four such examples.

$$\psi_1(x, y) = \frac{x + y}{2} + \frac{1}{2}|x - y| = \max(x, y).$$

Use of ψ_2 corresponds to belief by S that he will always win.

$$\psi_2(x, y) = \frac{x + y}{2} - \frac{1}{2}|x - y| = \min(x, y).$$

Use of ψ_2 corresponds to belief by S that he will always lose.

$\psi_3(x, y)$

$$= \frac{x + y}{2} + \alpha|x - y|, \quad \text{with} \quad |\alpha| < \tfrac{1}{2}.$$

This gambling function expresses belief in winning with probability $\frac{1}{2} + \alpha$.

$$\psi_4(x, y) = \frac{x+y}{2} + \frac{|x-y|}{2}\left(\frac{1-(x-y)^2}{1+(x-y)^2}\right).$$

In this last case, S likes to gamble if $|x - y| < 1$, and does not like to gamble if $|x - y| > 1$. Thus, for $|x - y|$ small, ψ_4 resembles ψ_1, and for $|x - y|$ large, ψ_4 resembles ψ_2.

It may be remarked that no S in the experiment had a utility of gambling surface nearly as simple analytically as those described by $\psi_1 - \psi_4$. In fact, none actually satisfied the additivity assumption of Hypothesis A.

EXPERIMENTAL PROCEDURE

The experimental set-up here reported was the same as that described in Suppes and Walsh (4). In fact, the same Ss (eight sailors from the Moffett Field Air Base) and an additional eight undergraduate students from Stanford University (six boys and two girls) were used. When first introduced to the experiment each S was told that he would be participating in a "gambling" situation designed to give information about how people make decisions involving various amounts of money. Each S was assured that the experimental results would remain confidential and that these results would not be used to make value judgments about his character or intelligence. The purpose of the experiment, Ss were told, was purely descriptive.

After these preliminary remarks had been made, each S, protected by $2.00 playing credit, was given a set of trial alternatives of the following sort:

A	B
x	z
y	w

where x, y, z and w are small amounts of money. S was asked to pick an option, column A or column B, and was told that the row value would be decided by his tossing an unbiased die. On each side of the die was a nonsense syllable instead of a number (to avoid the effect of previously conceived preference for certain numbers), and S was asked to assign one row value to three sides of the die, the second row value being then automatically assigned to the remaining three sides (1). Ss were assured that they would not lose more than their $2.00 credit during the experiment, and they were told that the average profit in previous runs had been about $2.50 for the six sessions.[3] In order to save time Ss were not allowed to toss for all offers presented in each session, but at the end of a period of play they were instructed to draw ten numbers at random from an envelope containing as many numbers as there were offers in the session and their wins were the results of these ten tosses.

Offers presented in the first two sessions consisted of various combinations of the seven amounts: −39¢, −23¢, −10¢, 2¢, 13¢, 27¢, 42¢, such that no option dominated, or was better in both values than its companion option. The eight sailors were given 42 offers, containing seven repetitions, and no "non-gambles," or pairs of alternatives of the form (x, y) and (z, z). The eight students were given 80 offers including all alternatives given to the sailors, plus 3 more repetitions and 35 non-gambles.

Offers yielding the values which were used in constructing the utility of gambling surfaces described earlier were composed of 27 "true" options consisting of various combinations of the 20 amounts: −48¢, −38¢, −34¢, −31¢, −25¢, −19¢, −18¢, −12¢, −6¢, 1¢, 4¢, 12¢, 14¢, 18¢, 21¢, 27¢, 33¢, 34¢, 40¢, 46¢, coupled with sure-thing offers approximately half-way between the two amounts of the "true" options, and varied until the point of indifference was found, as explained earlier. More specifically, the amount z was found

[3] The experiment consisted of six sessions. The results of the second two were used to obtain the utility of gambling surfaces as above explained. Data from the first two, for the sailors, were employed in another experiment (4) designed to measure monetary utility. The results of both the first and last two sessions for both sets of Ss were used for predictions.

such that the "true" option was chosen over (z, z), but $(z + 1¢, z + 1¢)$ over the "true" option.

The 70 offers of the last two sessions were composed of pairs of "true" options neither of which dominated the other in both rows. The options consisted of 90 amounts ranging in value from $-46¢$ to $+50¢$. Both the first two and last two sessions were used for making predictions.

RESULTS

Predictions were made for the data in sessions one and two and five and six on the basis of the utility of gambling model explained above and on the basis of an "actuarial" model, that is, the model consisting of the hypothesis that *S*s simply choose options so as to maximize expected monetary outcome. The predictive results for the data of these sessions were classified into 9 categories: (1) correct (winning) predictions for both actuarial and utility of gambling models (*WW*); (2) correct prediction for actuarial model, tie for utility of gambling (*WT*); (3) correct prediction for actuarial model, wrong (losing) prediction for utility of gambling (*WL*); (4) tie for actuarial model, win for gambling (*TW*); (5) tie for both models (*TT*); (6) tie for actuarial model, wrong prediction for gambling (*TL*); (7) wrong prediction for actuarial, right for gambling (*LW*); (8) wrong prediction for actuarial, tie for gambling (*LT*); (9) wrong prediction for both models (*LL*). The summary data comparing the actuarial and utility of gambling models for both sets of *S*s, sailors and Stanford students, are presented in Table 1, sailors first and students after. (Data for three of the students were omitted because these *S*s' choices were consistently actuarial.) Following Table 1 are two typical utility of gambling surfaces, those of *S*s #1 and #9.

Since no interesting conclusions about the two models can be drawn from the gross results recorded in Table 1, a more careful analysis of the data seems to be suggested. The question we are essentially concerned with is whether one of the models has greater predictive worth than the other. That is, we wish to test the null hypothesis that the two models (gambling and actuarial) are the same in predictive power. To do this we first tabulate for each *S* the disagreements in predictions for the two models, then find, for each *S*, the significance level for rejecting the null hypothesis, and finally we combine these significance levels and apply a χ^2 test. References for this type of statistical analysis are Mosteller and Bush (3) and Moses (2).

The predictive disagreements were re-

TABLE 1

PREDICTIVE COMPARISON OF ACTUARIAL AND GAMBLING MODELS

Subject	WW	WT	WL	TW	TT	TL	LW	LT	LL	Totals
1	44	3	9	13	0	3	17	3	20	112
2	40	4	17	3	1	10	7	3	27	112
3	53	2	8	12	2	2	16	0	17	112
4	37	2	13	8	1	6	16	1	28	112
5	44	2	17	8	1	6	13	1	20	112
6	47	2	13	13	0	2	22	1	12	112
7	47	1	10	10	0	5	24	0	15	112
8	34	1	18	5	2	9	14	3	26	112
9	82	2	16	7	0	5	14	2	22	150
10	77	2	19	7	0	5	29	0	11	150
11	96	7	12	7	2	2	6	1	17	150
12	91	3	18	7	0	5	3	0	23	150
13	98	2	9	5	0	7	7	0	22	150
Totals	790	33	179	105	9	67	188	15	260	1646

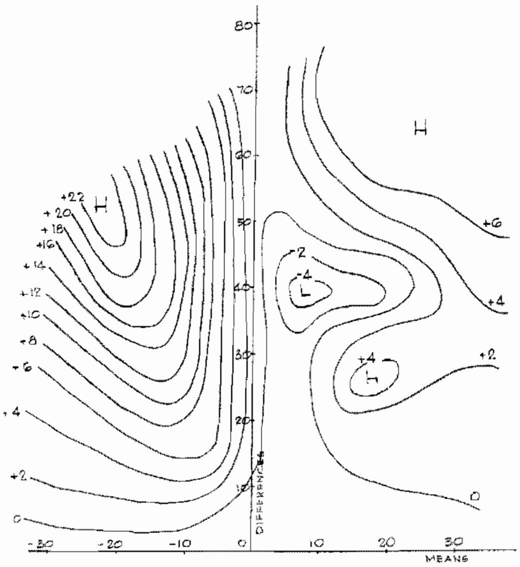

FIG. 1. Utility of Gambling Surface for Subject #1.

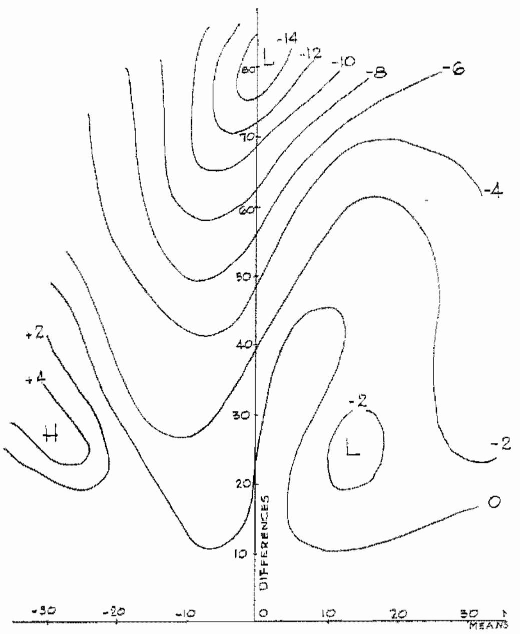

FIG. 2. Utility of Gambling Surface for Subject #9.

TABLE 2

PREDICTIVE DISAGREEMENT OF GAMBLING AND ACTUARIAL MODELS

Subject	Plus (Gambling)	Minus (Actuarial)	Number of Runs
1	31	15	26
2	12	32	18
3	28	12	18
4	25	21	22
5	22	25	23
6	36	17	26
7	34	16	18
8	22	28	27
9	23	23	30
10	36	26	32
11	14	21	24
12	10	26	18
13	12	18	18

corded in the following way. Offers for which both models made the same prediction were ignored. Offers for which the utility of gambling model predicted a correct choice, and the actuarial model predicted an incorrect choice or tie were given a plus (+). Any offer for which the gambling model predicted a tie and the actuarial model a loss was also assigned a plus. Offers for which, in the above description, the roles of the gambling and actuarial models are reversed were assigned a minus (−). The data are recorded in Table 2, the eight sailors first, then the five students. Let p be the probability of a plus. The exact significance level was found for the null hypothesis that $p = .5$ against the alternative that $p > .5$. Since a two-sided test would not properly account for the varying directionality of results, a one-sided test was used to provide a clear basis for combining significance levels. A similar one-sided test was made against the alternative that $p < .5$.

The 13 significance levels (for both sailors and students) were then combined, using, following the suggestion of Moses (2), the Lancaster correction for continuity

$$(1) \qquad \sum_{i=1}^{13} - 2 \log \left(\frac{p(X_i) + p(X_i + 1)}{2} \right)$$

for Fisher's statistic

$$(2) \qquad \sum_{i=1}^{13} - 2 \log p(X_i)$$

where $p(X_i)$ is the significance level and X_i the observed number of pluses for the i^{th} subject. Statistic (1) has approximately the χ^2 distribution with 26 degrees of freedom, two for each component.

The χ^2 results for the 13 *S*s together were $\chi^2 = 47.180$ against the hypothesis that $p > .5$, and $\chi^2 = 40.406$ against the hypothesis that $p < .5$. The first value is significant, for a two-sided test at the .01 level, the second, significant at the .05 level. Both tests resulted in a significance level $\leq .05$. Thus it seems advisable to partition *S*s into the two subgroups into which they naturally fall (sailors and students) and apply the χ^2 test to each subgroup separately.

The χ^2_{16} value for the sailors, against the hypothesis that $p > .5$ is 40.630, which is significant for a two-sided test at the .002 level. The χ^2_{10} value for the Stanford students against the same hypothesis is 6.55, which gives a significance level greater than .99. Testing the groups individually against the hypothesis that $p < .5$, the χ^2 values were 19.504 for the sailors and 20.902 for the students, with respective significance levels for a two-sided test of .5 and .05.

The clear-cut character of the statistical results for the sailors and students suggests that the development of a well-defined utility of gambling varies considerably in different cultural groups.

It might be objected that *S*s' choices between the given pairs of offers were probably not independent, and hence that the above-described statistical analysis is not applicable. However, since almost all of the offers, with the exception of seven repetitions for the sailors and ten repetitions for the students in the first two sessions, were different, there should have been no strong negative or positive recency effects. One partial check of independence is given by the following test on the number of runs of +'s and −'s. The basic idea is this, that if *S* has used some systematic method of making choices, then the number of runs recorded should be less than that expected on a random basis. For each subject we tested the null hypothesis that the +'s and −'s are randomly distributed against the alternative that there are too few runs. The observed number of runs for each subject is recorded in Table 2. The significance levels obtained were combined using statistic (1) as explained above. The result was $\chi^2_{26} = 14.309$, which for a one-sided test becomes significant at the .97 level. Thus our hypothesis of independence is well supported.

For another test of the utility of gambling model we may compare its predictive worth against that of the non-linear utility model presented in Suppes and Walsh (4). The first eight *S*s, the sailors, were used in both experiments. Predictive disagreements between the two models were treated in the same manner as predictive disagreements between the actuarial and utility of gambling models described above.[4] The individual significance levels were found for each subject for the null hypothesis $p = .5$ first against the hypothesis that $p > .5$ and second against the hypothesis that $p < .5$. Statistic (1) was again used to combine the individual levels. The χ^2_{16} value against the first alternative, that $p > .5$, was 53.922, which is significant (for a two-sided test) at beyond the .001 level. Against the alternative that $p < .5$, χ^2_{16} was 27.25 which is significant at .05. Thus, although both significance levels are $\leq .05$, the results favor the non-linear utility model.

Finally, there are some qualitative remarks which seem worth making about the utility of gambling surfaces.

(i) Only two *S*s, both students, had a surface which was everywhere positive (for the range of options tested). The fact that none of the eight sailors had such a surface is evidence against the popular conception that service men will "gamble on anything." (Yet one sailor was indifferent between the options [−2¢, −2¢] and [+4¢, −$5.00].) No *S*, sailor or student, had a surface which was always negative. These results suggest that any simple categorization of *S*s into the two classes: having a taste for gambling, having a distaste for gambling, is bound to be grossly distorting.

(ii) Six sailors and three students had a high positive utility of gambling for options with large negative means and medium standard deviations. Four sailors and two students had a negative utility of gambling for options with approximately zero means and large standard deviations. Seven sailors and two students had noticeably stronger gradients in the area defined by options with negative means than in the corresponding positive area. The existence of these quasi-uniformities of behavior should be useful in constructing a more complicated and sophisticated theory of choice behavior.

(iii) In constructing the individual utility of gambling surfaces (each based on 27 observations), it was often difficult to draw contour lines adequately accounting for each observation. It should be interesting to investigate what kind of results of the sort described in the preceding paragraph would be obtained on the basis of a very large number of observations with the contour lines drawn to fit "local" mean values of the observations.

[4] Pluses were assigned to correct predictions of the utility model, minuses to correct predictions of the gambling model.

SUMMARY

This study reports experimental results for a utility of gambling model. The model

is designed to yield a direct measurement of utility of gambling. The behavioral postulate of the model is that *S*s choose among options so as to maximize the *sum* of the expected monetary value and the utility of gambling.

The *S*s were eight sailors from Moffett Field Air Base and eight undergraduate students at Stanford University. Each *S* participated in six sessions, which primarily consisted of choosing between pairs of options having small gains or losses of money as outcomes, depending on the throw of a die. For each *S* 27 subsets of offers yielded 27 observations used to construct a utility of gambling surface.

The predictive worth of the utility of gambling model was compared with the actuarial model (maximization of expected money value) and found to be significantly better for the sailors and significantly worse for students, on the basis of a χ^2 test which combined individual significance levels.

REFERENCES

1. Davidson, D., Suppes, P., & Siegel, S. *Decision making: An experimental approach.* Stanford: Stanford Univer. Press, 1957.
2. Moses, L. E. Statistical theory and research design. *Ann. Rev. of Psychol.*, 1956, 7, 233–258.
3. Mosteller, F. & Bush, R. R. Selected quantitative techniques. In G. Lindzey (Ed.) *Handbook of social psychology.* Cambridge, Mass.: Addison-Wesley, 1954. 289–337.
4. Suppes, P. & Walsh, K. A non-linear model for the experimental measurement of utility. *Tech. Report No. 11* (ONR Contract NR 171-034), Stanford University, August, 1957.
5. von Neumann, J. & Morgenstern, O. *Theory of games and economic behavior.* (2nd ed.) Princeton: Princeton Univer. Press, 1947.

(Manuscript received July, 1958.)