

## Prediction of Bids for Two-Outcome Gambles in a Casino Setting

BARBARA GOODMAN, MARK SALTZMAN, AND

WARD EDWARDS

*University of Southern California*

AND

DAVID H. KRANTZ

*University of Michigan*

An experimental game was conducted in a Las Vegas casino. Ninety two-outcome gambles were presented in turn, and the player's monetary evaluation of each gamble was determined by the Marschak bidding procedure. Ordinal tests supported the monotonicity and the cancellation properties that are necessary for any from the family of expectation models. A number of different parametric expectation models were tested by a least-squares method. The bids were well predicted by the expected value (EV) of the gambles. More elaborate predictors, in which parameters were estimated to represent the individual player's subjective probabilities and/or utilities, merely capitalized on noise; cross-validation showed that no real improvement could be obtained over the EV predictor. Furthermore, no single feature of the bet nor any linear combination of them did as well as EV.

Traditional descriptive models of choices among bets have been variations on the notion that subjects attempt to maximize expected utility, calculated either with "objective" (whatever that means) probabilities (EU models) or with subjective probabilities (SEU models). Such models have had modest but not spectacular success in predicting decisions in simple laboratory gambling situations (for reviews of this large literature, see Becker & McClintock, 1967; Edwards, 1954c, 1961; Rapoport & Wallsten, 1972; Slovic, Fischhoff, & Lichtenstein, 1977). Because the success was modest, and perhaps also because some psychologists do not believe that maximizing expected value captures the essence of a descriptive theory of this type of behavior, workers in the field have urged other

The research reported here was sponsored by the Wood Kalb Foundation. Space and some facilities for the research were made available by the Four Queens Hotel and Casino in Las Vegas, NV. We wish to thank Amos Tversky for his help in designing this experiment and John Ponticello for collecting the data. Requests for reprints should be sent to Ward Edwards, Social Science Research Institute, University of Southern California, Los Angeles, CA 90007.

kinds of approaches. Payne (1975; Note 1) argued for process models quite different in character from SEU. Coombs and his associates (1970a, b, 1975) combined expected value (EV) maximization with an undefined notion of risk. Lichtenstein and Slovic (1971) have criticized the behavioral relevance of EV. Anderson and Shanteau (1970) and Svenson (1975) have argued for different rules to aggregate the relevant numbers.

This study has various motives. The older studies have used small stakes and have been conducted in university laboratory settings. This one raises the stakes and uses a real-life gambling setting. We thought that possibly the higher stakes and greater realism of the situation might lead to more successful prediction—for example, by extending the range of the relevant portion of the utility function for money and thus including more of its curvature. We also wanted to explore the extent to which convincing certainty equivalents for simple bets can be elicited; the question is significant for applications to decision analysis.

This paper reports a study of bids for gambles of the form  $xpy$  (obtain  $x$  with probability  $p$ , otherwise, obtain  $y$ ), conducted in the Four Queens Hotel and Casino, Las Vegas, Nevada. The bids were analyzed by means of a hierarchy of traditional expectation models—from expected value (EV), through various weaker models in which standard probabilities and/or values are replaced by personal probabilities and/or utilities (Edwards, 1962). We find that EV accounts for much of the variance, and at present we do not know of any way to improve the predictions obtained from EV in this situation.

## METHOD

### *Apparatus and Procedure*

The apparatus included a PDP-7 computer, a cathode ray tube (CRT) display, a set of push-button controls for the subject, a set of push-button controls for the experimenter, and a roulette wheel. The experimenter was a professional dealer and pit boss at the Four Queens Casino.

Each subject was run individually. After a subject was taught the game and played some practice bets, he or she chose to play with chips having one of the following values: 10¢, 25¢, \$1, \$5, or \$10. Once a choice was made, the chip value could not be changed during the play of that game.

However, if a subject chose to play another round of the game, a new choice could be made. The subject was, of course, free to stop playing the game at any time.

Ninety bets were displayed to each subject, one at a time, on the CRT. A typical bet is displayed in Fig. 1. This bet is translated as follows. The numbers of the standard roulette wheel were divided into six groups of six numbers each. Group 1 included the numbers 1 through 6. Group 2 included the numbers 7 through 12, and so on through Group 6, which included the numbers 31 through 36. The 0 and 00 were not a part of this

---

**Pick Two Groups****Yours - Win 30****Others - Lose 6**

---

FIG. 1. Display of a typical bet.

game. When either one of them came up, the roulette wheel was operated again. Thus, the probability of any one number was  $1/36$  (odds 35 to 1 against) and the probability of any group of six numbers was  $1/6$  (odds 5 to 1 against). "PICK TWO GROUPS" means that the subject could choose any two of the six groups. If a number in one of the two groups came up, 30 chips were won; otherwise 6 chips were lost.

All communications with the computer by either the dealer or a player were transmitted via a push-button network placed in front of each person. The computer, in addition to displaying the bets, kept track of what was going on, relayed a message to the subject and to the dealer if either of them made a procedural error, recorded the data, and kept financial records.

When a bet was displayed on the CRT, the subject's task was to bid for this gamble. The sale was conducted according to the Marschak bidding method (Becker, DeGroot, & Marschak, 1964). In this version of the method the subject stated a bid. Then the roulette wheel was operated, and the number that came up was fed into the computer. The computer used this number and a table stored inside it to specify a selling price. The gamble was then sold to the subject if and only if the selling price was at least as favorable to the subject as was the bid. If the sale took place, it took place at the computer's price, not at the subject's bid. The subject's bid served in effect as a cutoff point or threshold level. This cutoff point indicated the value at which the subject was indifferent between playing and not playing the bet. The subject's best strategy was to name his or her true indifference point.<sup>1</sup>

Because the sale took place at the computer's price, the game could have overall positive expectation for the subject. In order to neutralize that advantage, and also to discourage the subject from using a strategy of always making absurdly low bids, mostly not playing but occasionally

<sup>1</sup> If a person bid too little for certain bets, then the opportunity was probably missed to play some of these bets where the transactions had positive expected utility. Similarly, if a person bid too much for particular bets, then certain of these bets were probably played and those transactions had negative expected utility.

getting a real bargain, we charged a two-chip-per-bid bidding fee. With this two-chip bidding fee, each complete transaction (bidding fee, selling procedure, and play of the gamble) had an expectation of zero, if the subject named as a bid either the EV of the gamble or the EV plus one chip. For any other strategy, the expectation of the complete transaction for the subject was negative.

If a sale took place, the appropriate number of chips was exchanged. Then the subject selected the groups and the roulette wheel was operated again to play the gamble. After the subject received the winnings or paid the loss, the next bet was displayed. If a sale did not take place, the computer immediately went on to display the next gamble.

*Experimental Design*

The game consisted of 90 bets of the form  $x_i p_j y_k$  where  $x_i$  is the amount the subject received if the event with probability  $p_j$  occurred and  $y_k$  is the amount the subject received (or paid off) if the complementary event occurred. Five probability levels were used—1/6, 2/6, 3/6, 4/6, and 5/6. At each probability level there were 18 bets. These bets are displayed in matrix form in Fig. 2. The blank cells indicate bets presented to the subject. For example, the bet represented by the asterisked cell is 30,  $p_j$ , -6 which means there is a probability of  $p_j$  of winning 30 chips and a probability of  $1 - p_j$  of losing 6 chips.

*Subjects*

Thirty-five different subjects, mostly male, completed all bids. Nine of these subjects completed a second play of the game on another day. All subjects were customers of the Four Queens Casino who volunteered to play. Each person knew that this was a research game. The subjects were run individually and a typical play of the game took about 3 hr.

$x_i \backslash y_k$	0	12	18	24	30
-18					
-12					
-6					*
0					
12					

FIG. 2. Experimental design, where the blank cells indicate bets presented to the subject  $p_j = \text{Constant value}$ .

DATA ANALYSIS

We performed both an ordinal analysis (in which bids are only used to establish a rank ordering of the 90 gambles) and a series of regression analyses, using different expectation models to predict numerical values of bids.

All expectation models imply certain ordinal properties. We tested two: monotonicity and double cancellation (Krantz, Luce, Suppes, & Tversky, 1971; Tversky, 1967).

For these stimuli, the subjective ordering on each dimension can be assumed the same as the natural ordering. Monotonicity requires that the ordering of gambles be the same in each row of the matrix in Fig. 2 and likewise in each column. Within a row,  $y_k$  and  $p_j$  are constant and the bids should increase as  $x_i$  changes from 0 to 30. A similar argument applies in each column. And third, with  $x_i$  and  $y_k$  fixed, the bids should increase as  $p_j$  increases (cutting across matrices like that shown in Fig. 2).

Monotonicity predicts the order of 450 pairs of bids; 27 row and 27 column pairs for each of five matrices and 10 probability pairs for each of the 18  $(x_i, y_k)$  combinations.

Double cancellation uses the observed order of two pairs of gambles to predict the ordering of a third pair. Any  $3 \times 3$  matrix formed by the intersection of three  $x_i$  (ordered with  $x_i$  increasing) and three  $y_k$  (ordered with  $y_k$  increasing) in which bids exist for the relevant cells yields a possible test. There are 22 possible tests in each  $p_j$  matrix. An example of a testable matrix is shown in Fig. 3. If the pairs connected by single arrows are ordered in the same sense, then the pair connected by the double arrow is predicted to be ordered in that same sense. Specifically, let  $a$  through  $i$  be the bids for the games  $(18, p_j, -18), \dots, (30, p_j, -6)$ , respectively. (See Fig. 3.) Then if  $d > b$  and  $h \geq f$ , or if  $d \geq b$  and  $h > f$ , then  $g > c$ ; alternatively

$x_i \backslash y_k$	18	24	30
-18	a	b	c
-12	d	e	f
-6	g	h	i

FIG. 3. Example of a matrix which yields a test for double cancellation.  $p_j =$  Constant value.

Set A: No Free Probability Parameters	Set B: 2 Free Probability Parameters (2 fpp)
$Q(p_1) = 1/6$ $R(1-p_1) = 5/6$	$Q(p_1)$ free $R(1-p_1) = 1-Q(p_1)$
$Q(p_2) = 2/6$ $R(1-p_2) = 4/6$	$Q(p_2)$ free $R(1-p_2) = 1-Q(p_2)$
$Q(p_3) = 3/6$ $R(1-p_3) = 3/6$	$Q(p_3) = 3/6$ $R(1-p_3) = 3/6$
$Q(p_4) = 4/6$ $R(1-p_4) = 2/6$	$Q(p_4) = 1-Q(p_2)$ $R(1-p_4) = Q(p_2)$
$Q(p_5) = 5/6$ $R(1-p_5) = 1/6$	$Q(p_5) = 1-Q(p_1)$ $R(1-p_5) = Q(p_1)$

Set C: 5 Free Probability Parameters (5 fpp)	Set D: 10 Free Probability Parameters (10 fpp)
$Q(p_1)$ free $R(1-p_1) = 1-Q(p_1)$	$Q(p_1)$ free $R(1-p_1)$ free
$Q(p_2)$ free $R(1-p_2) = 1-Q(p_2)$	$Q(p_2)$ free $R(1-p_2)$ free
$Q(p_3)$ free $R(1-p_3) = 1-Q(p_3)$	$Q(p_3)$ free $R(1-p_3)$ free
$Q(p_4)$ free $R(1-p_4) = 1-Q(p_4)$	$Q(p_4)$ free $R(1-p_4)$ free
$Q(p_5)$ free $R(1-p_5) = 1-Q(p_5)$	$Q(p_5)$ free $R(1-p_5)$ free

FIG. 4. The four sets of probability constraints.

if  $d = b$  and  $h = f$ , then  $g = c$ . If the two pairs connected by single arrows are ordered in opposing directions, i.e., if  $d > b$  and  $h < f$  or  $d < b$  and  $h > f$ , no prediction can be made. (Note that none of these orderings are predictable from monotonicity considerations. For a discussion of predictions of this sort, see Coombs, Bezembinder, & Goode, 1967; Tversky, 1967.) All the expectation models make this prediction. Double cancellation predicts the order of a maximum of 110 pairs of bids per subject.

The regression analysis was based on prediction equations of the form:

Predicted bid for  $x_i p_j y_k$  is

$$\hat{Z}_{ijk} = U^{-1} [U(x_i)Q(p_j) + U(y_k)R(1 - p_j)], \tag{1}$$

where  $U$  is a utility function and  $Q$  and  $R$  are personal probability functions. Different models varied (i) the form assumed for  $U$  and (ii) the constraints on  $Q$  and  $R$ .

Four different sets of probability constraints, shown in Fig. 4, were tried. In set A,  $Q$  and  $R$  are the identity functions, i.e., personal probability = probability calculated on the basis of the symmetry of the roulette wheel. In set B, two additivity assumptions were imposed:

$$Q(p) + Q(1 - p) = 1, \tag{2}$$

and

$$R(p) + Q(1 - p) = 1.$$

In set C, only the second additivity assumption was imposed. Finally, in set D, there was no constraint. Thus, the number of free parameters was

For all Sets, i.e., 1 through 5, the parameter  $t$  represents either an  $x_i$ ,  $y_k$ , or bid amount.

Set 1: Objective Value Set

$$U(t) = t$$

Set 2: Exponential Family (exp)

$$U(t, \theta) = \begin{cases} e^{\theta t} - 1, & \text{if } \theta \neq 0 \\ t, & \text{if } \theta = 0 \end{cases}$$

Set 3: Cubic Family (cu)

$$U(t, a, b) = 1/3 t^3 - at^2 + (a^2 + b^2)t$$

Set 4: Power Family, Type 1 (p 1)

$$U(t, \alpha, \beta) = \begin{cases} t^\alpha, & t \geq 0 \\ -(-t)^\beta, & t < 0 \end{cases}$$

Set 5: Power Family, Type 2 (p 2)

$$U(t, \delta, \gamma) = \begin{cases} (1+t)^\delta - 1, & t \geq 0 \\ 1 - (1-t)^\gamma, & t < 0 \end{cases}$$

FIG. 5. The five utility functions.

0, 2, 5, or 10. In addition, the functions  $Q$  and  $R$  were constrained to be strictly increasing, in all four cases.

The five utility functions (all strictly increasing) that were tried are shown in Fig. 5. The exponential function has one free parameter,  $\theta$ . The cubic and both kinds of power functions constitute two-parameter families. The functions were all transformed linearly so that  $U(0) = 0$  and  $U(12) = 12$ .

The 20 prediction equations were formed by combining each of the four probability cases with each of the five utility functions. Note that the combination of probability set A with utility function 1 (linear) is the EV predictor. Sets B, C, and D combined with utility function 1 are subjectively expected value (SEV) models; Set A combined with functions 2–5 are expected utility (EU) models; and the remaining 12 predictors are SEU models. The prediction models had from 0 to 12 free parameters (12 in set D combined with utility functions 3–5).

The criterion was defined to be the minimum of the squared error sum (MSE)

$$D = \sum_i \sum_j \sum_k (Z_{ijk} - \hat{Z}_{ijk})^2, \tag{3}$$

where  $Z_{ijk}$  was the observed and  $\hat{Z}_{ijk}$  the predicted bid for gamble  $(x_i p_j y_k)$ .

For the 19 predictors with parameters, a computer search was conducted for parameter values that minimized  $D$ . Several different search procedures were tried. It became apparent that the minima of  $D$  are very flat. That is, almost the same value can be obtained with widely different parameters. Our search routine did not seem to encounter local minima and so we believe that the  $MSE$  values are close to their true minima.

## RESULTS

For the first half of this experiment we offered subjects the choice of playing for chips worth 25¢, \$1, \$5, and \$10. Later we added 10¢ chips to the permitted list. Of the 44 completed games, 28 were played for 25¢ chips; 11 were played for 10¢ chips; 2 were played for \$1 chips; and 3 were played for \$5 chips. We did not get many high rollers.

The largest amount won in any one completed game by a subject was \$200. The largest amount lost was \$485.

From the nine subjects who completed a repeat game we can get an estimate of the test-retest reliability. The average correlation coefficient between the repeat-game responses and the 90 bids from the first play was .89. The range went from .81 to 1.00 where the  $r$  equal to 1.00 was for the subject who bid EV perfectly 179 out of 180 times. Thus a subject's bid in a subsequent play of this game can be predicted reasonably well from his bids in the first play.

### *Ordinal Analysis*

Tests of monotonicity and double cancellation were performed on each of the 44 sets of complete data. The number of violations of monotonicity ranged (across data sets) from 0 to 75, with an average of 33. Out of 450 predictions per subject, there was an average of 7% violations with a range extending from 0 to 17%. In the strictest sense there was only one subject, subject 2, who did not violate the monotonicity criterion. However, it is hard to know how many violations to expect on the basis of "error." The only way to be sure that violations are important is to discover that they vary systematically with other variables. In fact, the violations of monotonicity seem to be concentrated in regions where the variable examined plays only a small role in determining the EV. For example, 44% of the 413 violations for  $x$  values were concentrated in the matrix with probability of  $x$  equal to 1/6, and the next largest number in the 2/6 matrix; 47% of the 483  $y$  value violations were concentrated in the matrix with 5/6 probability of  $x$  (1/6 of  $y$ ), with the next largest number in the 4/6 (2/6 of  $y$ ) matrix. The frequency of monotonicity violations for  $p$  was far higher when the range ( $x-y$ ) was low than when it was high. Thus, there were 17% violations with  $x-y = 6$ , but less than 3% when  $x-y = 48$ . The drop-off in between was smooth. The regularity of these results



suggests that the expectation models are not seriously invalidated by violations of monotonicity.

For each data set there is a maximum of 110 tests of double cancellation, not logically independent. The average number of tests that could actually be performed was 84.0 and the average number of violations was 13.8 or 16.5%. This percentage is higher than the one for monotonicity, but that may be because there are three sources of error in a cancellation test. There is systematic variation of cancellation violations with probability level. The percentage violations for  $p = 1/6$  is 5.3%,  $p = 2/6$  is 16.2%,  $p = 3/6$  is 31.0%,  $p = 4/6$  is 23.0%, and  $p = 5/6$  is 10.7%. This systematic variation may be a consequence of the systematic change in difference between EVs of the bets being compared as one moves away from  $p = 3/6$  in either direction.

### Regression Analyses

The results of an EV analysis (probability set A in Fig. 4, utility function 1 in Fig. 5), with no parameters estimated, are shown in Table 1,

TABLE 1  
EV MODEL STATISTICS FOR ALL SUBJECTS

Subject	MSE	r	b	Subject	MSE	r	b
S1	3421	.87	1.13	S15	190	.99	.98
S1*	1018	.95	1.00	S16	1791	.89	.82
S2	0	1.00	1.00	S17	18	1.00	.99
S2*	1	1.00	1.00	S18	1783	.91	1.06
S3	4797	.85	1.04	S19	1198	.93	.88
S4	1819	.96	.93	S20	150	.99	1.00
S5	1833	.87	.79	S21	394	.98	.94
S5*	1002	.93	.81	S22	872	.96	1.09
S6	1586	.89	.87	S23	191	.99	1.00
S6*	2302	.89	.95	S24	1225	.92	.84
S7	1878	.91	1.07	S25	686	.97	1.09
S8	2239	.87	.83	S26	1176	.96	1.13
S8*	931	.93	.91	S27	704	.96	1.06
S9	1020	.94	.97	S28	468	.97	1.02
S10	1583	.89	.86	S29	344	.98	1.04
S10*	1106	.93	.99	S30	1119	.94	.91
S11	1276	.91	.86	S30*	1798	.90	.81
S12	257	.98	.97	S31	818	.94	.94
S12*	199	.99	1.01	S32	1220	.95	.91
S13	1120	.94	1.06	S33	1145	.93	.95
S13*	1034	.93	.97	S34	937	.93	.83
S14	16	1.00	1.00	S35	1940	.88	.86

Note. An asterisk signifies that these data were collected from the second game for the subject in question.

where we give for each subject the *MSE* [or *D* value,  $= \frac{\sum \sum \sum (Z_{ijk} - EV)^2}{i j k}$ ], the value of the product-moment correlation, *r*, and the regression slope, *b*, of bids on EV.

Note that the gross fit of the EV model is excellent in every case. Nevertheless, subjects do differ with respect to goodness of fit; and for subjects for whom the fit is not perfect, one may hope to predict more accurately by using a model that takes into account personal probabilities or utilities. Therefore, the remaining 19 expectation models were tested on a sample of data sets. We chose ones for which the EV fit was not perfect (correlations ranging from .85 to .97) and for which the scatterplot showed some suggestion of nonlinearity.

The results of all 20 regression analyses, for the 11 selected data sets, are shown in Table 2.

The entries in Table 2 show the *MSE* for each model divided by the error for the EV model; they thus show the percentage of residual squared prediction error after estimating the parameters for the expectation model in question. Roughly speaking, the reduction in prediction error depends on the number of parameters estimated—from 5.6% (average) for 1 parameter (the exponential EV model in column 3) to about 40% for 12 parameters (the last three columns of the table).

How significant and how important are the improvements in fit obtained with multiparameter expectation models? The importance is obviously limited by the high correlations shown for EV (Table 1). Still, the reduction in prediction error by the use of multiparameter models could be very important for theories of choice, provided that these models really capture characteristics of individuals in their parameters.

Convincing evidence that the improvements over EV are both nonsignificant and unimportant can be obtained by examining the nine subjects for whom two complete data sets were collected. The question is, are the bids in the second data set better predicted by EV or by the best fitting multiparameter expectation model from the first session's data?

A good example is offered by subject 6 and subject 6\* whose analyses are presented in the third and fourth rows of Table 2. We see that on both occasions, the best multiparameter model (D4) produces over a 40% drop in *MSE*, compared with the EV prediction. For this subject, EV predictions are much better than test-retest reliability. The latter correlation is .81, while EV correlates .89 with each set of data (Table 1). If we use the D4 parameters from the first session to predict the second session's data, however, there is no improvement over an EV predictor; the correlation is .885.

A full set of comparisons of this type is given in Table 3. For each subject we compare the correlation coefficients between the second ses-

TABLE 2  
MINIMUM SQUARED ERRORS FOR 19 EXPECTATION MODELS

Subject	EV error	EU models					SEV models				
		Set 2:exp	Set 3:cu	Set 4:pl	Set 5:p2	Set B(2)	Set C(5)	Set D(10)			
S3	4797	.855	.850	.892	.897	.964	.621	.575			
S5	1833	.998	.998	.957	.949	.629	.622	.608			
S6	1586	.987	.987	.976	.986	.919	.762	.659			
S6*	2302	.761	.725	.647	.649	.991	.670	.621			
S10*	1106	.988	.906	.951	.950	.901	.879	.815			
S13*	1034	.988	.987	.993	.995	.875	.867	.722			
S18	1783	.967	.960	.834	.843	.948	.818	.765			
S24	1225	.954	.952	.869	.846	.701	.616	.541			
S25	686	.993	.897	.790	.789	.838	.628	.586			
S31	818	.943	.843	.840	.852	.870	.703	.622			
S33	1145	.972	.973	.975	.979	.896	.725	.649			
Average		.946 (1)	.925 (2)	.884 (2)	.885 (2)	.867 (2)	.719 (5)	.649 (10)			

TABLE 2—continued

Subject	SEU B(2)				SEU C(5)				SEU D(10)			
	Set 2:exp	Set 3:cu	Set 4:p1	Set 5:p2	Set 2:exp	Set 3:cu	Set 4:p1	Set 5:p2	Set 2:exp	Set 3:cu	Set 4:p1	Set 5:p2
S3	.849	.841	.886	.892	.615	.621	.572	.573	.576	.572	.556	.564
S5	.626	.619	.622	.624	.608	.609	.609	.613	.632	.607	.606	.606
S6	.900	.900	.868	.877	.762	.759	.726	.738	.673	.647	.580	.607
S6*	.756	.758	.639	.653	.663	.664	.649	.653	.619	.618	.594	.604
S10*	.887	.617	.728	.723	.819	.555	.763	.800	.727	.552	.680	.684
S13*	.864	.861	.811	.818	.806	.815	.791	.809	.691	.695	.711	.714
S18	.923	.896	.801	.810	.778	.776	.686	.661	.703	.712	.680	.655
S24	.669	.670	.668	.664	.406	.438	.429	.420	.407	.435	.467	.447
S25	.834	.809	.700	.708	.628	.611	.561	.570	.561	.555	.500	.528
S31	.829	.778	.661	.670	.700	.678	.638	.565	.622	.572	.553	.561
S33	.862	.845	1.000	.818	.649	.674	.552	.617	.555	.584	.569	.577
Average	.818 (3)	.781 (4)	.762 (4)	.751 (4)	.676 (6)	.655 (7)	.634 (7)	.638 (7)	.615 (11)	.595 (12)	.591 (12)	.595 (12)

Note. Each of these *MSEs* is expressed as a fraction of the squared prediction error of the EV model. The number of free parameters in a model is given in parentheses following the average fractional *MSE*. An asterisk signifies that these data were collected from the second game for the subject in question.

sion's data and three predictors: first-session data (test-retest), EV, and best first-session multiparametric model.

The table shows that, by and large, EV is substantially better than first-session bids. The EV model successfully removes a substantial random error component from the subject's bids. Moreover, it predicts well.

Out of the 44 possible bids for each of the bets, over all subject data sets, on the average 21.5 or about half were within one chip of EV. This does not mean that subjects simply bid EV routinely. The average of the ratios of range of bids to range between best and worst outcome of each bet was 0.79. Although subjects did not routinely bid EV, Table 3 indicates that their individual bids were highly correlated with EV and that none of the more complex models we tried were meaningful improvements over EV in predicting bids.

In view of the above results, our best guess is that the results of Table 2 cannot be interpreted as favoring any particular class of multiparametric models. (In any case, the present design would not have been appropriate for comparing SEV with EU models, because of the differing degrees to which their respective parameters are constrained by the data collected here. Each bid places a constraint on three utility values and on two subjective-probability values.)

## DISCUSSION

The purpose of this experiment was to determine which of the different expectation models best accounted for subjects' behavior in this gambling task according to our set of established criteria. In the light of the data, this was the wrong question to ask for many reasons.

First of all, on the average, 88% of the variance of the maximum buying

TABLE 3  
CORRELATIONS OF SECOND-SESSION BIDS WITH THREE PREDICTORS

Subject	First-session bids	EV	Best first-session model
S1*	.86	.95	.92
S2*	1.00	1.00	—
S5*	.86	.93	.94
S6*	.81	.89	.885
S8*	.83	.93	.90
S10*	.89	.93	.93
S12*	.97	.99	.99
S13*	.93	.93	.94
S30*	.89	.90	.88

*Note.* An asterisk signifies that these data were collected from the second game for the subject in question.

prices can be accounted for by the simplest model. Moreover, some fraction of the variance is irreducible. Therefore, almost no variance is left over for the more sophisticated models to explain.

Second, although the additional degrees of freedom of the subjective model allow a better fit, one must ask whether anything is to be gained by permitting the additional degrees of freedom. Certainly, the regression statistics did not improve much. Furthermore, for at least half of the subjects, the regression line for predicted bids against predictor values had slopes further from one, using subjective values than using EV. Additionally, once subjective parameters are allowed, one must arbitrarily decide which set of values to use, since it turns out that very different sets of subjective parameters yield about the same *MSE*.

This study shows that the EV model fits the data as well as the more complex expectation models. But nontrivial residuals from EV exist. On the average, half of the subjects bid within one chip of EV, and EV was the modal value of the distribution of bids for 81 out of the 90 bets. But intersubject variation was substantial. Slovic and Lichtenstein (1968) and Anderson and Shanteau (1970) have proposed several lines of thought in which linear combinations of the characteristics of a bet are used as predictors of behavior with respect to it. We explored a number of versions of this idea.

The range of chip values is the single most important determinant of the standard deviation of the bids for a bet. (*F* test significant beyond the .0001 level.) The probability level is the only simple feature of a bet (*F* test significant at the .05 level) that affects the skewness of the distribution of bids. Subjects overbid for the 3/6 probability bets (Edwards, 1953, 1954a, 1954b) and underbid for the 5/6 probability bets.

No simple feature of a bet,  $x$ ,  $y$ , or  $p$ , taken singly or in linear combinations can account for the data as well as EV does. In regression analyses of both individual and grouped data, no single dimension could even begin to compare with EV in explanatory power. Table 4 shows the percentage of variance of individual bids unaccounted for by various predictors, averaged over subjects. For comparison, it also presents the same calculations for the subject who always bid EV. The subjects clearly were not focusing primarily on  $x$ , or  $y$ , or  $p$ . The fact that 20% of the variance is unaccounted for by linear combinations of  $x$ ,  $y$ , and  $p$  (as compared with 11% for a perfect EV subject) rules out a variety of other simplistic wrong strategies. Introducing nonlinear terms did not help. Nor did using probability as five variables rather than one.

Of the residual variance not accounted for by EV, 14% on the average is explicable by linear combinations of  $x$ ,  $y$ , and  $p$ . This is almost certainly a typical regression artifact.

These statements are based on data aggregated over subjects. Perhaps

TABLE 4  
PERCENTAGE VARIANCE UNACCOUNTED FOR

	Percentage Variance Unaccounted for by				Linear combination of $x$ , $y$ , $p$
	EV	$x$ alone	$y$ alone	$p$ alone	
Mean subject	12	78	78	66	20
Perfect EV subject	0	79	79	56	11

individual subjects used idiosyncratic strategies that aggregate to look like EV. We repeated all these analyses on a single-subject basis, and found little reason to believe that subjects might be grouped by idiosyncratic strategies; there are no  $x$  subjects,  $y$  subjects, and  $p$  subjects, for example. However, the linear regressions for different subjects show some variation. There are some subjects for whom the greatest single factor contributing to the variance was  $x$ , while for other subjects  $y$  or  $p$  accounted for the most variance.

#### *Implications for the Marschak Bidding Procedure*

The Marschak bidding procedure, used in this study, has been widely advocated because it is the only available method for asking a subject to state a certainty equivalent for a bet that makes it optimal to state his or her true certainty equivalent. Since these data were collected, von Winterfeldt and Edwards (Note 2) have pointed out that decision-theoretical optima are universally flat—that is, that significant deviations from optimal strategy produce only relatively small reductions in EV. How much EV would a suboptimal subject lose in this experiment? Consider the bet (30, 0.5, -18), for which the losses from suboptimal bids are maximal. The least optimal acceptable bids are, of course, 30 and -18. The expected losses (in chips) associated with them are -2.5 and -2, respectively. The expected loss functions are smooth, of course. For that same bet, a bid of two chips above the higher optimal bid would lose one-sixth of a chip on the average and a bid of two chips below the lower optimal bid also would lose one-sixth of a chip on the average.

Deviations from optimal bidding cost virtually nothing in EV (or EU, or SEU; the arguments are the same for all expectation maximizing models). Yet EV was the overwhelming determiner of bids. Most subjects varied their bids around the EV in a way that we could not predict from utility and probability transformations. If suboptimality, random or systematic, is so cheap, why did EV account for 88% of the variance in bids?

We can think of two possible answers to this question. One is that subjects may have taken very seriously the instruction, emphasized by the detailed explanation of the properties of the Marschak procedure, that

they should bid exactly what the bet is worth to them. For near-linear utility functions and close-to-correct personal probabilities, combined according to an intuitive version of the appropriate calculation, the result would be bids on or close to the EV. This hypothesis would imply that the Marschak bidding procedure serves primarily an instructional function; its decision-theoretical optimal strategy properties are relevant only in reinforcing the instruction.

A second possible explanation is that these subjects could, at least approximately, calculate EVs. It may be that some significant percentage of the general population does use such a strategy, especially a significant portion of the population that visits Las Vegas casinos. However, it may also be that there were certain characteristics of this situation that made an EV strategy especially salient here. Some of the subjects were professionals, dealers, or pit bosses. Since we did not record personal information about our subjects, we do not know this percentage. There were 90 plays to each game. Subjects knowing that they would be playing a repetitive type game for a long time might be encouraged to develop some strategy to reduce the information processing loads placed on them. Furthermore, the payoffs and the probabilities were explicit. In fact, the payoffs were integer multiples of the probabilities, making the EV strategy easy to calculate. Finally, while the stakes were much larger than typically used in laboratory experiments, the total money involved may have represented only a small fraction of many players' total wealth. Any one or combination of these factors may account for an EV strategy being used by these subjects in this situation which would not generalize to quite different groups of people faced with quite different decision problems. This explanation is not alternative to the one proposed in the previous paragraph; it supplements it, and both might be true.

#### *Implications for Decision Analysis*

Most formal treatments of elicitation methods in decision analysis (e.g., Keeney & Raiffa, 1976; Raiffa, 1968) call for elicitation procedures that require respondents to state the certainty equivalents of imaginary bets. Edwards (1977) has criticized such procedures and proposed others of a more psychophysical nature, mostly out of distrust of choices among bets, real or hypothetical, as an effective way of obtaining information about opinions or values.

This study produced no basis for preferring other expectation models to EV, though that was its purpose. For all expectation models, including EV, it produced fairly high MSEs for some subjects. A priori, for these subjects at least, it seemed that an expectation model allowing some freedom in the utility or probability parameters should provide a better fit to the data than EV. It did not. Furthermore, the results enhance our



distrust for inferring subjective probabilities and utilities from elicited responses such as the certainty equivalents used in this experiment. This distrust arises because very different combinations of inferred subjective probabilities and utilities lead to the same or very similar goodness of fit. Therefore, it is not possible to determine which combination is "the" appropriate one. In such cases, the flat maximum phenomenon compounds the problem. If small divergences from EV are allowed, then an even larger set of possible combinations of subjective probabilities and utilities becomes acceptable.

A decision analyst might legitimately complain that this experiment used untrained subjects, a complex procedure, and stakes very small relative to those of most decision problems for which decision analyses are done. True. Yet we remain unconvinced that greater training and higher stakes will induce enough additional stability of assessment to compensate for the instability that we think would almost certainly be produced by the use, for example, of multidimensional payoffs. These data, though far from conclusive, should not enhance the confidence of those who use elicitation methods based on obtaining certainty equivalents of imaginary bets.

## REFERENCES

- Anderson, N. H., & Shanteau, J. C. Information integration in risky decision making. *Journal of Experimental Psychology*, 1970, **84**, 441-451.
- Becker, G. M., DeGroot, M. H., & Marschak, J. Measuring utility by a single-response sequential method. *Behavioral Science*, 1964, **9**, 226-232.
- Becker, G. M., & McClintock, C. G. Value: Behavioral decision theory. *Annual Review of Psychology*, 1967, **18**, 239-286.
- Coombs, C. H. Portfolio theory and the measurement of risk. In M. F. Kaplan & S. Schwartz (Eds.), *Human judgment and decision processes*. New York: Academic Press, 1975.
- Coombs, C. H., Bezembinder, T. G. G., & Goode, F. M. Testing expectation theories of decision making without measuring utility or subjective probability. *Journal of Mathematical Psychology*, 1967, **4**, 72-103.
- Coombs, C. H., & Huang, L. C. Polynomial psychophysics of risk. *Journal of Mathematical Psychology*, 1970, **7**, 317-338. (a)
- Coombs, C. H., & Huang, L. C. Tests of a portfolio theory of risk preference. *Journal of Experimental Psychology*, 1970, **85**, 23-29. (b)
- Edwards, W. Probability-preference in gambling. *American Journal of Psychology*, 1953, **66**, 349-364.
- Edwards, W. Probability-preferences among bets with differing expected values. *American Journal of Psychology*, 1954, **67**, 56-67. (a)
- Edwards, W. The reliability of probability preferences. *American Journal of Psychology*, 1954, **67**, 68-95. (b)
- Edwards, W. The theory of decision making. *Psychological Bulletin*, 1954, **51**, 380-417. (c)
- Edwards, W. Behavioral decision theory. *Annual Review of Psychology*, 1961, **12**, 473-498.
- Edwards, W. Subjective probabilities inferred from decisions. *Psychological Review*, 1962, **69**, 109-135.

- Edwards, W. How to Use Multi-attribute Utility Measurement for Social Decision Making. *IEEE Transactions on Systems, Man, and Cybernetics*, 1977, 7, 326-340.
- Keeney, R. L., & Raiffa, H. *Decisions with multiple objectives: Preferences and value tradeoffs*. New York: Wiley, 1976.
- Krantz, D. H., Luce, R. D., Suppes, P., & Tversky, A. *Foundations of measurement*. New York: Academic Press, 1971.
- Lichtenstein, S. C., & Slovic, P. Reversals of preference between bids and choices in gambling decisions. *Journal of Psychology*, 1971, 89, 46-55.
- Payne, J. W. Relation of perceived risk to preferences among gambles. *Journal of Experimental Psychology: Human Perception and Performance*, 1975, 104, 86-94.
- Raiffa, H. *Decision analysis: Introductory lectures on choices under uncertainty*. Reading, Mass.: Addison-Wesley, 1968.
- Rapoport, A., & Wallsten, T. S. Individual decision behavior. *Annual Review of Psychology*, 1972, 23, 131-175.
- Shanteau, J. An information-integration analysis of risky decision making. In M. F. Kaplan & S. Schwartz (Eds.), *Human Judgment and Decision Processes*. New York: Academic Press, 1975.
- Shanteau, J. & Phelps, R. H. Judgment and swine: Approaches and issues in applied judgment analysis. In M. F. Kaplan & S. Schwartz (Eds.), *Human judgment and decision processes: Applications in problem settings*. New York: Academic Press, 1976.
- Slovic, P., Fischhoff, B., & Lichtenstein, S. Behavioral decision theory. *Annual Review of Psychology*, 1977, 28, 1-39.
- Slovic P., & Lichtenstein, S. Relative importance of probabilities and payoffs in risk taking. *Journal of Experimental Psychology*, 1968, 78(no. 3, part 2).
- Svenson, O. A unifying interpretation of different models for the integration of information when evaluating gambles. *Scandinavian Journal of Psychology*, 1975, 16, 187-192.
- Tversky, A. Additivity, utility, and subjective probability. *Journal of Mathematical Psychology*, 1967, 4, 175-202.

#### REFERENCE NOTES

1. Payne, J. W. *A process tracing study of risky decision making: Examples of protocols and comments*. Complex Information Processing Working Paper No. 274, Carnegie-Mellon University, Pittsburgh, September 1974.
2. von Winterfeldt, D. & Edwards, W. *Flat maxima in linear optimization models*. (Engineering Psychol. Lab. Tech. Rpt. 011313-4-T) Ann Arbor: University of Michigan, 1973.

RECEIVED: October 2, 1978