The American Journal of Psychology, Vol. 75, No. 1. (Mar., 1962), pp. 35-44.

CHOICES AMONG UNCERTAIN OUTCOMES: A TEST OF A DECOMPOSITION AND TWO ASSUMPTIONS OF TRANSITIVITY

By BETTY J. GRISWOLD, Columbia University, and
R. DUNCAN LUCE, University of Pennsylvania

During the past few years, considerable attention has been devoted to mathematical descriptions of choices among uncertain outcomes (gambles) and the theories based on these descriptions have led to numerous empirical studies.[1] The guiding principle has been that choices can be represented as a maximization of expected utility. More exactly, the idea is to assign, for every subject (*S*), utilities to gambles and subjective probabilities to events so that (1) the utility of a gamble equals the expectation of the utilities of the constituent outcomes of the gamble relative to the subjective probabilities of the events upon which these outcomes are contingent; and (2) the ordering of the gambles by utility reflects *S*'s preferences. The earliest theories were not probabilistic, which made them difficult to test because, even in a short series of judgments, the *S*s often make intransitive and inconsistent choices. Whether this is due to some change in *S* during the experiment or whether it indicates a probabilistic determination of choices that permits certain local inconsistencies is not easy to decide. The second alternative is, however, simpler to formalize, hence it is explored first.

Our starting point is a probabilistic theory that has been presented in

* Received for publication November 18, 1958. This experiment was conducted at the Bureau of Applied Social Research, Columbia University, and was supported in part by National Science Foundation grants NSF-G2803 and G4506 and by National Institute of Health grant M-2293. Our thanks are due Professors Eugene Galanter and William J. McGill for advice concerning the design of this study.

[1] See J. S. Chipman, Stochastic choice and subjective probability, in D. Willner (ed.), *Decisions, Values, and Groups,* 1960, 70-95. C. H. Coombs, On the use of inconsistency of preferences in psychological measurement, *J. exp. Psychol.,* 55, 1958, 1-7; Donald Davidson and J. Marschak, Experimental tests of a stochastic decision theory, in C. W. Churchman and Philburn Ratoosh (eds.) *Measurement: Definitions and Theories,* 1959, 233-269; Donald Davidson, Patrick Suppes, and Sidney Siegel, *Decision Making,* 1957; and Frederick Mosteller and Philip Nogee, An experimental measurement of utility, *J. pol. Econ.,* 59, 1951, 371-404. Two general surveys of work in this area are contained in Ward Edwards, The theory of decision making, *Psychol. Bull.,* 51, 1954, 380-417, and R. D. Luce and Howard Raiffa, *Games and Decisions,* 1957.

the economic literature.[2] We will not attempt a detailed summary here, but several features must be mentioned.

First, an explanation of the notation: Lower case Latin letters 'a,' 'b,' . . . . denote pure outcomes, such as receiving a book or losing a sum of money, and lower case Greek letters 'α,' 'β,' . . . denote events whose occurrences are uncertain, such as "a six coming up on a throw of a die" or "it will rain in Boston tomorrow." A symbol of the form 'aαb' denotes the gamble in which 'a' is the outcome if the event α occurs and 'b,' if it does not. It is assumed that if two gambles, 'aαb' and 'cβd,' are presented to *S*, then there is a probability $P(\text{a}\alpha\text{b}, \text{c}\beta\text{d})$ that *S* will choose 'aαb' in preference to 'cβd.' Although it is not clear how to introduce expected utility into a probabilistic framework, the intuition that choices between two gambles can be decomposed by considering their constituent parts—outcomes and events—can be retained in a much weakened form.

Suppose that there is a probability $Q(\alpha,\beta)$ that *S* judges Event α more likely to occur than Event β. Moreover, suppose that *S*'s preferences between pairs of outcomes are statistically independent of his judgments about the likelihood of events, then it is not difficult to see that

$$P(\text{a}\alpha\text{b}, \text{a}\beta\text{b}) = P(\text{a, b})Q(\alpha, \beta) + P(\text{b, a})Q(\beta, \alpha) \cdot \cdot \cdot \quad [1]$$

We take Equation [1], known as the *decomposition-axiom,* as a basic assumption to be tested. The easiest way to do so is to note that it is equivalent to

$$Q(\alpha,\beta) = \frac{P(\text{a}\alpha\text{b}, \text{a}\beta\text{b})P(\text{a, b}) - P(\text{a}\beta\text{b}, \text{a}\alpha\text{b})P(\text{b, a})}{P(\text{a, b})^2 - P(\text{b, a})^2} \cdot \cdot \cdot \quad [2]$$

The right side of Equation [2] is expressed entirely in terms of preference-probabilities, and, although each of these probabilities depends upon the outcomes a and b, Equation [2] states that this particular function of the preference-probabilities is independent of a and b.[3]

The original purpose of the present study was: (a) to make this test in isolation; and (b) to test the whole of the theoretical structure given in the paper cited above. One feature of that structure—as of so many probabilistic models in psychology—is that practically all of the probabilities of a choice must be different from 0 and 1. This leads to doubt about the model at the outset, and it certainly implies that it is wrong when the outcomes are money, because the larger of two sums is always preferred. Nonetheless, we had hoped that we could choose outcomes for which the preferences would not be completely consistent. We selected for this purpose a set of four brands of cigarettes that lay between the fourth and the tenth rank on each *S*'s ordering of 30 popular bands. In addition, we studied gambles with money to see what difference might exist between the two classes of outcome.

---

[2] R. D. Luce, A probabilistic theory of utility, *Econometrica,* 26, 1958, 193-224. See also Appendix 1 of Luce and Raiffa, *op. cit.,* 371-384.

[3] The statistical issue in estimating $Q(\alpha,\beta)$ and in testing whether the right side of Equation 2 is independent of a and b are discussed in Luce, *op. cit.,* 217-219.

As will be seen, the choices of our *S*s exhibit many proportions that are 0 or 1, and so we were not successful in meeting the requirement of the over-all theory. Thus, the complete theory is an inadequate description of the results obtained from these *S*s. A more detailed analysis of the data is needed, however, before we can draw conclusions about the adequacy of the decomposition-axiom.

While the experiment was under way, another, more realistic, theory of choice was developed that rests primarily upon one fundamental axiom.[4] This axiom, when coupled with the decomposition-axiom, leads to certain predictions that can be used to test the theory.[5] Although the present experiment was not designed with this theory in mind, and so we cannot make some of the more powerful tests, it is suitable to check two weaker consequences.

The first is the well-known property of transitivity, namely:

$$\text{if } P(x, y) = 1 \text{ and } P(y, z) = 1, \text{ then } P(x, z) = 1, \cdots \qquad [3]$$

where x, y, and z are gambles. The second is strong stochastic transitivity,[6] *ie.*

$$\text{if } 1/2 < P(x, y) < 1 \text{ and } 1/2 < P(y, z) < 1,$$
$$\text{then } P(x, z) > P(x, y) \text{ and } P(x, z) > P(y, z). \cdots \qquad [4]$$

Strong stochastic transitivity is a generalization of the familiar idea that if x is one *jnd* larger than y and y one *jnd* larger than z, then x will be at least one *jnd* larger than z. In stating this requirement, we do not restrict ourselves to *jnd*-separations, but simply require that if there is a tendency to choose x over y and y over z, then the probability of choosing x over z shall not be less than either of the other two probabilities.

Our primary aim, then, will be to determine for each *S* whether the requirements of the decomposition-axiom, of transitivity, and of strong stochastic transitivity are met. Prior to that, however, it will be necessary to determine whether our *S*s exhibited any significant sequential effects, such as learning, because each choice was presented many times over a period of weeks. The choices of each *S* will, therefore, be analyzed separately.

Method: Subjects. The *S*s (5 in number—1 woman and 4 men) were students of Columbia University; each smoked at least 20 cigarettes per day. They were paid $1.00 for each experimental session and, in addition, they won a package of some brand of cigarettes and a sum of money ranging from 1 cent to 50 cents averaging 24.9 cents per session. The *S*s were naïve regarding the purpose of the experiment.

[4] Luce, *Individual Choice Behavior,* 1959, 6.
[5] *Ibid.,* 75-90.
[6] *Ibid.,* 19.

Procedure. On each of the 74 trials of a session E presented S with a choice between two gambles, such as "fifty cents if red occurs and twenty cents if not, or fifty cents if green occurs and one cent if not." A second class of gambles involved cigarette brands as outcomes. After S decided which of the two gambles he preferred, the outcome was determined by E on a pinball machine. The choice and outcome were recorded by E, but S was not paid off until the end of the session. E then selected at random one trial paid in cigarettes and one in money, and S received the recorded outcome.

E randomized the order of presentation of the gambles within each session by selecting, without replacement, numbered chips from a box. These chips were also used to determine which trials were actually paid off. The order in which a gamble was read was reversed from session to session and within any one session half the pairs had the larger expected value read first.

Gambles. Sixty-eight of the 74 choices were of the form: $a\alpha b$ vs. $c\beta d$, where the outcomes, a,b,c, and d, were not necessarily all different. In 34 of these 68 cases, the outcomes were sums of money, the amounts being 1, 10, 20, and 50; in the other 34, they were four brands of cigarette. The events, α and β, were characterized in terms of one of four colors occurring in the pinball machine. The six remaining choices were between pairs of the four cigarette brands.

Sessions. Each session consisted of 74 choices between a pair of gambles. Whenever possible, six sessions were run per week, but there was never more than one per day. The first five sessions were pre-training and they are not included in the analyses. Following that, data were collected for 50 sessions from S_1, S_3, and S_4; for 45 sessions from S_2; and for 32 sessions from S_5.

During the first and at the end of Session 17 and 34, and the final session, S was required to rank order 30 popularly priced types of cigarettes (18 brands, some in variations of regular or king size and of filter or not) and to answer questions about his smoking habits and preferences. On the basis of the first ranking, four brands were chosen for each S from the upper third of his range and these were used as rewards throughout the experiment.

Apparatus. The pinball machine consisted of a 12×18-in. tilted board. A glass marble when released at the top of the board would wend its way among the 22 bumpers of the board into one of 10 compartments at the bottom. One roll of the marble would determine the occurrence or non-occurrence of a given event, such as red or not red.

Each compartment was labelled by two different colors. Each of the four colors, red, green, black, and silver, appeared on five compartments. The red event, for example, occurred if the marble dropped into any of the five red compartments; it did not occur if the marble dropped in one of the other five. It will be recalled that a gamble paid off as a consequence of a given event.

The bumpers were so located that the marble entered the compartments with unequal frequencies, thus making it difficult for S to calculate the probability of an event. The colors were assigned in an attempt to produce events with objective probabilities of 0.33, 0.45, 0.55, and 0.67. Prior to the experiment proper, we rolled the marble 500 times to get relative frequencies for the four colors and the stability of the mechanism was checked from 200 randomly chosen trials for each event during the latter half of the experiment. These two sets of relative frequencies are shown in Table I. It will be noted that the largest difference between the first and the second test is about two standard deviations and the other three differences are much less.

Results: (1) Consistency of preferences. The theory proposed states that almost no preferences can be consistent. The percentages of cases for which the observed proportions were 0 or 1 are summarized in Table II. For four of the five *S*s, a substantial fraction, more than 32%, of the choices was perfectly consistent; therefore, we conclude that this theory cannot be applied generally. It was deemed not worthwhile to attempt to check it for the remaining *S*, *i.e.* S_2.

(2) Sequential effects. Any experiment that extends over two and a half months, that involves events with initially unknown probabilities, and that

TABLE I

RELATIVE FREQUENCIES OF THE FOUR EVENTS ON WHICH THE OUTCOME OF A GAMBLE DEPENDED

Event	500 pre-experimental rolls	200 experimental rolls
α	.672	.615
β	.546	.540
γ	.445	.465
δ	.337	.335

TABLE II

PERCENTAGE OF PAIRS OF GAMBLES FOR WHICH ONE ALTERNATIVE WAS ALWAYS PREFERRED TO THE OTHER

S	Cigarettes (6 cases)	Cigarette gambles (34 cases)	Money gambles (34 cases)
1	83.5	32.4	76.5
2	66.7	0.0	17.7
3	66.7	32.4	55.9
4	100.0	32.4	38.3
5	100.0	67.7	79.5

has cigarette brands as outcomes, is an experiment for which it would be unwise to assume without investigation that the choices will have constant probabilities. We shall examine this question in two ways.

First, we ask whether there is a consistent trend in the proportion of choices of gambles from the first to the second half of the experiment. Secondly, we ask whether the run-structure of the responses suggests a lack of independence in the choices. In both analyses, we separate the cigarette data into those choices involving at least one pair of brands for which *S*'s preferences were not invariably the same and into those choices that did not involve any inconsistent pair of brands. We expect that experience with the brands may well have modified preferences for a brand of cigarettes over time, and this is most likely to have happened with pairs of brands for which S was initially inconsistent.

Using the data from the first half of the experiment, we estimated the probabilities of each choice. Pairs of gambles with a proportion of 0 or 1 were not considered; all other pairs were taken in the order of the pair which was chosen more than half the time. For these pairs, the proportion of choices was also calculated from the second half of the data. The number of cases for which the proportion from the second half was larger, the same, or smaller was determined and a sign test applied. The resulting p-values are shown in Table III. Where there were inconsistencies in the preference for a brand, there is a significant (5% level) change in the choice of gambles involving these pairs, but no significant

TABLE III

PROBABILITY, BASED ON A SIGN TEST, THAT THE PROPORTION OF CHOICES CHANGED FROM THE FIRST TO THE SECOND HALF OF THE EXPERIMENT BY CHANCE

The number of cases on which the test is based is shown in parentheses.

	Outcome of gamble		
S	Brands not consistently preferred	Brands consistently preferred	Money
1	.05 (10)	>.25 (11)	>.25 (8)
2	.05 (18)	>.25 (15)	>.25 (22)
3	.05 (7)	.25 (11)	.25 (8)
4	—	<.01 (22)	.01 (20)
5	—	.05 (10)	.25 (7)

effect is found either in the gambles involving consistently chosen brands or in the gambles for money. Thus, apparently, for S_1, S_2, and S_3 there were some changes of preference among the brands, but no systematic changes in their judgments of the relative likelihood of the events used in the gambles. S_4, however, exhibits a significant (1% level) changes in both classes of gambles; the direction is toward greater consistency, suggesting that his estimate of the likelihood of the various colors improved with experience. Finally, S_5 exhibits a significant (5% level) change in the gambles for cigarettes, but not in the gambles for money. One suspects that she altered the relative spacing of her preferences for some brands without, however, affecting her choices of the gambles paid off by cigarettes, which remained consistent.

To look into more local effects, we examined the runs of responses to gambles for which the over-all proportion of choices was different from 0 and 1. The results are summarized in Table IV. S_2, S_3, and S_5 gave runs of responses above chance (5% level) to slightly more pairs than one would expect. For example, S_2's gambles with payoffs in money led to 3 significant cases out of 28, whereas by chance there should have been 1.4. The discrepancies are so small, however, that we doubt that there were serious deviations for these *S*s. S_4, whose choices differed in the first and second halves of the experiment, also exhibits a significant number of runs in more than the expected number of cases. So does S_1. In almost all cases, there were too few rather than too many runs, thus suggesting that there was a perseveration effect.

Although there are some deviations in the responses of our *S*s from a model of independent Bernoulli observations, there are still enough cases where this model appears adequate to warrant analysis of the data in terms of the three hypotheses that were discussed earlier.

(3) Transitivity. For both the brands and money, there were 23 distinct triplets of gambles in which *S* made choices between all three pairs. Of

TABLE IV

THE NUMBER OF PAIRS OF GAMBLES THAT EXHIBITED A SIGNIFICANT (5% LEVEL) NUMBER OF RUNS

	Cigarettes		Money	
S	No. with more than one run	No. with a significant deviation	No. with more than one run	No. with a significant deviation
1	8	6	23	5
2	36	1	28	3
3	25	3	15	1
4	23	3	21	5
5	11	1	7	1

the data for these triplets, some satisfy the hypotheses of Equation [3] (transitivity-property) provided that estimates $\hat{P}$ are substituted for probabilities *P*, some appear to satisfy the hypotheses of Equation [4] (strong stochastic transitivity) and the rest satisfy neither set of hypotheses and will be ignored. The data satisfying the hypotheses of the transitivity-property can be summarized as shown in Table V. Of the 56 cases involving money, only one violates transitivity. Of course, this may not be an actual violation because the observed 0 and 1 proportions may not, in fact, signify 0 and 1 probabilities. We conclude that transitivity cannot be rejected for money-gambles.

The picture is less favorable for transitivity in the cigarette data, particularly for S_1 and S_3. It is interesting, however, that all the violations for these two *S*s involve at least one pair of brands for which *S* did not have a consistent preference. As we noted earlier, there were changes over time in the data for these gambles, and so the test of transitivity may not be legitimate.

(4) Stochastic transitivity. Parallel data for a test of strong stochastic transitivity are given in Table VI. The hypotheses underlying such a test were only met in the data for gambles paid off with cigarettes. There are 12 apparent violations in 47 cases. It is not easy to know how to test this assumption because, even if it were true, we would expect some apparent violations to arise from sampling errors. No one has yet worked out the sampling distribution, even for particular assumptions about the dependence of $P(x,z)$ upon $P(x,y)$ and $P(y,z)$.

TABLE V

NUMBER OF VIOLATIONS OF THE ASSUMPTION OF TRANSITIVITY

	Cigarettes		Money	
S	No. cases	No. violations	No. cases	No. violations
1	2	2	13	0
2	0	0	13	1
3	3	3	5	0
4	2	0	4	0
5	13	1	21	0

TABLE VI

NUMBER OF VIOLATIONS OF THE STRONG STOCHASTIC ASSUMPTION OF TRANSITIVITY IN GAMBLING FOR CIGARETTES

S	No. cases	No. violations	*P* under the null hypothesis
1	6	2	.100
2	23	8	.002
3	7	1	.007
4	10	1	.000
5	1	0	.333

We may show, first, that the data exhibit some structure. As a null hypothesis, suppose that each of the six orderings of the three proportions of choices is equally likely, then two of the six will be consistent with strong stochastic transitivity. The binomial probabilities of our observations under this hypothesis are shown in Table VI, and it is clear that we must reject this null hypothesis.

Looking at the violations more carefully, two points seem important. First, of the 12 violations, only 3 (one for S_1 and two for S_2) are violations to the extent of more than two standard deviations of the estimated probabilities. Second, of the 11 violations exhibited by S_1, S_2, and S_3, only one does not involve a pair of brands for which that S was inconsistent. As we have seen earlier, the stability of the probabilities of choice between such gambles is in doubt.

On the basis of these considerations, we conclude that these data are not adequate to reject the strong stochastic transitivity-condition; but we would not claim that the support for it is strong.

(5) Decomposition-axiom. The decomposition-axiom, in the form of Equation [2], is used to estimate the postulated judgment of likelihood of events, $Q(\alpha,\beta)$, from the proportions of choices of gambles, $\hat{P}$. For each

TABLE VII

VALUES OF Q ESTIMATED FROM THE DECOMPOSITION-AXIOM (EQUATION [2])
Events are denoted by Greek letters.

S		Cigarettes			Money		
		β	γ	δ	β	γ	δ
1	α	.970	1.000	.960*	.996	1.000	1.000
	β		.760	.990		.770	.980
	γ			.955			.960
2	α	.363	.567*	.933	.213	.589	.956
	β		.667	.889		.856	.889
	γ			.889†			.960
3	α	.990	.880	.990	.984	.980	.996
	β		.880	.980		1.000	.990
	γ			.965			.988
4	α	.912	.980	.976	.928	.980	.968
	β		.870	.910		.810*	.940
	γ			.832			.852*
5	α	.956†	1.000	.938†	.994	1.000	1.000
	β		.969	.891†		.985	1.000
	γ			.856			.769

* Significant at 5% level. † Significant at 1% level.

outcome pair (a,b) for which we have estimates of the P-terms, Equation [2] gives us an estimate of $Q(\alpha,\beta)$. The means of these estimates are shown in Table VII. If the axiom is correct, the several estimates of each Q should be approximately the same. To test this, we take the over-all mean as the expected value, convert all of the proportions to frequencies, and do a χ^2 test. Only gambles involving money and cigarette brands that were consistently preferred have entered into these calculations.

Of the ten subtables of Table VII, seven do not have an over-all χ^2 significant at the 5% level. The S_1-cigarette table is significant at the 5% level. The S_4-money and the S_5-cigarette tables are significant at the 1% level. It will be recalled that we found some indications of instability and dependence of responses in the choices of S_4 and of S_5 when the gambles are paid off by cigarettes (Table III), in which cases the test of the decomposition-axiom is not suitable. Ignoring the S_4- and S_5-cigarette subtables (with their 1% levels), we note that the remaining eight tables include 48 estimates. Of these, two are significant at the 5% level and one at the 1% level, which is about what one would expect by chance. This evidence appears to support the decomposition-axiom, at least for small payoffs when the preferences are stable over time.

One point of interest is S_2's reversal of the relative likelihood of the events (colors) α and β from the order given by the other *S*s and from their objective probabilities of occurring. It is a little difficult to see how he could have misjudged which event was more likely after hundreds of observations, but he did.

Summary

A gambling experiment employing both money and cigarettes as the outcome was run to test three assumptions now current in a stochastic theory of choice. Five *S*s chose between 74 pairs of gambles in each of from 32 to 50 daily sessions. Probabilities of choice were estimated from these data for each *S*.

Some sequential changes in the choices of the *S*s were established. For one *S*, they appeared about equally in both the cigarette and money data, suggesting that he altered his judgment of the relative likelihood of the events during the experiment. For the other four *S*s, sequential effects appeared mainly in the cigarette data, and they seemed to be closely correlated with changes of brand preferences. It was found that gambles involving inconsistently preferred brands accounted for most of the sequential effects. There was no evidence that these *S*s altered their judgments about the likelihood of the events on which the outcome of the gamble depended.

Ignoring those pairs of gambles involving at least one pair of inconsistently preferred brands, we found strong support for the transitivity-assumption and more modest support for the strong stochastic transitivity- and decomposition-assumptions.