

Action Selection and Likelihood Ratio Estimation by Individuals and Groups¹

BARBARA C. GOODMAN²

University of Michigan

This study investigates the shifts between individual and group performance in a choice dilemma, a gambling, and a Bayesian likelihood ratio estimation task. Twenty seven male subjects performed each task alone. Six four-man leaderless groups were formed and repeated the each task. Three subjects performed the task alone a second time. Finally, all 27 subjects repeated each task again alone. The choice dilemma task decisions reproduced previously found patterns of shifts. Groups preferred higher variance gambles than did the average of pregroup individuals. The post-group likelihood ratio estimates of 22 of the 24 test subjects resembled their group's estimates more closely than they resembled their own pregroup estimates. Both group and individual correlations between measures of performance in all three tasks were low.

INTRODUCTION

Some investigations have shown that groups make riskier decisions than individuals do. Stoner (1961) called this phenomenon "the risky shift." Specifically, if unanimous group decisions are more risky than the average of the initial individual decisions, the group exhibits a risky shift. Brown (1965), Burnstein (1969), and Kogan and Wallach (1967) have reviewed this topic. In most experiments, Ss initially perform the tasks alone, then repeat it in a group, and then repeat it again alone; this is an individual-group-individual (I-G-I) design. A consistent finding in I-G-I studies is that final (postgroup) individual decisions conform more closely to the decisions of the group than to the initial (pregroup) individual decisions.

¹The research reported here was undertaken in the Engineering Psychology Laboratory, Institute of Science and Technology, University of Michigan. The research was supported by the National Aeronautics and Space Administration Grant NGL 23-005-171. This research was submitted to the University of Michigan in partial fulfillment of the requirement for a doctoral degree. The author would like to thank her graduate advisor, Ward Edwards; the other members of her committee, Clyde H. Coombs, Walton M. Hancock, David H. Krantz, and Robert B. Zajonc; and also Cameron R. Peterson for their advice and encouragement.

²Requests for reprints should be sent to Barbara C. Goodman, Engineering Psychology Laboratory, Institute of Science and Technology, University of Michigan, HSRI Building, 2901 Baxter Road, Ann Arbor, Michigan 48105.

The choice dilemma paradigm has been the usual task. In this paradigm, the individual or group acts as an advisor in hypothetical decisions. Only in studies by Lonergan and McClintock (1961), Pruitt and Teger (1969), Rettig (1966), Wallach, Kogan, and Bem (1964) have Ss won or lost money (or experienced other real consequences) as a result of their decisions. Whether risky shifts will occur in real life situations is still an open question.

Thus far differences between individuals and groups in decision making have been studied only in action selection tasks. This study, using a within-S design, examines the differences between individuals and groups in two different action selection tasks and in one inference task. This kind of design permits the exploration of differences in performance between the same individuals and groups within several different tasks. It also permits a study of the interrelationships between performance on the different tasks within individuals and groups.

Bayes's theorem is the optimal rule to use in order to revise one's prior opinions about the states of nature (Edwards, Lindman, and Savage, 1963). In its odds-likelihood ratio form, it expresses the impact of a new datum on prior opinion as a (set of) likelihood ratios (LRs). The main advantage of estimating an LR rather than the separate probabilities or likelihoods is that an LR measures only those features of a datum that are relevant to the hypotheses being considered. Edwards, Phillips, Hays, and Goodman (1968) have proposed a diagnostic system, PIP, based on LR estimates.

An information processor is called conservative when his posterior opinion differs from his prior opinion less than Bayes's theorem prescribes. A great deal of research has shown that in certain situations people are conservative. Reviews of this literature can be found in DuCharme (1969), Edwards (1968), and Peterson and Beach (1967). The issue remains: Are groups more or less conservative than individuals in an inference task? In the foregoing context, this study considers three questions: (1) Will a risky shift occur in an action selection task in which the participants bet their own money? (2) Does individual or group behavior in several tasks that might all be supposed to require responses to "risk" show the kind of coherence that would permit the assumption of a general risk-seeking or risk-avoid trait? (3) Are individuals more or less conservative than groups?

METHOD

Subjects

Subjects were 27 male volunteers, 17 graduate students and 10 undergraduates from several disciplines, but none were psychology majors.

Subjects were paid \$30 plus their winnings or minus their losses from the gambling task.

Tasks

Three tasks were used in order to contrast decisions of individuals and groups in action selection and LR estimation, and to search for a possible underlying phenomenon of risk preference.

1. *Choice dilemma task.* I selected the same twelve choice dilemma situations as used by Stoner (1968) to determine a *S*'s (or group's) preferences in hypothetical situations. In each problem, an imaginary person must choose between two acts. The risky action, if successful, produces a high payoff, and if unsuccessful, a low one. The sure thing produces an intermediate payoff. The subject must "advise" the imaginary decision maker what the lowest probability of success is for which the risky act should be chosen. Stoner (1968) summarizes the twelve situations.

2. *Likelihood ratio estimation task.* I used a task devised by Ward Edwards and Gloria Wheeler (unpublished) to compare *S*s' and groups' LR judgments. This task uses seven-inch sticks. Each stick has a blue and a yellow part. The blue and yellow occur in various proportions in different sticks. The *S*s are shown two blue and yellow colored drawings, each of a sample of approximately 100 sticks from one of the populations being considered. Populations A and B have Gaussian (normal) distributions with mean lengths of blue of 4.5 inches and 2.5 inches, respectively. Each population has a standard deviation of 1.25 inches of blue. To prepare the drawings the cumulative normal distribution for each population was divided into 100 equally likely parts. 1/100 of the population falls between every two division or boundary points. The mean lengths of blue at the boundary points, randomly rearranged, comprise the drawing that displays each population. Three lengths are missing from each drawing because it was too difficult to produce the very small amounts of color required to represent the boundary points at the tails of the distribution. Thus each drawing is a picture of a random arrangement of what looks like a sample from a normal distribution but in fact has been chosen so that the "sample" precisely depicts the population. A line is drawn across each drawing showing the population mean. The *S*s are told that each drawing shows a random sample of sticks from its population, even though this was not the case.

After suitable instruction, *S* sees a partly blue-partly yellow stick. With the drawings representing the two populations as guides, he is asked to state from which population (A or B) this stick is more likely to have been sampled and, in a ratio form, how much more likely. This is a likelihood ratio estimate. It is repeated with 80 different sticks.

TABLE 1
THE GAMBLES COMPRISING THE GAMBLING GAME^a

<i>S</i> (5)	<i>S</i> (10)	<i>S</i> (15)	<i>S</i> (20)	<i>S</i> (25)
(\$1, -90¢)	(\$1, -80¢)	(\$1.50, -\$1.20)	(\$2, -1.60)	(\$2.50, -\$2)
(50¢, -40¢)	(60¢, -40¢)	(\$1, -70¢)	(\$1, -60¢)	(\$1.50, -\$1)
(30¢, -20¢)	(50¢, -30)	(70¢, -40¢)	(80¢, -40¢)	(\$1, -50¢)
(25¢, -15¢)	(40¢, -20¢)	(50¢, -20¢)	(60¢, -20¢)	(80¢, -30¢)
(20¢, -10¢)	(30¢, -10¢)	(40¢, -10¢)	(50¢, -10¢)	(65¢, -15¢)
(15¢, -5¢)	(25¢, -5¢)	(35¢, -5¢)	(45¢, -5¢)	(55¢, -5¢)
(5¢, 5¢)	(10¢, 10¢)	(15¢, 15¢)	(20¢, 20¢)	(25¢, 25¢)

^a The flip of a coin determines the outcome of a gamble (heads wins, tails loses). The first number of each pair is the amount to be won if that gamble is played and a head comes up. The second number of each pair is the amount to be lost.

3. *Gambling task.* I selected the gambles designed by Coombs and Huang (1970) to determine a *S*'s (or groups') preference in a gambling situation. *S*s or groups are shown, one set at a time, the five sets of seven gambles displayed in Table 1. They are asked to rank order from most to least preferred the seven gambles within each set. Following this, the most preferred choice from one of the sets is randomly selected and played. The flip of a coin determines the outcome. The *S*s are given money to wager, but they could lose more than they are given.

Procedure

The first three of the nine sessions of the experiment were individual; the second three were group; the last three were individual again. In sessions 1, 4, and 7, *S*s performed the choice dilemma task. In sessions 2, 5, and 8, they performed the LR task. And in sessions 3, 6, and 9, they performed the gambling task. *S*s worked at their own pace, and were not given prior information about the I-G-I design.

Twenty four test *S*s did the tasks alone and in groups. These test *S*s were randomly assigned into four-man, leadership groups. Three *S*s were individual controls, performing each task three times alone. There were no group controls.

The groups were formed after the initial individual sessions and remained unchanged. Each group sat around a desk. There was no fixed seating arrangement. Group sessions were tape recorded with *S*s' knowledge. *E* remained present and answered only procedural questions. Unanimous decisions were required.

For the choice dilemma task, the pre- and postgroup individual decisions were made outside the laboratory. Subjects were given the problems with instruction not to discuss the material. The three control

Ss performed the first and third sets of problems at leisure, and the second performance in the laboratory. For each problem, *S* checked the lowest probability for which he would advise taking the higher risk alternative. During the group session, each *S* received a booklet containing the choice dilemma problems. They discussed each problem until they reached a unanimous decision. Each man recorded the group decisions in his booklet. Three weeks intervened between successive choice dilemma sessions.

Each *S* was individually trained to estimate LRs. This instruction typically took 30 minutes to an hour. The training consisted of a written explanation, with examples, of the inferential process, followed by actual example tasks, with feedback of the correct values. In the last example task the *S* estimated LRs for five data which might have been sampled from either of two Gaussian (normal) distributions. One distribution was the heights, in stocking feet, of an empirical sample of American men between the ages of 18 and 79; the other was the same for American women. For each datum the *S* first estimated whether that height was more likely to have been sampled from the population of men, or the population of women. Then he estimated how much more likely in a ratio sense. The experimenter presented all five data before giving any feedback. The five heights were presented in the following order: 5'2", 6'1", 5'6", 5'8", and 4'10". The corresponding correct LRs are 1:10, 725:1, 1.3:1, 7.5:1, and 1:120. I explained to *S* the specific task using blue and yellow sticks when I felt satisfied that he understood how to estimate LRs.

The LR estimation procedure was as follows. The *S* was shown 80 sticks, one at a time, in a special holder that presented the entire length of blue and fellow against a white background. The subject decided whether a stick was more likely to have been sampled from the predominantly blue or the predominantly yellow population. Then he decided how much more likely in a ratio sense. The subject recorded both responses (qualitative and quantitative) on a sheet of paper. When *S* turned the paper over, *E* removed that stick from the holder, placed it out of sight, and put the next stick into the holder. The order in which the sticks were presented *never* changed. The individual and group sessions were performed in the same manner. The groups discussed each stimulus. When a unanimous decision was reached, each person recorded the value on a sheet of paper. Three weeks intervened between the pre-group individual sessions and the group session and two weeks between the group session and the postgroup individual sessions. The same relative spacing occurred for the individual sessions of the control Ss.

A brief explanation of the gambling task format preceded the initial

individual sessions. All the individual gambling sessions followed the same procedure. Each *S* received \$1.50 when he first arrived. He was told that one gamble would be played at the end of the session, and that this would be selected at random from his most preferred choices among the different sets of gambles. He was given, however, the option of selecting his second most preferred choice, if the possible loss from his first choice was greater than \$1.50. Then *S* was handed five decks of IBM cards, with one gamble printed on each card. The random order within a deck was held constant throughout the experiment. The decks were given one at a time in order of increasing wager. Within each deck, *S* was asked to rank order the gambles from most to least preferred. When *S* finished the rank orderings, a gamble was randomly selected, played, and the appropriate payoff was made.

At the beginning of the group sessions each member received \$1.50. The *E* explained that the one gamble to be played at the end of the experiment would be selected in the same manner as it had been in the individual sessions. Each person stood to win or lose the full amount of the gamble finally selected. The sets of gambles were given one at a time to each person. When the group arrived at a unanimous preference ranking, each member arranged his cards according to the group decision and then turned these cards over to *E*, who handed out the next deck of cards. When groups finished the rank orderings, a gamble was randomly selected, played, and the appropriate payoffs made. Two of the six groups consisted of three *S*s instead of four. The two *S*s who did not attend their respective group sessions served as additional control *S*s for this task. Three weeks intervened between the pregroup individual and group sessions. There was a two-week interval between the group and postgroup individual sessions. This same spacing was maintained for the control *S*s.

RESULTS

Changes Between Pregroup Individual Performance and Group Performance

The Choice Dilemma Task

Both risky and cautious shifts occurred. A risky shift is a positive difference resulting from the subtraction of the group Risk Score from the pregroup average individual Risk Score. A negative difference in the same subtraction defines a cautious shift. Table 2 summarizes the mean shifts in the performance of the six four-man groups compared to pregroup individual performance.

TABLE 2
SUMMARY OF THE GROUP SHIFTS ON THE CHOICE DILEMMA PROBLEMS

Problem number	Mean [of the pregroup average individual] Risk Scores ^a	Mean group Risk Scores	Mean shift	<i>t</i>	Number of groups showing risky shift	Number of groups showing cautious shift
1	6.41	5.50	.91	1.22	4	2
2	5.62	6.00	-.38	.34	3	3
3	3.17	1.67	1.50	2.51	5	1
4	3.67	3.17	.50	.69	3	3
5	2.83	2.67	.16	.33	3	3
6	4.54	3.83	.71	.85	3	3
7	4.12	3.67	.45	.86	4	2
8	8.29	9.33	-1.04	2.87*	1	5
9	2.67	1.17	1.50	3.05*	5	1
10	3.33	3.00	.33	.39	2	3
11	6.08	5.83	.25	.37	3	3
12	6.00	5.83	.17	.20	3	3
Total					39	32

^a Risk Score equals probability of success multiplied by 10.0.

* $p < .05$ two-tail t test.

Stoner (1968) found, for the same task with a larger sample, 30-33 groups of 4-7 members, statistically significant risky shifts in problems 3-5, 7, 9, and 10, and statistically significant cautious shifts in problems 2 and 8. The mean pregroup individual Risk Scores obtained in this study resembled his. The Pearson product-moment correlation between these scores, listed in Table 2, and the mean Risk Scores in his study was .95. Moreover, the direction of the shifts coincided with Stoner's findings, even though most of the shifts in this study were not statistically significant. The Pearson-product-moment correlation between the twelve mean shifts, across groups, in the two studies was .84. Thus there is excellent qualitative agreement between these two sets of results.

The Gambling Task

The riskiest gamble, the one with the greatest possible loss within each set, is scored 1, the next 2, etc., until 7. The number corresponding to the *S*'s most preferred choice in a set is called the Level of Risk Score (LOR) for that set. A risky shift is a positive difference resulting from the subtraction of the group LOR from the pregroup average individual LOR. A negative difference in the same subtraction defines a cautious shift. Table 3 summarizes the mean shifts in the performance of the

TABLE 3
SUMMARY OF THE MEAN LEVEL OF RISK SCORES AND THE GROUP SHIFTS FOR
THE GAMBLING TASK

Set	Mean [of the pregroup average individual] Level of Risk score ^a	Mean group Level of Risk score	Mean shift	<i>t</i>	Number of groups showing risky shift	Number of groups showing cautious shift
<i>S</i> (5)	4.08	2.83	1.25	1.29	5	1
<i>S</i> (10)	3.92	3.17	.75	.69	3	2
<i>S</i> (15)	3.95	2.50	1.45	1.68	4	2
<i>S</i> (20)	4.03	2.00	2.03	3.02*	5	1
<i>S</i> (25)	4.36	2.67	1.69	2.91*	5	1

^a Two of the six groups have 3 members each, the other four groups have four members each.

* $p < .05$, two-tail t test.

groups. Although both risky and cautious shifts were found in all sets of gambles, the mean shift was risky. Considering only the four-member groups, there were 18 instances of risky shift, one instance of a cautious shift, and one instance of no shift.

*The Likelihood Ratio Estimation Task*³

The Group Conservation Score (GCS) is the slope of the regression line relating the group log likelihood ratios (LLR) to the pregroup average individuals LLRs where the individual scores are the predictor variable. This value is a quantitative measure of conservatism between groups and individuals. If it is greater than 1, the group is less conservative than the individual; and if it is less than 1, the group is more conservative than the individual.

Although each individual or group estimated 80 LRs per session, 79

³ Analyses for LR and odds responses use logarithmic transformations of the data because the logarithmic transformation of Bayes's theorem results in an equation in which the log likelihood ratio (LLR) is added to the logarithm of the prior odds to obtain the logarithm of the posterior odds. Thus information of a given diagnosticity, LR, on a log scale changes the log of the prior odds a constant amount, irrespective of what the log of the prior odds may be. Neither the LR scale, nor the odds scale, nor the probability scale has this property. Therefore, analyses using untransformed odds and LR scales would result in changes in prior odds, from an application of Bayes's theorem, being dependent upon the size of the prior odds.

TABLE 4
GROUP CONSERVATISM SCORES AND CORRESPONDING CORRELATION COEFFICIENTS
FOR THE LR ESTIMATION TASK

Group	Correlations between average pregroup LLRs and group LLRs	Group Conservatism Scores	t^a
1	.986	.922	4.460**
2	.945	.882	3.394**
3	.890	.673	8.322**
4	.952	.470	30.966**
5	.983	1.071	3.147*
6	.975	1.320	9.283**

^a Test of the null hypothesis that the regression slope equals 1.000.

** $p < .01$, two-tail t test.

*** $p < .002$, two-tail t test.

are shown in all the data analyses. The one response which was eliminated was for a nearly all blue stick which frequently generated an extreme LR far out of the range of the other 79.

The GCSs and corresponding correlations between the average pregroup estimates and group estimates for this task are shown in Table 4.

Since the correlation coefficients were high, the relation between group values and the average pregroup individual values can be represented by a linear approximation. The GCSs indicate the difference between group and individual performance. Four of the six groups were more conservative than the average of the individuals within those groups.

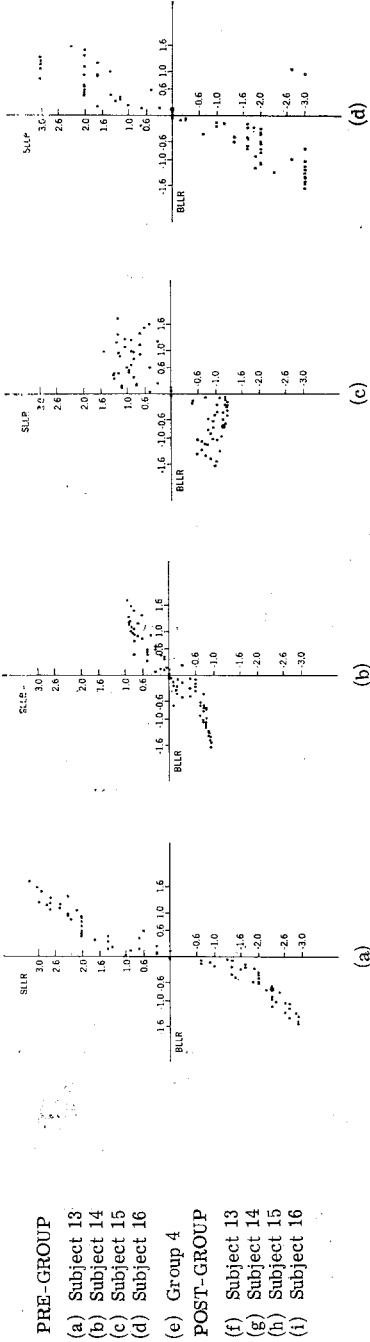
Conformity

Likelihood Ratio Estimation Task

This task is ambiguous since Ss have insufficient information to permit calculation of the Bayesian values. The SLLR-BLLR scatterplot compares subjective performance with calculated values, where SLLR is the subjective log likelihood ratio value and BLLR is the Bayesian log likelihood ratio value. This plot was made for each individual and group session. An example of a set of graphs for one group is shown in Fig. 1 where (a)–(d) are the pregroup individual graphs, (e) is the group graph, and (f)–(i) are the postgroup graphs.

For all groups, the postgroup individual estimates showed striking conformity to those of the group regardless of the individual's pregroup performance or that of the group itself. There were noticeable changes in the shape of the function relating the SLLR and BLER estimates in

PRE-GROUP INDIVIDUAL SESSIONS



POST-GROUP INDIVIDUAL SESSIONS

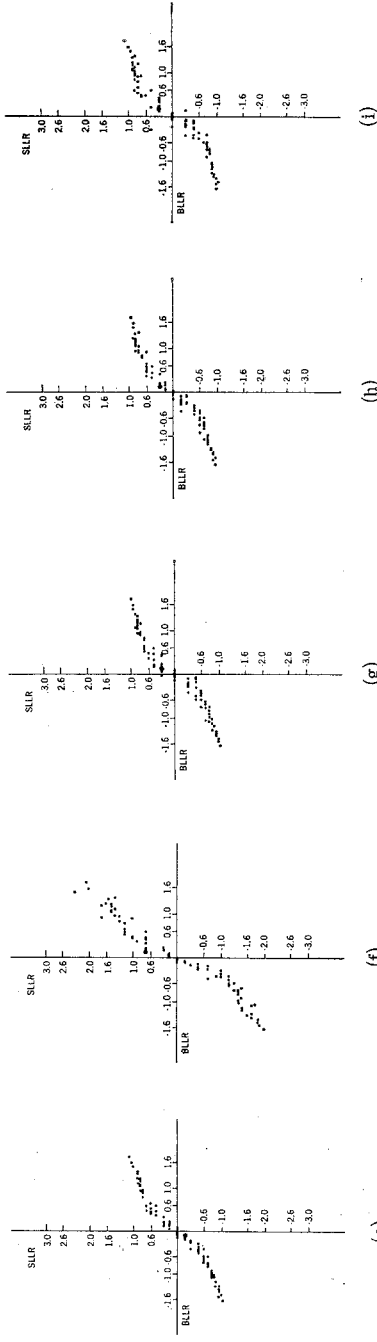


FIG. 1. Log Likelihood Ratio Scatterplots. Logarithms of the estimated likelihood ratios as a function of the logarithms of the Bayesian likelihood ratios for Group 4 subjects. In the top row are the plots for the pre-group individual sessions. In the bottom row are the group consensus plot and the post-group individual session plots.

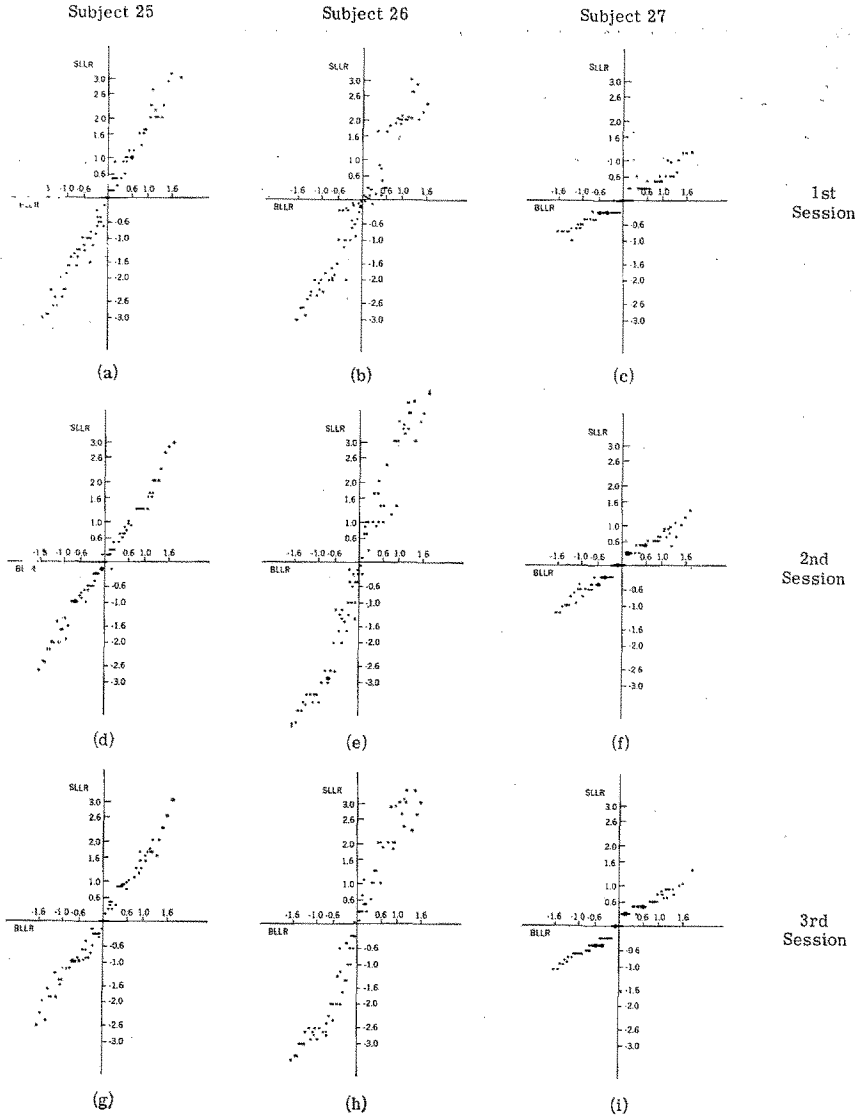


FIG. 2. Log Likelihood Ratio Scatterplots. Logarithms of the estimated likelihood ratios as a function of the logarithms of the Bayesian likelihood ratios for the three control S 's. In the top row is the plot for each S 's first session. In the middle row are the plots for the second sessions. The last row contains the third session plot for each control S .

the range of SLLR values estimated and in the variance of the individual estimates.

The graphs in Fig. 2 are the scatterplots for the three control Ss. Graph sets (a)–(c), (d)–(f); and (g)–(i) are the SLLR-BLLR plots for the 1st, 2nd, and 3rd individual sessions, respectively. Graph sets (a), (d), and (g); (b), (e), and (h); (c), (f), and (i) each represent one S.

The shape of the functions and the range of estimates for each control S were similar from session to session.

The degree of this conformity is further illustrated by the Sum of the

TABLE 5
SUMS OF SQUARES OF THE DIFFERENCES BETWEEN FINAL AND INITIAL LLRs,
AND BETWEEN FINAL AND GROUP (OR 2nd SESSION) LLRs ACROSS THE 79
STICKS; AND THE RATIOS OF THESE TWO SUMS OF SQUARES

Group number	Subject number	SSD Between final and initial LLRs	SSD between final and group (or 2nd Session)	Ratio of final, initial SSD to final, group (or 2nd session) SSD
1	1	12.02	2.60	4.63
1	2	31.79	10.72	2.97
1	3	37.08	4.02	9.22
1	4	91.70	17.62	5.20
2	5	10.76	12.45	0.87
2	6	22.18	14.78	1.50
2	7	81.88	24.61	3.33
2	8	40.41	7.63	5.30
3	9	19.20	69.47	0.28
3	10	136.82	1.65	82.96
3	11	23.08	4.44	5.19
3	12	80.36	71.12	1.13
4	13	59.36	27.18	2.18
4	14	3.06	1.34	2.28
4	15	20.52	0.81	25.20
4	16	205.62	6.13	33.53
5	17	143.03	10.84	13.20
5	18	9.67	5.21	1.86
5	19	17.49	3.00	5.84
5	20	89.97	6.86	13.12
6	21	32.26	4.19	7.70
6	22	28.98	8.88	3.27
6	23	37.48	7.24	5.18
6	24	20.08	13.46	1.49
Control	25	12.89	3.57	3.59
Control	26	38.68	25.51	1.52
Control	27	2.81	0.94	2.98

Squares of the Difference (SSD) between the postgroup individual and the group LLRs and also between the postgroup individual and the pregroup individual LLRs across the 79 sticks. The SSDs between the final and initial session LLRs, and between the final and group (or 2nd) session LLRs, as well as the ratios of the two values are shown in Table 5. The greater the value of the ratios shown in Table 5 for the test Ss, the greater the individual acceptance of the group rule. This ratio was larger than one for 22 of the 24 individuals, and was larger than two for 18. The ratios for the three control Ss indicate that their estimates converged on their own standard in the absence of an external source of information or group norm. The SSDs between the final and initial LLRs clarify the picture even more. These SSDs for the two control Ss with high ratios (#25, #27) are smaller than the corresponding SSDs for almost all of the test Ss.

TABLE 6
BETWEEN SESSION ABSOLUTE DIFFERENCES OF THE RISK SCORE SUM FOR
TOTAL PROBLEM SET (RST)

Subject number	Absolute difference between final and initial RST	Absolute difference between final and group (or 2nd session) RST	Subject number	Absolute difference between final and initial RST	Absolute difference between final and group (or 2nd session) RST
1	0.9 ^a	0.1	15	0.6	1.2
2	2.3 ^a	1.6	16	2.8 ^a	0.5
3	0.8 ^a	0.6	17	3.1 ^a	0.7
4	1.0 ^a	0.2	18	0.5 ^a	0.4
5	1.4 ^a	0.7	19	0.6 ^a	0.2
6	0.8 ^a	0.7	20	1.7 ^a	1.3
7	0.7	1.4	21	0.7	1.0
8	0.8 ^a	0.1	22	0.4	0.9
9	0.8 ^a	0.4	23	0.9 ^a	0.1
10	1.4 ^a	0.6	24	0.8 ^a	0.1
11	1.4 ^a	0.9	Control Ss		
12	0.2	1.5	25	2.0 ^b	0.0
13	0.6	1.4	26	1.1 ^b	0.5
14	0.2	1.4	27	1.4 ^b	0.8

^a Signifies that the Postgroup minus Pregroup absolute difference is the larger difference, i.e., instances showing conformity.

^b Signifies that the 3rd Session minus 1st Session absolute difference is the larger difference.

Action Selection Tasks

An RST is the sum of the Risk Scores across all 12 choice dilemma problems. The smaller the value, the riskier is the response. Table 6 summarizes the absolute differences between the final and initial session RSTs and between the final and group (or 2nd) session RSTs for the choice dilemma task. Seventeen of the 24 test *Ss* showed conformity in this task, whereas 22 of the 24 *Ss* showed conformity in the LR task. Therefore, there is a trend toward a greater acceptance of the group decisions in the LR task than in the choice dilemma task. As in the LR estimation task, the third session individual values for the three control *Ss* resembled the second session values more closely than they did the first session values.

An ALOR is the average of the five Level of Risk (LOR) Scores

TABLE 7
BETWEEN SESSION ABSOLUTE DIFFERENCES OF THE AVERAGE LEVEL OF
RISK (ALOR)

Subject number	Absolute difference between final and initial ALOR	Absolute difference between final and group (or 2nd session) ALOR	Subject number	Absolute difference between final and initial ALOR	Absolute difference between final and group (or 2nd session) ALOR
1	0	1.2	16	4.6 ^a	0.4
2	2.4	4.0	17	0.4 ^a	0.2
3	2.8 ^a	0.6	18	5.6 ^a	0.2
4	2.6 ^a	0.0	19	4.4 ^a	0.2
5	4.4 ^a	2.2	20	0.0	0.2
6	0.0	2.2	21	3.6	5.0
8	0.4	0.8	22	0.8 ^a	0.4
9	0.0	1.2	23	0.2 ^a	0.0
10	0.0	4.8	Control <i>Ss</i>		
11	3.8 ^a	0.0	7 ^b	5.2 ^c	0.0
12	0.0	1.2	24 ^b	0.2 ^c	0.0
13	4.8 ^a	0.2	25	0.4	0.4
14	0.0	5.0	26	2.4 ^c	0.6
15	0.4	0.4	27	1.2 ^c	0.4

^a Signifies that the postgroup minus pregroup absolute difference is the larger difference, i.e., instances showing conformity.

^b This *S* did not participate in the group session for this task. *S* served as an additional control *S* for this task.

^c Signifies that the 3rd session minus 1st session absolute difference is the larger difference.

across the five different sets of gambles. Table 7 presents the absolute differences between the final and initial session ALORS and between the final and group (or 2nd) sessions ALORs for the gambling task. Eleven of the 22 test Ss showed conformity in this task. Individual acceptance of the group decision is least in the gambling task. The third session performance of four control Ss, out of five, closely resembled their second performance more than their first.

Correlations Between Tasks Within Individual and Groups

Action Selection Tasks (Choice Dilemma and Gambling Tasks)

Many dependent variables are possible when scoring an individual or group performance of choice dilemma problems. Five nonorthogonal dependent variables were studied. Individuals and groups, separately, were rank ordered with respect to risk on each of the five variables. Individuals and groups were also rank ordered with respect to risk according to their ALOR on the gambling task. There was little correlation between performance in these two action selection tasks as measured by the Spearman Rank Order Correlations.

Likelihood Ratio Estimation Task

Two nonorthogonal dependent variables were studied. One was the Likelihood Ratio Slope (LRS). It is the slope of the regression line relating subjective log likelihood ratios (SLLR) to the corresponding Bayesian log likelihood ratios (BLLR). The greater the LRS the less conservative are the subjective estimates. The other variable measured the amount of deviation from the correct values. It was the difference of the final cumulative odds for the estimated and calculated values resulting from the particular sample of 79 sticks used.

Individuals and groups, separately, were rank ordered on both of the above variables, with respect to conservatism in the case of the LRS and with respect to the extent of the difference from the Bayesian value in the case of the other variable. Spearman Rank Order Correlations were calculated between the variables in the LR estimation task and those of the action selection tasks. There is no evidence for a substantive relationship between performance in the different tasks.

Comparison-Individual And Group Subjective LR Estimates Compared to Bayesian Values

It is possible to grade performance when a task has objective criteria. The LR estimation task is such a task. Behavior is conservative when

people do not extract as much certainty from new information as is available within the data. The LRS is the dependent variable used to measure conservatism. Behavior is conservative when the LRS is less than 1, and radical when LRS exceeds 1.

The pregroup individual LRSs ranged from .519 to 3.364 with a mean of 1.546 and a standard deviation of .640. The group LRSs were less extreme. They varied from .761 to 1.895 with a mean of 1.457 and a standard deviation of .383. The mean postgroup individual LRS of 1.488 was very close to the other mean LRSs. The corresponding standard deviation was .358, while the individual LRSs ranged from .720 to 2.072. The average correlation coefficient in all three cases was at least .930.

The regression statistics for the three control subjects were as follows: correlation coefficients of at least .950 in all cases but one; first session LRSs of 1.878, 2.006, and .599; corresponding second session LRSs of 1.563, 2.890, and .714; and corresponding third session LRSs of 1.498, 2.434, and .707.

Several conclusions can be drawn from these analyses. First, the high correlations show good qualitative agreement between the estimated LRs and the Bayesian values, i.e., *qualitatively* the individuals and groups performed the task well. Secondly, most groups and individuals were not conservative, since most LRSs exceeded 1. They tended to err in the radical direction, extracting too much certainty from data presented.

DISCUSSION

Although the results of this administration of the choice dilemma task are not statistically significant, perhaps because a small number of groups were used, they agree with those of Stoner (1968). The repetition of Stoner's findings confirmed the reproducibility of the risky shift under conditions of this study.

The gambling task used in this investigation is the only experiment I know of which includes all of the following conditions: an I-G-I design, unanimous group decisions, and Ss experiencing the consequences of their decisions in all sessions. A risky shift was found in each set of gambles. However, the risky shift is a measure based on average deviation. And, as in the choice dilemma problems, there were divergent groups showing less risk than the average pregroup individual. In any event, this study demonstrates that a risky shift can occur in an action selection task in which Ss wager their own money. Thus it is consistent with the findings of Pruitt and Teger (1969).

Slovic (1962) found low intercorrelations within individuals between different risk taking measures. Similarly, the present study also found

small rank order correlations within individuals (or groups) between two action selection tasks, each with a nonambiguous definition of greater risk. Presumably these two tasks were measuring different behavioral characteristics. At the present time, there is no single, generally acceptable definition of risk. Moreover, there is no experimental evidence of a general risk-seeking or risk-averse behavioral trait.

Likelihood ratio estimation and action selection are two different processes and produce unrelated behavior within the same individuals and groups. The low correlations between performance in the inference task and performance in the action selection tasks support the hypothesis that different processes and different individual and group characteristics might be involved in the performance of the two kinds of tasks.

A striking result of this study was the extent of the conformity in the LR estimation task. This conformity, two weeks after the group session, occurred irrespective of the group performance and of the *S*'s pregroup individual estimates. What caused it? I believe there were two main factors. First of all, the group had to evolve and specify a *modus operandi* in order to generate unanimous estimates for the 80 sticks. Having established an explicit rule, they remembered it. Secondly, *Ss* accepted the group rule and were satisfied with it. The informational social influence defined by Deutsch and Gerard (1955) as "an influence to accept information obtained from another as evidence about reality" appears to have been a major influence towards conformity.

Different strategies were used to arrive at an LR estimation rule. Common to all of these strategies was, first, a sequential appraisal of individual opinion. Then, relevant information was discussed. All groups but one realized that it was not possible to calculate the true scale with only the information available. Moreover, all groups determined that consistency was the primary task requirement. All rules reflected this, so that if a second stick was more blue than the first, then it was more likely for the second stick to have come from the predominantly blue population.

All individuals, prior to the group session, had realized that the appropriate LR for the stick with equal amounts of blue and yellow was one. Moreover, each person had realized the symmetric nature of the task. Once a scale value was decided upon for a particular stick in one population, this same LR value was applied to the corresponding stick in the other population.

Five of the six groups decided on their scale by setting several reference points and interpolating from there. The reference points for three groups were determined by deciding upon specific values for the first

few sticks that were presented. In this case, the order of presentation of the sticks was probably important. The other two groups set up their scales by determining LR values for the average length of blue, $4\frac{1}{2}$ ", and for the extreme length.

The sixth group used a different approach for determining their scale. Their rule went as follows: Divide the ratio of the length of blue to the length of yellow by the ratio of length of yellow to the length of blue, if blue is the predominant color. Reverse the lengths, if yellow is the predominant color. This rule results in the following transformation: If x is the estimated length of the predominant color and y is the other length, then $LR = x^2/y^2$. For the wrong reason, this rule results in the correct responses for most of the sample.

In addition to each group evolving a specific procedure, the rules were enforced to different degrees in different groups. If a member of the group proposed an estimate clearly contrary to the rules, the other members questioned his evaluation of the amount of blue on the stick and/or his selection of the appropriate LR for that amount of blue.

Conformity was greater in the LR estimation task than in either the choice dilemma or the gambling task. Presumably, *Ss* were more uncertain about their own responses in this task than in the other tasks, since the most critical task variable influencing the extent of conformity behavior seems to be *Ss*' uncertainty about the correctness of his own response (Boomer, 1959; Kelley & Lamb, 1957; Seaborne, 1962; Sherif, 1935; Suppes & Schlag-Rey, 1962; Wiener, 1958; Zajonc & Morrisette, 1960).

One implication of these results for studies in LR estimation is that if you want a task which requires that *Ss* have confidence in the accuracy of their estimates, then don't use this data generating process with the present displays.

Qualitatively, the *Ss* did a good job of performing the inference task. They understood the nature of the process. They were consistent. Quantitatively, the *Ss* did not do very well. Most *Ss* were nonconservative. Most groups were nonconservative. This nonconservatism may have resulted from the feedback given in the training procedure, or there might be something about this data generating process which encourages people to extract more certainty from information than they should.

Thus if accuracy is an important requirement in an inference task of this kind, it is important that *Ss* learn the appropriate scale.

Four groups were more conservative than was the average individual within that group on the pregroup session; the other two groups were not. Therefore, it is not possible to conclude from this study whether groups are more or less conservative than individuals.

There are several implications of these results with regard to man's general capability as an intuitive statistician. First of all, when the data generating process is well specified and orderly, people can perceive the relative diagnosticity of different data extremely well. Secondly, untrained people are not religious about the scale they are using. They can easily be persuaded to use a different one. Perhaps conservatism as a general phenomenon can be extinguished with a very few minutes of appropriate training.

This study, like many, answers some questions, while raising new ones. If *Ss* were very uncertain about the correctness of their own responses in an LR estimation task, would it be possible to get a conformity effect without the necessity of the group interaction? A testable hypothesis is that the conformity behavior found in the particular inference task used in this study is caused solely by cognitive factors. Suppose one makes available to an *S*, after he has performed the task alone, an authoritative rule consistent with his intuitions about the nature of the task. Then the *S* uses this rule in a repetition of the task. Two weeks later, the *S* repeats the task again. Will he have internalized the rule given to him?

How should we train subjects to translate their intuitions about LRs onto a numerical scale, so that they understand the units of measurement, and consequently have confidence in the transformation of their subjective judgments into numbers? What training methods, displays, and responses modes will be useful?

REFERENCES

- BOOMER, D. S. Subjective certainty and resistance to change. *Journal of Abnormal and Social Psychology*, 1959, **53**, 323-328.
- BROWN, R. *Social psychology*. New York: Free Press, 1965.
- BURNSTEIN, E. An analysis of group decisions involving risk (the "risky shift"). *Human Relations*, 1969, **22**, 381-395.
- CARTWRIGHT, D. & ZANDER, A. (Eds.) *Group dynamics*. New York: Harper & Row, 1960.
- COLLINS, B. E. & GUETZKOW, H. *A social psychology of group processes for decision-making*. New York: Wiley, 1964.
- COOMBS, C. H. & HUANG, L. C. Tests of a portfolio theory of risk performance. *Journal of Experimental Psychology*, 1970, **85**(1), 23-29.
- COOMBS, C. H. & MEYER, D. E. Risk preference in coin-toss games. *Michigan Mathematical Psychology Program Technical Report MMPP 68-1*. Ann Arbor, Mich.: University of Michigan, 1968.
- DEUTSCH, M. & GERARD, H. B. A study of normative and informational social influences upon individual judgment. *Journal of Abnormal and Social Psychology*, 1955, **51**, 629-636.
- DUCHARME, W. M. A review and analysis of the phenomenon of conservatism in human inference. *Applied Mathematics and Systems Theory Technical Report, #46-15*. Houston, Texas: Rice University, 1969.

- EDWARDS, W. Conservatism in human information processing. In B. Kleinmuntz (Ed.) *Formal representation of human judgment*. New York: Wiley, 1968. Pp. 17-52.
- EDWARDS, W., LINDMAN, H., & SAVAGE, L. J. Bayesian statistical inference for psychological research. *Psychological Review*, 1963, **70**(3), 193-242.
- EDWARDS, W., PHILLIPS, L. D., HAYS, W. L., & GOODMAN, B. C. Probabilistic information processing systems: Design and evaluation. *IEEE Transactions on Systems Science and Cybernetics*, 1968, **SSC-4**, 248-265.
- KELLEY, H. H. & LAMB, T. W. Certainty of judgment and resistance to social influence. *Journal of Abnormal and Social Psychology*, 1957, **55**, 137-139.
- KOGAN, N. & WALLACH, M. A. *Risk taking: A study in cognition and personality*. New York: Holt, Rinehart and Winston, 1964.
- KOGAN, N. & WALLACH, M. A. Risk taking as a function of the situation, the person, and the group. In *New directions in psychology III*. New York: Holt, Rinehart and Winston, 1967, 111-278.
- LONERGAN, B. G. & McCLINTOCK, C. G. Effects of group membership on risk-taking behavior. *Psychological Reports*, 1961, **8**, 447-455.
- LORGE, I., FOX, D., DAVITZ, J., & BRENNER, M. A survey of studies contrasting the quality of group performance and individual performance, 1920-1957. *Psychological Bulletin*, 1958, **55**, 337-372.
- NORDHØY, F. Group interaction in decision-making under risk. Unpublished Masters thesis, Massachusetts Institute of Technology, School of Industrial Management, 1962.
- PETERSON, C. R. & BEACH, L. R. Man as an intuitive statistician. *Psychological Bulletin*, 1967, **68**, 29-46.
- POLLATSEK, A. & TVERSKY, A. A theory of risk. *Michigan Mathematical Psychology Program Technical Report MMPP 69-6*. Ann Arbor, Mich.: University of Michigan, 1969.
- PRUITT, D. G. & TEGER, A. I. The risky shift in group betting. *Journal of Experimental Psychology*, 1969, **5**, 115-126.
- RETTIG, S. Ethical risk taking in group and individual conditions. *Journal of Personality and Social Psychology*, 1966, **3**, 648-654.
- SEABORNE, A. E. M. Group influence on the perception of ambiguous stimuli. *British Journal of Psychology*, 1962, **53**, 287-298.
- SHERIF, M. A study of some social factors in perception. *Archives of Psychology*, 1935, **9**, No. 187.
- SLOVIC, P. Convergent validation of risk taking measures. *Journal of Abnormal and Social Psychology*, 1962, **65**, 68-71.
- SLOVIC, P. The assessment of risk taking behavior. *Psychological Bulletin*, 1964, **61**, 220-223.
- STONER, J. A. F. A comparison of individual and group decisions involving risk. Unpublished Masters thesis, Cambridge, Mass.: Massachusetts Institute of Technology, School of Industrial Management, 1961.
- STONER, J. A. F. Risky and cautious shifts in group decisions: The influence of widely held values. *Journal of Experimental Social Psychology*, 1968, **4**, 442-459.
- SUPPES, P. & SCHLAG-REY, M. Analysis of social conformity in terms of generalized conditioning models. In J. Criswell, H. Solomon, and P. Suppes. (Eds.) *Mathematical methods in small group processes*. Stanford, Calif.: Stanford University Press, 1962. Pp. 334-361.
- WALLACH, M. A., KOGAN, N., & BEM, D. J. Diffusion of responsibility and level of

risk taking in groups. *Journal of Abnormal and Social Psychology*, 1964, **68**, 263-274.

WIENER, M. Certainty of Judgment as a variable in conformity behavior. *Journal of Social Psychology*, 1958, **48**, 257-263.

ZAJONC, R. B. Conformity. In *International Encyclopedia of the Social Sciences*. New York: Crowell-Collier, 1968. Pp. 253-260.

ZAJONC, R. B. & MORRISSETTE, J. The role of uncertainty in cognitive change. *Journal of Abnormal and Social Psychology*, 1960, **61**, 168-175.

RECEIVED: January 25, 1971