

INFORMATION INTEGRATION IN MULTIPLE-CUE JUDGMENTS

Berndt Brehmer

and

Paul Slovic

University of Uppsala

Decision Research  
A Branch of Perceptrics, Inc.

Send Proofs To:

Berndt Brehmer  
Department of Psychology  
University of Uppsala  
Box 227  
S-751 04 Uppsala  
Sweden

Short Title:

Information Integration  
in Judgment

Journal of Experimental Psychology:  
Human Perception and Performance,  
in press.

### Abstract

When people attempt to integrate multiple cues into a single judgment, does the cognitive load produced by the integration process lead to simplification of the cue-judgment relationships? Three experiments tested the hypothesis that the strain of integrating information is reduced by treating non-linear cues as though they were linearly related to the object being judged. The experiments differed with respect to task content, number of cues and functional relations between cues and judgments. The results did not support the hypothesis. Scale values derived from single stimulus scales did not differ from those obtained with information integration procedures; there were no significant deviations from bilinearity when the cue values were compared across the two scale types. This finding was interpreted as support for the assumption in information integration theory that cue values are translated into judgment relevant subjective values before the integration process.

# INFORMATION INTEGRATION IN MULTIPLE-CUE JUDGMENTS

Berndt Brehmer

University of Uppsala

and

Paul Slovic

Decision Research

A Branch of Perceptronic, Inc.

Many important judgment tasks require that a person combine information from multiple cues to make an inference about some criterion variable. In medical diagnosis, for example, the physician must use information provided by symptoms, laboratory tests, and physical examination to infer the nature and severity of the patient's illness. To make such an inference correctly, the physician must (a) know the rules relating the cues to the various diagnostic possibilities and (b) be able to apply those rules.

It seems intuitively obvious that if people know which rules or policies are correct, they will be able to apply them. However, this may not be so. A number of studies have shown that when the task gets complex, people's judgments and decisions often reflect their desired rules and policies imperfectly, due to the action of random error and systematic biases (e.g., Brehmer, 1971; Hammond & Summers, 1972; Lichtenstein & Slovic, 1971, 1973; Slovic & MacPhillamy, 1974). Faithful implementation of one's desired judgmental policy appears to involve a degree of cognitive skill that often exceeds human capabilities.

The present study is concerned with one specific aspect of cognitive skill, people's ability to integrate information from several cues. In particular, we are interested in learning whether the subjective values

of the component cues become distorted during the integration process.

The possibility of distortion arises because the requirement to integrate information will tend to increase cognitive load. Given that the capacity to process information is limited, such an increase in load may necessitate a corresponding decrease in load in some other respect. For example, since people find it difficult to use cues that are non-linearly related to the judgment criterion (Brehmer, 1971; Brehmer & Qvarnstrom, 1976; Hammond & Summers, 1972), integration of information may lead to a "linearization" of all individual cue-criterion relations such that judgments after integration would not reflect the person's beliefs about the non-linearity of these functions.

Some support for the "linearization hypothesis" comes from the work of Björkman (1965) who had college students predict the motion time of an iron ball rolling down an inclined chute. Judgments in this task should be proportional to the square root of the ratio between distance along the chute and the sine of the angle of inclination. Although the students were undoubtedly familiar with the concept of acceleration, they were unable to take it into account, even when they were instructed that time is not proportional to distance and were given 30 feedback trials. Instead, their judgments were linearly related to distance and the height at the far end of the chute.

The findings of a "linearization" effect due to the integration process would have important implications regarding the use of information integration theory (Anderson, 1974) to obtain subjective scale values. This procedure requires people to integrate information from different cue dimensions, the values of which are presented according to a factorial design. If the results show that people combine information in an additive way, then the mean responses for the levels of each

dimension (computed across the values of the other dimensions) are on an interval scale. These procedures thus provide validated interval scales for the stimulus dimensions, but the test for validity can be performed only at the expense of having people integrate information. As noted above, however, this may lead to distortion, and the scale values obtained with the integration procedure may deviate from the "true" scale values because of problems in performing the integration. It is the purpose of the present study to investigate this problem.

Three experiments were performed. The first experiment used a task in which subjects could use their prior knowledge. The other two employed fictitious medical tasks in which the experimenter gave subjects information about the basic relations to be used in the judgments. In each experiment, the stimulus scale values obtained from an integration procedure were compared with those obtained when the cues were presented individually. Because some types of integration tasks produce context effects and other forms of non-additivity (Birnbaum, 1973, 1974), care was taken to select tasks for which an additive integration rule was likely to be valid.

### Experiment 1

#### Method

Subjects. The subjects were 55 women and 42 men undergraduate students from the University of Oregon. They were paid for participating.

Procedure. The subjects were asked to evaluate the attractiveness that various jobs would have for them. Each subject evaluated jobs under three different conditions. In Condition SL, subjects were given only the dimension of salary level on which to base their judgments. In Condition CT, they evaluated jobs solely on the basis of commuting time. These two conditions will also be referred to as the "single-cue"

conditions. In Condition SL + CT, a "double-cue" condition, subjects had both SL and CT available to use conjointly when making their judgments. The dimensions of salary and commuting time were chosen because of their salience to the criterion of job attractiveness and because of the a priori likelihood that each would relate to this criterion in a monotonic but nonlinear fashion.

In Condition SL, subjects were instructed as follows:

How important is salary as a characteristic of a job? Assume that you are considering a series of jobs, all of which were of satisfactory interest to you with regard to the work they entailed, chance for growth or promotion, etc. Assume that a salary of \$2,000 a year made the job have a value of 10 on your scale of job attractiveness. Indicate the scale value of attractiveness that the same job would have, given the various salaries listed below. Use 2,000 = 10 as a standard. So, for example, if \$4,000 was 3 times more attractive than \$2,000, you would give it a value of 30. Use any number up to 1000 for your judgments.

The subjects then evaluated five jobs with salary levels of \$4,000, \$8,000, \$12,000, \$16,000 and \$20,000.<sup>1</sup>

When judging on the basis of CT, the subjects were instructed:

Next assume that the jobs had satisfactory salary, interest value, etc., but differed with respect to commuting time. Also, assume that 1 1/2 hours--1 way--commuting time (driving) was equal to a value of 10 on a scale of attractiveness. How attractive would the jobs be if they had one of the following commuting times? Use any numbers up to 1000 as an upper limit for your judgments.

The five jobs evaluated in this Condition CT had commuting times of 1 1/4 hours, 1 hour, 45 minutes, 30 minutes, and 15 minutes.

The instructions for Condition SL + CT were as follows:

Next we would like you to judge some more jobs. This time you will have both types of information--salary and commuting time--on which to base your judgments. Assume that the jobs are satisfactory in all other respects. As a baseline to help you judge, assume that a job with a salary of \$2,000 and a one-way commuting time (by car) of 1 1/2 hours is equal to an attractiveness rating of 10. You can use any numbers you wish as long as they are less than 1000.

In Condition SL and CT, all five jobs were presented on a single page. In Condition SL + CT, there were 25 jobs to evaluate. These 25 were constructed by combining all five levels of SL with the five levels of CT in a factorial design. Only one job was presented on each page of the stimulus booklet. The standard job (salary of \$2,000, commuting time of 1 1/2 hours) was also presented on each page. The standard was assigned a value of 10, the same value assigned to the individual components, in order to induce an averaging set in the subjects.

The subjects evaluated each job in Condition SL and CT three times. They evaluated each of the jobs in Condition SL + CT twice. They judged 10 practice stimuli prior to evaluating any of the jobs in Condition SL + CT.

### Results

Additivity. For Condition SL + CT, a 5 (levels of SL) by 5 (levels of CT) analysis was performed on each subject's responses to test for deviations from additivity. Eleven out of the 97 subjects showed significant interactions indicating that they did not combine the information from the two cues additively. Since for these subjects, it cannot be shown that their scale values are on an interval

scale level, they were excluded from further analyses.

Utilization of single cues vs. cues in combination. For the remaining 86 subjects, the marginal mean response for each cue value and replicate was computed to yield scale values for the cues for the integration task. The cue values were then compared to those obtained with the single cues in a 2 (tasks: single stimulus task vs. integration task) x 5 (cue values) analysis of variance with two replicates. For these analyses, the first replicate of the single stimulus conditions was discarded to obtain equal numbers of observations for each cue. For cue SL, 16 analyses yielded significant conditions by cue values interactions, and for cue CT, 22 analyses yielded such interactions. Six subjects yielded significant interactions for both cues. These results show that 32 of the 86 subjects had significant deviations from parallelism. However, deviation from parallelism does not necessarily indicate differences in scale values, but the interactions may stem from differences in slope, which may come about because the subjects follow an averaging model rather than an adding model when integrating information from the two cues. Therefore, only those interactions involving significant deviations from bilinearity can be considered real cases in which there are systematic differences in scale values. For each of the 38 interactions in the above analyses, therefore, a test for deviations from bilinearity was performed using the program developed by Shanteau (1977). The results of these analyses showed 3 cases of significant deviations for cue SL and 4 cases of significant deviations for cue CT. Examination of the data from each of these cases produced no evidence that integration led to simplification of the task.

---

Mean scale values for the subjects who showed no significant interactions are shown in Figure 1. Although there were no differences in the linearity of the scale values, the divergent slopes in the right-hand

---

panel suggest that commuting time received relatively less weight than salary level in the integration task.

-----  
Insert Figure 1 about here  
-----

### Discussion

The results of the present experiment do not present any compelling evidence that integration of information leads to distortion. Only 5 of 86 subjects exhibited any systematic differences between the scale values obtained in the integration condition and those obtained in the single stimulus conditions. Even for these five subjects there was no evidence that information was used in a simpler way in the integration condition. Consequently, the results of Experiment 1 do not support the hypothesis that the requirement to integrate information leads to a cognitive simplification of the task. To some extent, this may be due to the fact that the subjects were using information in a simple way to start with. Trend analyses performed on the scale values obtained under the various conditions of the experiment indicated that fewer than half of the subjects had significant ( $p < .05$ ) non-linear relations between objective and subjective scale values. It may thus be that the task used in this experiment was not complex enough to generate any real need for cognitive simplification. To test this hypothesis, Experiment 2 was performed.

### Experiment 2

Experiment 2 differs from Experiment 1 mainly in that it used a fictitious medical task. Therefore, the subjects could not use their preexperimental knowledge but had to rely on the information supplied by the experimenter about the functional relations in the task. Complex relations could thus be introduced.

## Method

Subjects. Twenty-one undergraduate students from the University of Umeå were paid to serve as subjects.

Judgment task. The judgment task required the subjects to infer the severity of a disease called Brunswik's Egoni from two symptoms: Lavidity hormone and UCE level in the blood. The subjects were instructed that these symptoms were independent, so that a low level of the one symptom could be accompanied by a high level of the other and vice versa. They were further informed that both symptoms were equally important. Finally, they were informed about the functional relation between each cue and the criterion, which was a linear function in the one case and a J-shaped relation in the other. Information about these functional relations was given both graphically and verbally. Specifically, the subjects were shown graphs displaying a linear function and a J-shaped function. There were no numbers on the axes of graphs, so these graphs gave information only about the general nature of the function. In addition, they were told that one function was linear, meaning that the higher the level of Lavidity hormone the more severe the disease. They were also told that the other function was J-shaped, meaning that there was a value of UCE leading to a minimum severity, and that values above and below this minimum indicated a more severe disease, but that the effect was greater for values above the minimum than for values below it.

Each cue could take one of five different levels, and information about the level of the cue was given in the form of a bar, the height of which indicated the cue level. The severity of the disease was measured in terms of the days of illness and could vary from 1 through 40.

There were two kinds of cases intermixed in the judgment task. For some cases, the subjects were given the values of both cues. For other cases, they were given information about only one of the cues.

The subjects made four judgments for each of the twenty-five possible combinations resulting from a factorial combination of both cues, and four judgments for each value of the single cues, i.e., a total of 140 judgments.

Procedure. The instructions described the general nature of the task as stated above. The subjects were instructed not to compute the criterion values, but to give their intuitive impressions. Before making any judgments, the subjects were shown a case with the highest and lowest values of the symptoms.

The judgment task was presented in booklets, and the subjects recorded their responses on special answer sheets. They were allowed to work through their booklets at their own pace, and an experimental session required about one hour.

### Results

The first replicate was considered as practice and discarded from the analyses. The results for the remaining three replicates were analyzed in the same way as the results from Experiment 1.

Additivity. For each subject, a 5 (level of cue 1) by 5 (levels of cue 2) analysis of variance with three replicates was conducted to assess the additivity of the integration rule. The results of these analyses showed no evidence of systematic non-additivity. Only one subject yielded a significant interaction, and for this subject, inspection of the interaction indicated that it was due to a single errant judgment rather than to a systematic non-additive rule. Consequently, it seems reasonable to conclude that the scale values obtained from the integration task have interval properties, and if the more complex integration task yields interval scales, the simpler task of single stimulus judgments should also have yielded such scales.

Comparison of scale values from the integration task and the single stimulus task. For the integration task, scale values were computed for each replicate, and a 2 (tasks: single stimulus task vs. integration task) by 5 (cue levels) analysis of variance was performed for each cue and subject. There were no significant interactions, indicating that the scales obtained by the two methods were parallel (see Figure 2). Consequently, there is no evidence that integration caused distortion in this experiment.

-----  
Insert Figure 2 about here  
-----

### Discussion

The results of Experiment 2 are similar to those of Experiment 1 in that they yield no evidence that information integration leads to any change in the scale values. This result is obtained despite the fact that in Experiment 2, subjects had to use (and also used, as evidenced by plots of their scale values) a very complex J-shaped relation. However, it may be that the task was still too simple in that it contained only one non-linear relation. If task demands were increased by adding another non-linear cue, greater needs for simplification might arise, and evidence of distortion might be obtained. Experiment 3 was designed to test this hypothesis.

### Experiment 3

Experiment 3 also used the Brunswik's Egoni task, but to increase complexity, an additional cue called ELUR was introduced making the task a three-cue task instead of a two-cue task. In one condition of the experiment further complexity was introduced by requiring the subjects to assign different weights to the cues.

### Method

Subjects. Twenty undergraduate students from the University of Umeå were paid to serve as subjects.

Judgment task. As mentioned above, a new cue was added to the Brunswik Egoni task. The relation between UCE and severity of disease was a positive linear function, that between Lavidity and severity of disease a U-shaped relation, and that between ELUR and severity of disease a J-shaped function. Half the subjects were told that all cues were equally important. The other half were told that UCE was twice as important as the other cues.

Instruction procedure. The instructions and procedure were the same as in Experiment 2 except for the changes made necessary by the addition of the third cue. To reduce the number of judgments, only two levels, levels 2 and 4, of the linear cue were used in the integration task. Five practice trials were added and after these, the subjects made judgments for two replicates of the  $2 \times 5 \times 5 = 50$  cases of the integration task, and the  $3 \times 5$  cases for the single stimulus conditions. That is, the subjects made 130 judgments in all.

### Results

Additivity. Only two subjects yielded significant interactions in the  $2 \times 5 \times 5$  analysis of variance used to assess additivity. Inspection of the interactions showed, as in Experiment 2, that these interactions were due to single deviant responses, rather than the systematically non-linear integration rules. Thus, we may be confident that the scale values obtained in this experiment have interval properties.

Comparison of the scale values from the integration task and the single stimulus task. There were no interactions in the 2 (conditions) by 5 (cue levels) analyses of variance performed for each subject and non-linear cue. Thus there was no evidence that integration led to any change in the cue values (see Figure 3).

-----  
 Insert Figure 3 about here  
 -----

## Discussion

The results of this experiment are consistent with those of the earlier experiments in that there is no evidence that the requirement to combine information from different cues leads to any systematic change in scale values. Despite the added complexity of a third non-linear cue and unequal weights, subjects reproduced the function rules equally well in both the integration and single stimulus tasks. Since the error terms were generally small and the test for deviations from bilinearity is based on all but one of the degrees of freedom for the interaction, it does not seem likely that lack of statistical power is an explanation for the results.

Thus, several experimental tasks ranging from simple to quite complex, have failed to yield any evidence of distortion due to integration. This does not, of course, mean that there never will be any distortion of this type. It is possible that under conditions providing more extreme cognitive strain (e.g., time pressure, tasks involving verbal rather than quantitative stimuli) distortion will occur.

Whether or not there will be any distortion is, of course, very much dependent on how the subjects actually perform the integration. If, as is assumed in Anderson's theory (e.g., Anderson, 1974) subjects first translate the cue values into scale values and then integrate, cue-criterion function form may not affect cognitive load because the functions disappear when the cue values are translated into scale values. Indeed, from Anderson's point of view, cue-criterion function form should not affect the integration operation. Only the subjective scale values count, and the subject may not be cognizant of any non-linear relations in the task when integrating the information. According to this theory, then, increased cognitive load does not necessarily lead to any simplification of the functional relations

in the task, though it may lead the subject to ignore some of the cues.

Since the present experiments have yielded no evidence of distortion in the integration process, they suggest that the scale values derived from information integration theory methods are resistant to the sort of bias hypothesized in the introduction. The results also suggest that single stimulus methods may give results that are as good as those obtained by the integration task. However, we could not have discovered this had we not had the results from the integration task with which to compare the single stimulus results. Furthermore, the integration task provides an opportunity for assessing the validity of the scale values, which the single stimulus task does not do.

## References

- Anderson, N. H. Information integration theory: A brief survey.  
In D. H. Krantz, R. C. Atkinson, R. D. Luce & P. Suppes (Eds.),  
Contemporary Developments in Mathematical Psychology, Vol. II.  
Measurement, Psychophysics and Neural Information Processing. San  
Francisco: Freeman, 1974.
- Birnbaum, M. H. Morality judgment: Test of an averaging model with  
differential weights. Journal of Experimental Psychology, 1973,  
99, 3, 395-399.
- Birnbaum, M. H. The nonadditivity of personality impressions. Journal  
of Experimental Psychology, Monograph, 1974, 102, 3, 543-561.
- Björkman, M. Studies in predictive behavior. Explorations into predic-  
tive judgments based on functional learning and defined by estimation,  
categorization and choice. Scandinavian Journal of Psychology, 1965,  
6, 129-156.
- Brehmer, B. Subjects' ability to use functional rules. Psychonomic  
Science, 1971, 24, 259-260.
- Brehmer, B. & Qvarnstrom, G. Information integration and subjective  
weights in multiple-cue judgments. Organizational Behavior and  
Human Performance, 1976, 17, 118-126.
- Hammond, K. R. & Summers, D. A. Cognitive control. Psychological Review,  
1972, 79, 58-67.
- Lichtenstein, S. & Slovic, P. Reversals of preference between bids and  
choices in gambling decisions. Journal of Experimental Psychology,  
1971, 89, 46-55.
- Lichtenstein, S. & Slovic, P. Response-induced reversals of preference  
in gambling: An extended replication in Las Vegas. Journal of  
Experimental Psychology, 1973, 101, 16-20.

Shanteau, J. POLYLIN: A FORTRAN IV program for the analysis of multiplicative (multilinear) trend components of interactions.

Behavior Research Methods & Instrumentation, 1977, 9, 381-382.

Slovic, P. & MacPhillamy, D. J. Dimensional commensurability and cue utilization in comparative judgment. Organizational Behavior and Human Performance, 1974, 11, 172-194.

## Footnotes

We are indebted to Roger Johansson and Jan Kuylenstierna for assistance in the experimental work and to Norman H. Anderson, James Shanteau and two anonymous referees for their comments on an earlier version of the paper. The study was supported in part by a grant from the Swedish Council for Social Science Research, in part by Contract N00014-68-C-0431 with the Office of Naval Research, and in part by Contract N00014-79-C-0029 (ARPA Order No. 3668) to Perceptronics, Inc.

Requests for reprints may be sent to Berndt Brehmer, Department of Psychology, University of Uppsala, Box 227, S-751 04, Uppsala, Sweden, or to Paul Slovic, Decision Research, 1201 Oak Street, Eugene, Oregon 97401, USA.

1. Experiment 1 was conducted several years ago when even the lowest of these salary levels was not unreasonable. If the experiment were to be replicated at the present time, without adjusting the salaries for inflation, the low levels might trigger non-additive evaluations of the sort found by Birnbaum (1974).

## Figure Captions

Figure 1. Mean scale values for the subjects not showing any interactions in the cue value by type of scale analysis. Results for Salary Level are shown in the left panel, and results for Commuting Time in the right panel.

Figure 2. Mean scale values for the two cues in Experiment 2.

Figure 3. Mean scale values for the U-cue and J-cue in Experiment 3.