

## Test of an Additive Model of Social Inference

Timothy Anderson  
Kansas State University

Michael H. Birnbaum  
University of Illinois at Urbana-Champaign

Ratings of a hypothetical person's attitude or likely behavior toward a group are a nonadditive function of his/her attitudes or behaviors toward subgroups. The effect of a person's attitude toward one of several subgroups is inversely related to the total number of attitudes or behaviors toward different subgroups. Addition of neutral instances raises the ratings of sets of low-valued items and lowers the ratings of high-valued sets. These critical interactions violate the basic independence assumption of a general class of additive models. Furthermore, the diminished effect of one item due to the total number of items did not depend on whether the items were redundant in value, contrary to the expectancy-impact, values of instances model. The results were qualitatively consistent with an averaging model, although quantitative tests suggested that sets containing a greater number of items may receive reduced absolute weight.

Knowing a person's attitude toward Negroes and Jews, one could form an inference of the person's general attitude toward minorities. Gollob, Rossman, and Abelson (1973) theorized that social inferences of this sort obey an additive model in which the resultant inference of one's attitude toward a group is assumed to be the sum of the values of instances of the attitudes or behaviors toward subgroups. For example, if a man despises Jews and feels superior to Negroes, the additive model predicts that the inferred general attitude, dislikes minorities, is the sum of the values of these two instances.

The additive models studied in the past have assumed that the effect of a piece of information will be independent of the other information with which it is combined. Thus, a given piece of information should produce the same directional effect on the social inference, independent of other instances in the set. In contrast, averaging models of information integration (Anderson, 1971, 1974a, 1974b; Birnbaum, 1973, 1974; Birnbaum,

Wong, & Wong, in press) predict that a neutral-valued instance could produce opposite directional effects in combination, lowering the judgment of a set of positive instances and raising the judgment of a negative set. Certain experimental designs do not allow the distinction between constant-weight averaging models and additive models. When appropriate designs have been used, research in several information integration domains has ruled out additive models in favor of averaging formulations (Anderson, 1971; Birnbaum, 1974; Birnbaum et al., in press).

It is important to note that the "additive, values of instances model" of Gollob et al. (1973) is not additive in the traditional sense and can account for a result that would ordinarily be interpreted as critical evidence against adding models. Gollob et al. (1973) assume that the impact of redundant information will be less. This assumption is termed the *expectancy-impact hypothesis*: Expected information should have smaller effects.

The adding model, with parameter invariance, predicts that if  $a + c > a$ , then  $b + c > b$ . However, information integration experiments often find that  $J(L + LH) > J(L)$ , yet  $J(H + LH) < J(H)$ , where  $L$  is a low-valued item,  $H$  is a high-valued item,  $LH$  is a combination, and  $J$  represents the judgment. The data of Gollob et al. (1973) show that the addition of a disconfirming-confirming combination ( $HL$ ) to a disconfirming

---

This project was initiated at Kansas State University with the support of Kansas Bureau of General Research. We thank Harry Gollob, Robert Abelson, Robert Wyer, and Gerald Clore for their helpful criticisms.

Requests for reprints should be sent to Michael H. Birnbaum, Department of Psychology, University of Illinois at Urbana-Champaign, Champaign, Illinois 61820.

instance (L) raises the rating from 3.58 to 4.57, whereas the addition of the same HL combination to an H item lowers the rating from 6.85 to 6.67. This result would ordinarily be interpreted as inconsistent with additive models. Two points that have not been noted previously deserve emphasis: (a) The values of instances model (Gollob et al., 1973) can account for this result; (b) the values of instances model is not equivalent to the averaging model and can be tested against it with a suitable experiment.

The present experiment provides a more critical test of additive models of social inference, allowing a differentiation of the averaging models from the additive, values of instances model.

## METHOD

### Instructions

The instructions were similar to those of Gollob et al. (1973) and included the following:

The purpose of this experiment is to investigate how people use various sorts of information in drawing conclusions about other people. For example, I am interested in learning how people answer questions such as: "If you know that a man respects Japanese, how likely do you think it is that the man, generally speaking, respects foreigners?" In past work of this type it has been found that, rather than use the names of real groups of people such as Japanese and foreigners, it is useful to ask the questions about an imaginary group of people called Antuvians. (Gollob et al., 1973, p. 23)

Each Antuvian was said to belong to one, and only one, of 12 equal-sized subgroups with names such as Buvians, Cuvians, Duvians, and so on. Thus, the 12 subgroups form a mutually exclusive, exhaustive partition of Antuvians.

The subjects were instructed that each set of instances contained information about a man's attitudes or actions toward one or more subgroups of Antuvians. Their tasks were to rate the likelihood of this man *liking* or *helping* Antuvians, based on the set of instances describing his *feelings* or *actions* toward the subgroups. For example, if you know that a man loves Guvians, feels superior to Huvians, and ignores Kuvians, what is the likelihood that the man, generally speaking, likes Antuvians? If you know that a man protects Guvians, criticizes Juvians, and associates with Huvians, what is the likelihood that the man helps Antuvians?

The subjects were instructed to rate their inferences of liking and helping using 9-point scales with categories ranging from 1 (very, very unlikely) to 9 (very, very likely), in which 5 was designated as uncertain.

### Stimuli

The sets were composed of 46 instances that were chosen on the basis of preliminary scaling of 132 items to fit categories of low (L), slightly negative (M-), neutral (N), slightly positive (M+), or high (H) evaluation. Half of the 46 items described *manifest* or observable behaviors, and the other half described *subjective* or unobservable feelings. The following are examples of the items chosen for each category, with their mean ratings from the preliminary scaling experiments:

1. subjective—L, despises (1.76); M-, feels superior to (3.35); N, is indifferent toward (5.06); M+, approves of (6.76); H, loves (8.29); and
2. manifest—L, hurts (1.76); M-, criticizes (3.41); N, talks about (4.82); M+, listens to the ideas of (6.41); H, protects (7.88).

### Design

There were 60 sets of instances generated from a  $5 \times 3 \times 4$  (value of first instance  $\times$  number of subsequent instances  $\times$  value of subsequent instances) factorial design, in which the first instance had levels of value of L, M-, N, M+, and H; the number of subsequent instances was 0, 2, or 6 items, and subsequent instances had values of M-, N, M-M+, and M+. Subsequent instances of 2(M-M+) had one item of each value; similarly, 6(M-M+) had three M- and three M+.

There were two replications of the entire design, one using subjective instances and the rating of *liking*, the other using manifest instances and the rating of *helping*. Within each replication, the factor for the first instance used the same item at each level of value throughout. The experimental design is also portrayed in Figure 2, where the mean ratings are reported averaged across the stimulus replicates.

### Procedure

The 120 trials were printed in random order on 22 pages which were shuffled in different orders to form booklets. The cover page of each booklet included written instructions and an explanation of the rating scales. Subjects were instructed to glance through the booklet to familiarize themselves with the task before beginning. They were permitted to work at their own pace and completed the task in 45 minutes to 1½ hours.

### Subjects

The subjects were 50 Kansas State University undergraduates who received extra credit in a psychology course.

## RESULTS

Figure 1 plots the mean rating of social inference as a function of the value of the first instance. Panel A plots the results for sets of three instances, with a separate curve

for subsequent pairs of instances of M-, M-M+, N, or M+ value. Panel B plots the results for sets of seven instances in similar fashion. Since there were no important differences between the subjective and manifest replications when graphed separately, the data are averaged across replicates. All of the following analyses were performed separately on each replicate with the result that both sets of data supported the same theoretical conclusions as the averaged data.

Except for sets including neutral subsequent instances, the curves in each panel are nearly parallel; the interaction is nonsignificant,  $F(8, 392) = .73$ . However, with the inclusion of neutral instances, the interaction (the N curves crossing the M-M+ curve) is statistically significant,  $F(12, 588) = 5.07$ . This interaction would be inconsistent with simple additive or constant-weight averaging models but could be explained in terms of the averaging model by allowing subsequent N items to carry less weight than items of more extreme value.

Figure 2 plots the mean social inference as

a function of the value of the first instance, with a separate curve for each number of subsequent instances. Each panel shows the results for adding subsequent instances of different value. The interaction between the number of subsequent instances and value of subsequent instances is significant,  $F(6, 294) = 119.15$ , and is consistent with both averaging and adding models. The simple additive model predicts that the curves in Figure 2 should be parallel; that is, the effect of the value of the first instance (slopes) should be independent of the number of subsequent instances. Instead, the slopes in Figure 2 are a monotonically decreasing function of the number of items, consistent with averaging models,  $F(8, 392) = 32.84$ . Furthermore, this dependence of the effect of the value of the first item on the number of items appears independent of whether the subsequent items are redundant or nonredundant with the first. The three-way interaction of Value of the First Instance  $\times$  Value of Second Instance  $\times$  Number of Subsequent Instances is statistically significant but of trivial magnitude,

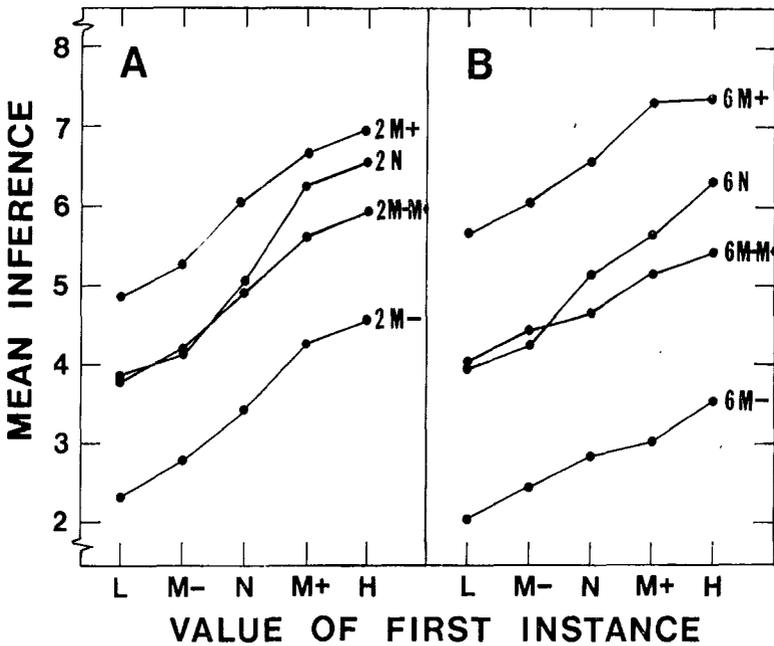


FIGURE 1. Mean ratings of social inference as a function of the value of the first instance. Curve parameters indicate that the value of the subsequent instances were either moderately unfavorable (M-), neutral (N), moderately favorable (M+), or mixed (M-M+). Panel A plots results for sets containing 2 subsequent instances. Panel B plots results for sets of 6 subsequent instances.

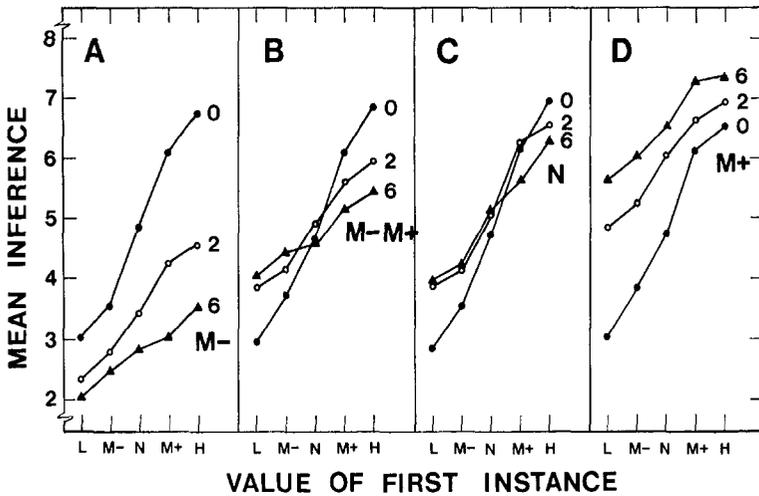


FIGURE 2. Mean ratings of social inference as a function of the first instance, with a separate curve for each number of subsequent instances. Different panels show results for adding subsequent instances of different value. Curves labeled 0 in separate panels represent different judgments of single first instances.

$F(24, 1176) = 1.55$ . The data do not support the prediction of the expectancy-impact, values of instances model, that the curves in Figure 2 would be parallel when nonredundant subsequent items are included.

DISCUSSION

The present data are inconsistent with additive models and the values of instances model of social inference proposed by Gollob et al. (1973). Social inferences appear to be better explained by the averaging model, which has also bested additive models in other information integration tasks (Anderson, 1971, 1974a, 1974b). Much of the recent discussion of additive and averaging models has perhaps been confused by a failure to specify sufficiently the assumptions of the models. The following sections define the additive and averaging models, expand on their psychological implications, explain why the general class of additive models is ruled out by the present data, and describe how these models would have to be modified to account for the present data.

Adding Model

Figure 3 depicts two alternative models of the social inference process. The additive model views the process of social inference as

an attempt to map out a space in which Buvians, Cuvians, Duvians, and the like, correspond to subsets of Antuvians, represented by the entire area. Each circle represents one instance. The value of each instance is represented by the number of plus or minus signs within the circle. The value of the social inference is assumed to be the difference between the total value of the positive and negative instances (i.e., the difference between the number of positive and negative signs in the diagram). Each symbol (sign) can be viewed as representing a certain number of Antuvians.

In the case where the subsets are (psychologically) mutually exclusive, this model is an additive model with parameter invariance. If there is no overlap among the sets, the model can be written

$$\Psi = \sum_{i=1}^k s_i, \tag{1}$$

where  $\Psi$  is the social inference based on  $k$  instances, and  $s_i$  is the value (either positive or negative) of the  $i$ th instance. This model assumes parameter invariance; that is, the value of  $s_i$  is independent of the value and number of other instances. This can be seen by considering Figure 3 for the case of mu-

tually exclusive sets (upper left-hand panel).

The simple (no overlap) additive model predicts that the effect of any instance will be independent of the value and number of other instances with which it is paired. The effect of the first instance is represented by the slopes in Figures 1 and 2. The additive model requires that all of the slopes be the same (i.e., that all of the curves in Figures 1 and 2 be parallel). Instead, Figure 1 shows that the first item has a greater effect when paired with  $N$  items than when paired with  $M$ — $M+$  combinations. Figure 2 shows that the effect of the first item is inversely related to the number of subsequent items. These critical crossover violations of independence rule out the simple additive model.

*Values of Instances Model*

Gollob et al. (1973) suggested an additive model that can be represented in these terms by allowing the sets to overlap. This plausible assumption is illustrated in the left-hand middle panel of Figure 3. Although the subjects are instructed that the sets are mutually exclusive, they would be reasonable to expect (by induction) that attitudes toward one subgroup might apply to others. Thus, when one hears that a man hates Buvians, it would be natural to induce that he may also hate some Cuvians and Duvians as well. Consequently, the additional information that the man hates Cuvians would not have as great an impact. The left-hand middle panel of Figure 3 illustrates an example in which each piece of information would be worth 7 units (plus signs) alone, but when added to another item of the same sign, it increases the total by 5, and when added to two other items, its increment is only 4. The increments are represented as those portions of the union that exclude the other sets (i.e., the increment is the measure of the intersection of a set with the complement of the union of the other sets).

The additive model of Gollob et al. (1973) can be written

$$\Psi = \sum_{i=1}^n s_i + \sum_{j=1}^m s'_j, \quad [2]$$

where  $\Psi$  is the social inference of  $n$  favorable instances and  $m$  unfavorable ones,  $s_i$  is the

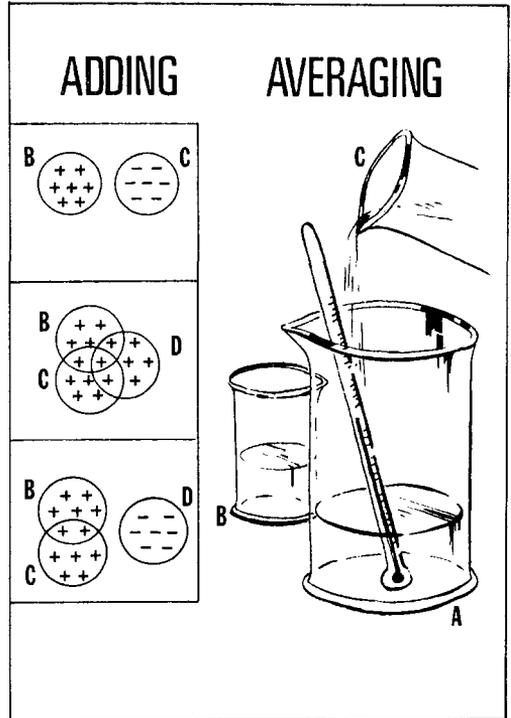


FIGURE 3. Additive and averaging models of social inference: Left-hand panels portray social inference as a measure of the union of sets, the difference between the number of favorable and unfavorable aspects. Overlaps (intersections) among same-signed instances allow for redundancy effects. Right-hand panel models the social inference as the resultant temperature of a mixture of solutions. The basic quality of an averaging model is illustrated by the fact that the effect of the temperature of Solution C will be inversely related to the total amount of liquid in the mixture.

value of the  $i$ th favorable instance, and  $s'_j$  is the value of the  $j$ th unfavorable instance. The model allows the value of  $s_i$  to depend on the particular phrase and on the other favorable instances; similarly, the value of  $s'_j$  depends on the other unfavorable instances. However, the model assumes no interaction between favorable and unfavorable instances.

The additive, values of instances model (Equation 2), by permitting the redundancy effect, can account for some of the crossovers (e.g., Figure 2, Panel B) that would previously have been assumed to rule out additive models. The left-hand, lower panel of Figure 3 illustrates how addition of an HL combination would lower the rating of an H

from 7 to 5; whereas, addition of the same HL combination to an L would raise the rating from  $-7$  to  $-5$ . This violation of independence is allowed because the second H or L in the set carries reduced value. Redundancy therefore allows a crossover, while still retaining an adding function for the integration process.

The additive, values of instances model predicts no interaction between the number of unfavorable instances and the value of favorable instances included in the same set. Therefore, it would predict that the curves in Figure 2, Panel A should be parallel for items of N, M+, and H value. Instead, the slopes are inversely related to the number of subsequent M- items. Similarly, the values of instances model predicts that the curves in Figure 2, Panel D should be parallel for L, M-, and N items. Instead, the effects of the first unfavorable item (the slopes) are inversely related to the number of favorable (M+) items. This result shows that the diminishing marginal effects of additional instances are better explained by an averaging model, in which the effect of any one instance is washed out by other instances, rather than by an expectancy impact model, in which the additive effect of an item would be presumed to depend on its redundancy.

The crossover in Figure 2, Panel C also presents great difficulty for the values of instances model. When adding instances of N value, redundancy should not occur; therefore, the effect of favorable or unfavorable information should be unaffected by the number of N instances. For an additive model to explain the neutral-item crossovers in Figure 2, Panel C and Figure 1, it would have to be assumed that items changed value and sign in linear contrast with the other items in the set.

This requirement would be consistent with a theory of social inference that is analogous to the idea of the *measure of the union* of several sets. The measure of a union of two sets is the sum of the measures of the two sets minus the measure of the intersection. One attraction of this approach, aside from the elegance of set theory, is that the intersection (redundancy) may be predictable from the difference in value. Thus, if a neu-

tral item can be represented by a set with elements of both positive and negative value, the intersection with a positive item will be positive so the contribution of a neutral item would be negative. Similarly, a neutral item will produce a positive effect when combined with a negative item since the intersection would be negative.

This modification seems unattractive for two reasons: First, it seems unlikely that judges would think, for example, that "talks about Buvians" should have a negative implication when the person also "listens to the ideas of Cuvians" but a positive value when combined with "disagrees with Cuvians." Second, previous research in information integration has shown that ratings of single items show assimilation to the other items, rather than the contrast that would be postulated to save the additive model (Anderson, 1966, 1971, 1974a). Nevertheless, this set theory version seems to deserve further scrutiny in empirical studies.

#### *Averaging Model*

The right-hand panel of Figure 3 models the process of social inference as the temperature of a mixture of solutions. Each piece of information is represented by a portion that will be added to the mixture. Thus, each solution has a temperature ( $s_i$ ) and a volume ( $w_i$ ) corresponding to scale value and weight. The initial impression corresponds to the fluid (presumably lukewarm) already in the mixing beaker. The resultant social inference,  $\Psi$ , of  $k$  items is presumed to correspond to the temperature of the mixture, which will be given by the averaging quotation

$$\Psi = \frac{\sum_{i=0}^k w_i s_i}{\sum_{i=0}^k w_i}, \quad [3]$$

where  $w_0$  and  $s_0$  are the volume and temperature of liquid already in the beaker.

The averaging model accounts for three important qualitative features of the data, whereas neither the simple adding model nor the values of instances model can explain all of the features of the data. First, the effect of the first instance (the slopes in Figure 2) should be inversely related to the number of items in the set. In the temperature model

(Figure 3), the effect of the temperature of Solution C will be inversely related to the total amount of fluid in the beaker.

Second, the crossover in Figure 1 is explained by postulating that neutral items are less diagnostic and have less weight. In the temperature analogy, it would be as though N-value solutions (of intermediate temperature) had less fluid. Consequently, the first item has greater effect (slopes in Figures 1 and 2) when the subsequent items are N.

Third, if the weights of M+ and M- items are equal and the weight of the first item is constant, the averaging model predicts the parallelism for the M-, M+, and M-M+ curves in Figure 1. The additive model (Equation 1) also predicts this parallelism. The expectancy-impact values-of-instances model, however, predicts a small interaction due to redundancy that does not appear in Figure 1.

Consistent with the averaging model, the slopes of Figure 1 vary inversely with the total number of items in the set, and the distances between the curves in Figure 1, Panel B increase as the slopes decrease. However, the averaging model would require some modification to give an accurate quantitative account of the slopes. The average slopes in Figure 2 (arbitrarily setting the abscissa variation to unity) are 3.81, 2.26, and 1.72. If the absolute weight of each item is constant, the slopes are proportional to  $w/(w_0 + w)$ ,  $w/(w_0 + 3w)$ , and  $w/(w_0 + 7w)$ . Hence, the reciprocals should be linearly related to 1, 3, and 7. Instead, the reciprocals are .26, .44, and .58. This result can be accounted for by a configural-weighted averaging model that allows absolute weight to decline with increasing set sizes. This configural-weight model was fit to the 60 means in Figure 2 using a least-squares criterion, with the aid of the STEPIT subroutine (Chandler, 1969). The mean squared discrepancy was .024, and the deviations appeared unrelated to set size. The estimates of the absolute weights, setting  $w_0 = 1$ , were  $w_1 = .984$ ,  $w_2 = 1.20$ , and  $w_3 = 2.61$ , for the first instance, sets of 2, and sets of 6 subsequent instances, respectively. These parameters are far from being proportional to 1, 2, and 6.

It might be thought that the configural weighting effect is analogous to the expectancy-impact hypothesis, since both are violations of the assumption of parameter invariance. However, the expectancy-impact hypothesis predicts that the effect would depend on whether the added information is of the same or different sign, whereas the configural weighting effect depends only on set size.

The experiment of Gollob et al. (1973) was not designed to differentiate the values of instances model from the averaging model. It is worthwhile to note that the averaging model can account for the pattern of results in their Table 3 without the expectancy-impact hypothesis. The negatively accelerated set-size function (one argument for expectancy impact) is a mathematical consequence of the averaging model. Although the present study does not support an expectancy interpretation, the general concept of expectancy may prove useful for explaining other effects in judgment (Birnbbaum, 1975).

The averaging model can thus account for all of the important qualitative features of the data. It can explain both the parallelism for M+ and M- items and the crossovers in Figure 1, the shifts in the location of the crossovers in Figure 2, and the variation of the slopes as a function of the number of items. To give a complete quantitative account, however, the model may require a configural weighting revision, in which the absolute weights diminish as the number of items in the set increases. In the temperature analogy, this would amount to a cook who added less of each solution when there were many solutions to add. With this simple modification, the averaging model can give a complete account of the present data.

#### REFERENCES

- Anderson, N. H. Component ratings in impression formation. *Psychonomic Science*, 1966, 6, 279-280.
- Anderson, N. H. Integration theory and attitude change. *Psychological Review*, 1971, 78, 171-206.
- 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. 2). San Francisco: W. H. Freeman, 1974. (a)
- Anderson, N. H. Integration theory applied to social attribution. In L. Berkowitz (Ed.), *Advances in*

- experimental social psychology*. Volume 7. New York: Academic Press, 1974. (b)
- Birnbaum, M. H. Morality judgment: Test of an averaging model with differential weights. *Journal of Experimental Psychology*, 1973, *99*, 395-399.
- Birnbaum, M. H. The nonadditivity of personality impressions. *Journal of Experimental Psychology*, 1974, *102*, 543-561. (Monograph)
- Birnbaum, M. H. Expectancy and judgment. In F. Restle, R. Shiffrin, N. J. Castellan, H. Lindman, & D. Pisoni (Eds.), *Cognitive theory* (Vol. 1). Hillsdale, N.J.: Erlbaum, 1975.
- Birnbaum, M. H., Wong, R., & Wong, L. Combining information from sources that vary in credibility. *Memory & Cognition*, in press.
- Chandler, J. D. Subroutine STEPIT—Finds local minima of a smooth function of several parameters. *Behavioral Science*, 1969, *14*, 81-82.
- Gollob, H. F., Rossman, B. B., & Abelson, R. P. Social inference as a function of the number of instances and consistency of information presented. *Journal of Personality and Social Psychology*, 1973, *27*, 19-33.

(Received February 17, 1975)