*Memory & Cognition*
*1976, Vol. 4 (3), 330-336*

# Combining information from sources that vary in credibility

MICHAEL H. BIRNBAUM
*University of Illinois, Champaign, Illinois 61820*

and

REBECCA WONG and LEIGHTON K. WONG
*University of California, San Diego, La Jolla, California 92037*

**Models describing the role of source credibility in information integration were tested in two experiments. In the first experiment, subjects estimated the value of used cars based on two cues: blue book value and an estimate provided by one of three friends who examined the car. The three sources were described as differing in mechanical expertise. In the second experiment, subjects rated the likeableness of persons described by either one or two adjectives, each adjective contributed by a different source. The sources differed with respect to the length of their acquaintance with the person to be rated. In both experiments, credibility of the source magnified the impact of the information he provided. Further, this multiplicative effect of a source was inversely related to the credibility of the other source, in violation of additive or constant-weight averaging models, but consistent with a relative-weight averaging model.**

Attitudes toward issues or persons are often based on inconsistent information provided by sources that differ in credibility. The Watergate scandal of the Nixon Administration provides one example. In televised Senate hearings, different members of the White House staff and the Committee to Reelect the President gave contradictory testimony about the extent of White House involvement in clandestine attempts to wiretap the Democratic headquarters and to obstruct the investigation by covering up the facts. At that point in time, *Dean* suggested that the President was involved in the Watergate cover up and *Haldeman* implied that he was not. Before secretly made tapes of White House conversations became available, millions of Americans combined this inconsistent information to form their own impressions of the likelihood that the President was involved. The Watergate scandal gave new meaning to the notion of an unimpeachable source.

Wartime communications provide another illustration. Suppose *Egypt* reported shooting down 10 Israeli planes and *Israel* reported the loss of only 2. The problem for the military decision maker is to combine this uncertain information to estimate the true number. But the psychologist has two theoretical problems: (a) to specify how sources affect the subjective value of the information they provide, and (b) to explain how the communications provided by the different sources are combined to form integrated judgments. Anderson (1971) and Rosenbaum and Levin (1968, 1969) have proposed averaging models of information integration in which the weight of a piece of information depends upon the credibility of its source. In the Middle East example, the estimated aircraft loss could be theorized to be the average of the Egyptian and Israeli values, falling closer to the source thought to be more credible.

This paper extends developments in integration theory (Anderson, 1971, 1974) to test simple mathematical descriptions of human judgment. The basic conceptualization views man as an intuitive statistician (Peterson & Beach, 1967) who subjectively aggregates varigated information, weighting bits according to their importance and credibility. The approach is to study simple situations in which the relevant variables (value and credibility of the information) are under experimental control, to test implications of explicit models for specific situations, and to speculate about the general implications for understanding social judgment.

This research was initiated while the first author held a National Institute of Mental Health postdoctoral fellowship at the Center for Human Information Processing, University of California, San Diego. These experiments were portions of undergraduate research projects. Thanks are due Norman H. Anderson for his supervision and support, provided by Grant GB-21028. We thank Clairice T. Veit and Robert Wyer, Jr., for comments, and Barbara Rose for help with pilot research. Requests for reprints should be addressed to Michael H. Birnbaum, Department of Psychology, University of Illinois at Urbana-Champaign, Champaign, Illinois 61820.

## EXPERIMENT I
## VALUE OF USED CARS

A used car is an entity of well-known uncertainty. It may have an undiscovered defect that will shortly

make it worthless. On the other hand, it may have hidden virtues that would cause it to give good service for many years. Judging the monetary value of a used car usually requires the combination of many informational cues: year, make, model, mileage, general condition, etc. In Experiment I, the subject's task was to estimate the value of used cars, based on just two cues: the blue book value (BBV) and an estimate (EST) provided by one of three friends who examined the car. The three friends (SOURCES) were described as unbiased, but differing in their understanding of automobile mechanics.

A reasonable model for this task would be given by the following equation:

$$R = w_{BBV}s_{BBV} + w_{EXP}s_{EST}, \quad (1)$$

where R is the judged worth, $w_{BBV}$ and $s_{BBV}$ are the weight and scale value of the blue book value, $w_{EXP}$ is the weight of the friend's estimate, presumably depending on his expertise, and $s_{EST}$ is the scale value of his estimate. Equation 1 predicts an interaction between expertise and estimate. Since it predicts no interaction between blue book value and the other factors, Equation 1 is termed an additive model. In Experiment I, Equation 1 is also equivalent to a constant-weight averaging model.

Another model which might describe the judgment process is a relative-weight averaging model. In this model, the value of an object would be the weighted average of the estimated values provided by different sources, with the absolute weight depending upon the credibility of the source. The additional assumption is a principle of relativity—the greater the *absolute* weight of one piece of information, the less the *relative* weights of the other information. This relative-weight averaging model can be written:

$$R = \left(\frac{w_{BBV}}{w_{BBV} + w_{EXP}}\right)s_{BBV} + \left(\frac{w_{EXP}}{w_{BBV} + w_{EXP}}\right)s_{EST}, \quad (2)$$

where R is the judged value of the car, $w_{BBV}$ and $s_{BBV}$ are the absolute weight and scale value of the BBV, $w_{EXP}$ and $s_{EST}$ are the weight of the source and the scale value of the source's estimate, respectively. It is assumed that the greater the friend's mechanical expertise, the greater his credibility ($w_{EXP}$). The blue book is considered a source with credibility $w_{BBV}$. In Equation 2, the absolute weights of the blue book value and the source's estimate are divided by the sum of the absolute weights.

Since the relative weights of the source and the BBV sum to one, Equation 2 can be rewritten:

$$R = s_{BBV} + w'_{EXP}(s_{EST} - s_{BBV}), \quad (3)$$

where $w'_{EXP} = w_{EXP}/(w_{BBV} + w_{EXP})$ is the relative weight of the source. Equation 3 shows that when $s_{EST} = s_{BBV}$, $R = s_{BBV}$. But when the EST and BBV do not agree, the response varies in proportion to the deviation between them according to the relative weight (credibility) of the source. The greater $w'_{EXP}$, the closer will be the response to the source's estimate. Formally, the model can be analyzed as a multilinear model, predicting that source combines multiplicatively with EST and BBV, but that the other effects are additive.

## Method

**Instructions.** The subjects were instructed that the purpose of the experiment was to study how people combine information to estimate the values of used cars. They were to attempt to estimate "true" value (neither over- nor underestimate), based on the blue book value and a friend's estimate of the value of the same car.

The blue book value (BBV) was described as a standard "fair" price that is determined by such factors as year, model, make, and mileage. It was remarked that BBV is widely relied upon by businesses, but that BBV might not describe a particular car; i.e., for a given type of car, some particular car might be more valuable than another.

The three friends were described as unbiased sources who were trying their best to estimate true value, based on a 30-min inspection and test drive. The three friends differed in their mechanical abilities. They were described as low-, medium-, and high-expertise sources by separate paragraphs that described their training and mechanical skills. The low-expertise friend was described as a competent person who drove a car regularly and had purchased cars for himself. The medium-expertise friend had taken some classes in auto shop and could make some repairs himself. The high-expertise friend was described as an "expert mechanic" whose hobby was the repair and modification of sports cars. The three friends were described as sensible but fallible; good or bad points of a car might go unnoticed in such an examination.

**Design.** There were 60 trials generated from a 4 by 3 by 5, BBV by SOURCE by EST, factorial design in which the levels of BBV were \$350, \$450, \$550, and \$650; the sources (friends) were low-, medium-, and high-expertise; the levels of friends' estimates were \$300, \$400, \$500, \$600, and \$700. In addition, there were 6 trials that paired BBV of \$250 with an estimate of \$200, or BBV of \$750 with an estimate of \$800, combined with each source.

The 66 trials were randomly intermixed and printed in random order in booklets. The first page of the booklet was a questionnaire that ascertained that subjects all had valid driver's licenses. The next pages contained the written instructions, followed by 14 representative practice trials (that included the range of values and differences among the independent variables), then the 66 experimental trials.

**Subjects.** The subjects were 50 University of California, San Diego undergraduates who were enrolled in lower division psychology classes.

## Results

Figure 1 shows the mean judged value as a function of the source's estimate with a separate curve for each level of source expertise. Each panel contains the data for a different level of BBV. For example, the leftmost panel of Figure 1 shows that when the BBV is \$350 and the friend's estimate is \$700, the judged value is either \$449, \$525, or \$592, depending on whether the source had low, medium, or high credibility. As predicted by both models, the slopes of the curves depend upon the source. The greater the expertise of the source, the greater the effect of his estimate.

The relative-weight model successfully predicts the locations of the crossovers in each panel. Equation 3 implies that when $s_{BBV} = s_{EST}$, there will be a crossover at $R = s_{BBV}$. Since the stimuli and responses are in the very familiar monetary unit *dollars*, it seems reasonable to suppose that the scale values for BBV, EST, and judged value are not far from their numerical monetary values. Under this simplifying assumption, Equation 3 predicts that the curves will cross when BBV = EST. As can be seen by the dashed lines in Figure 1, these assumptions give a good account of the crossovers.

Figure 2A shows mean judged value as a function of the source's estimate, averaged across BBV, with a separate curve for each source. The abscissa values have been spaced according to the marginal means. Consistent with the multiplicative relationship between source and estimate dictated by both models, the data (solid points) fall close to the predicted pattern (straight lines).

Figure 2B shows mean judged value as a function of BBV (spaced on the abscissa according to the BBV marginal means) with a separate curve for each source. The additive or constant-weight averaging model (Equation 1) predicts no interaction for Figure 2B. According to the relative-weight model, the *greater* the weight of the source, the *less* the relative effect of the BBV. Thus, the slopes of Figure 2B should have the opposite ordering of those in Figure 2A, forming a bilinear set of curves. Again, the empirical data (points) fall close to the bilinear predictions (straight lines in Figure 2B). Equation 3 nicely accounts for this important qualitative result that requires an *averaging* rather than an *additive* interpretation.

Figure 2C shows judged value as a function of the source's estimate with a separate curve for each level of BBV. Both models predict that the curves should be parallel. Although the bottom three curves appear roughly parallel, the curve for BBV = $650 shows a clear discrepancy. For example, when BBV = $650 and the friend says the car is worth $300, the judged value is lower than predicted, as if the subject imagined that the friend had located a major defect in the car that would not be reflected in the blue book value. This divergent interaction is statistically significant, $F(12,588) = 12.98$, $MS_e = 1083$, but does not seem overly serious. It does suggest that the model should be revised to allow weight to vary with the scale value, differential weighting, or with the configuration of estimates, configural weighting (Birnbaum, 1974). Consistent with Equation 3, the three-way interaction between source, BBV, and estimate was nonsignificant, $F(24,1176) = 1.14$; $MS_e = 789$.

The marginal means represent the functional values for estimate and BBV. As can be seen from the

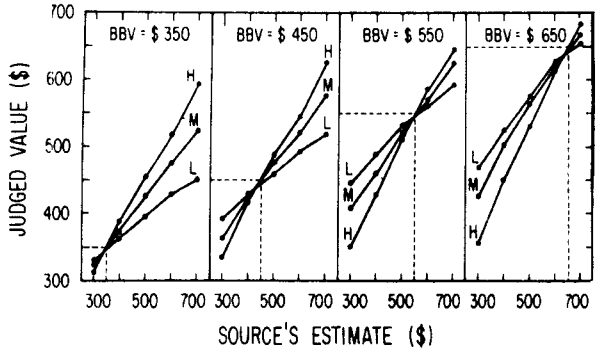


**Figure 1. Mean judged value of used cars as a function of the source's estimate with a separate curve for each level of source expertise (L = low, M = medium, H = high); each panel presents data for a different level of blue book value (Experiment I).**

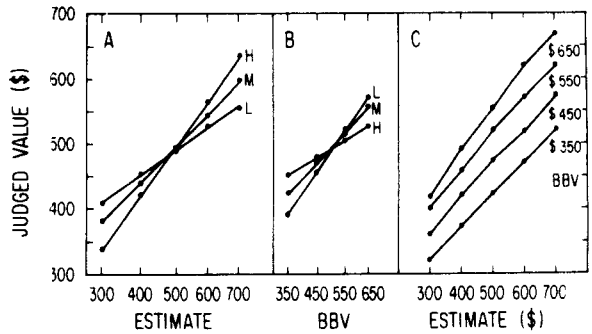


**Figure 2. Model analyses: (A) Mean judged value as a function of source's estimate with a separate curve for each level of source, averaged over blue book value; (B) mean judged value as a function of blue book value (BBV) with a separate curve for each level of source; (C) mean judged value as a function of source's estimate, with a separate curve for each level of blue book value (Experiment I).**

spacing in Figures 2A and 2B, the subjective value of money derived from the model would be a negatively accelerated function of dollar values. For example, the difference between $600 and $700 is less than the difference between $300 and $400.

The weights for the low-, medium-, and high-expertise sources were estimated to be .145, .303, and .745, respectively, compared to an arbitrary value of .255 for the weight of the BBV.

## Discussion

Experiment I shows that a conceptually simple model based on a relative-weight averaging mechanism can give a good account of some rather complex and interesting results. These results simultaneously rule out the additive model and the constant-weight averaging model. Although the additive models seem reasonable a priori possibilities, they cannot account for the critical finding (Figure 2B) that the effect of blue book value is *inversely* related to the credibility of the source. It is interesting that the impact of an "absolute" standard

like the blue book value depends on the credibility of the friend who examines the car. The relative-weight averaging model can easily account for this type of contextual effect since it predicts that the relative weight of a piece of information is inversely related to the sum of the weights.

## EXPERIMENT II
## SOURCE CREDIBILITY IN IMPRESSION FORMATION

One could argue intuitively that the evaluation of used cars and the formation of personality impressions might involve different processes of information integration. For example, if one source estimates that a car is worth $700 and another source estimates the value at $250, both cannot be simultaneously correct. Hence, the seeming contradiction may induce the subject to average the two discrepant estimates. However, there would be no logical contradiction if one source described a person as *loyal* and another described him as *malicious*. It is therefore of great interest to examine whether deductions from unidimensional algebraic models applicable to used car judgment would also find empirical support in personality impression formation.

Experiment II tests among three plausible models of source credibility in impression formation. The models differ in two important respects: the adding model predicts that the impact of a communication is independent of the number of communications, whereas the averaging models predict that the impact is inversely related to this number. The constant-weight model assumes that the impact of a piece of information depends on the credibility of its source, but is independent of the credibility of other sources. The relative-weight model, an extension of Equations 2 and 3, predicts that the effect of a piece of information varies directly with the credibility of its source and *inversely* with the credibility of other sources.

### The Models

It is assumed that the adjectives can be represented by scale values on a likeableness continuum and that the longer a source has been acquainted with the target person, the greater will be his credibility. In all of the models, credibility is represented by weight, which multiplies the scale value of the information. The values of the adjectives and weights of the sources are assumed to be independent of the adjectives and sources with which they are combined.

The *additive model* (Anderson, 1971, Equation 6) can be written:

$$R = w_0 s_0 + w_1 s_1 + w_2 s_2, \quad (4)$$

where R is the overall impression of likeableness, $w_1$ and $w_2$ are the weights of the two sources that presumably depend on length of acquaintance, $s_1$ and $s_2$ are the scale values of the adjectives provided by the two sources. Wyer (1974) has recently argued that this model is descriptive of source credibility effects in impression formation.

The *constant-weight* averaging model can be written:

$$R = (aw_0 s_0 + bw_1 s_1 + cw_2 s_2)/(a + b + c), \quad (5)$$

where $s_i$ is the scale value of the adjective provided by source i, and $w_i$ is the weight representing the credibility of source i; $w_0$ and $s_0$ refer to the weight and scale value of a postulated initial impression. The source-adjective communications (ws combinations) are then averaged (with weights a, b, and c) to form the overall evaluation. This equation is termed the constant-weight model since a, b, and c are assumed to be independent of source credibility.

When there are exactly two communications, this formulation is equivalent to the averaging model of Rosenbaum and Levin (1968, p. 169), and would not be distinguishable from the *adding* model. Experiment I used exactly two communications, explaining why Equation 1 encompassed both adding and constant-weight averaging models. Experiment II includes single source-adjective statements (setting c = 0), which permit one to distinguish adding vs. averaging models (Equations 4 vs. 5) by allowing a comparison of sets of one and two communications.

The relative-weight averaging model (Anderson, 1971) can be written:

$$R = (w_0 s_0 + w_1 s_1 + w_2 s_2)/(w_0 + w_1 + w_2), \quad (6)$$

where $w_1$ is the *absolute* weight of the first source and $w_1/(w_0 + w_1 + w_2)$ is the *relative* weight of the first source when paired with the second source. The relative-weight model differs from Equation 5 in the important respect that the relative weight of a piece of information is directly related to the credibility of its source and *inversely* related to the credibility of other sources. For single source-adjective statements, $w_2$ would be set to zero.

### Method

The subject's task was to read either one or two adjectives that described a person and to rate how much he would like such a person. Each adjective was attributed to a different source who had known the person for either *one meeting, 3 months, or 3 years.* For example, how much would you like a person who would be described by an *acquaintance of 3 years* as *understanding* and an *acquaintance of one meeting* as *blunt*? The ratings were made on a 1-19 scale with labels varying from 1 = dislike very very much, to 19 = like very very much, with 10 specified as the neutral point.

**Stimuli.** There were two sets of adjectives of low (L), medium (M), or high (H) likeableness. The Set 1 adjectives were *malicious,*

*solemn*, and *understanding*. The Set 2 adjectives were *phony*, *blunt*, and *loyal*.

**Design.** The three adjectives from Set 1 and Set 2 were combined separately with the three levels of source (*one meeting*, *3 months*, or *3 years*) to form two 3 by 3, Source by Adjective, factorial designs: Source 1 by Adjective 1 and Source 2 by Adjective 2. These two designs contained 18 *single* source-adjective communications.

The nine single source-adjective communications from each 3 by 3 design were then combined to form a 9 by 9 factorial design producing 81 *pairs* of source-adjective statements. The 9 by 9 design can thus be seen as a (3 by 3) by (3 by 3), (Source 1 by Adjective 1) by (Source 2 by Adjective 2), design where the numbers 1 and 2 refer to the first and second sources, respectively.

**Procedure.** The 18 sets of single source-adjective communications and 81 sets of pairs were randomly intermixed with 4 anchor sets (each consisting of 4H or 4L adjectives attributed to acquaintances of 3 years). These 104 trials were printed in random order in booklets with a cover page containing written instructions and the response scale. The subjects were instructed to read through the list before beginning.

**Subjects.** The subjects were 50 University of California, San Diego undergraduates.

## Results

Figure 3A shows the mean ratings of the single source-adjective communications for Set 1. The data are plotted against the adjective value, with a separate curve for each level of source. The figure shows that the rating is higher when a source of longer acquaintance provides a positive trait and lower when he provides a negative trait. Set 2 data are similar. The slopes of the curves are directly related to the source's length of acquaintance with the target person. This crossover interaction is predicted by the multiplicative relation between weight (source) and scale value (adjective) in all of the models. Since the abscissa values have been spaced according to the marginal means, the multiplicative model also predicts that the curves should be linear, intersecting at a common point. This graphical prediction (straight lines) appears to be in reasonable agreement with the data (solid points).

Figure 3B shows the mean ratings for the Source 1 by Adjective 1 communications, averaged over the Source 2 by Adjective 2 combinations. This figure is directly analogous to Figure 3A and shows that the crossover interaction between source and adjective is also obtained in the four-factor design. Again, the abscissa spacing corresponds to marginal means, with straight lines depicting the bilinearity prediction and solid points for the mean judgments. Similar results were also obtained for the Source 2 by Adjective 2 combinations graphed separately.

Comparison of Figures 3A and 3B tests adding vs. averaging models. According to the adding model, $R = w_0s_0 + w_1s_1 + w_2s_2$; hence, for single adjectives, $R = w_0s_0 + w_1s_1$; therefore, the effect of a given variation in $s_1$ should be independent of the number of items in the set. Thus, the ordinate variation, $\Delta R$, in Figures 3A and 3B should be the same. Instead, $\Delta R$ is less in Figure 3B for sets of two adjectives. According to the relative-weight averaging model, $\Delta R$ would be proportional to $w_1/(w_0 + w_1)$ in Figure 3A and $w_1/(w_0 + w_1 + w_2)$ in Figure 3B. Thus, Figures 3A and 3B are inconsistent with the additive model (Equation 4), but remain consistent with either the constant-weight averaging model (Equation 5) or the relative-weight averaging model (Equation 6).

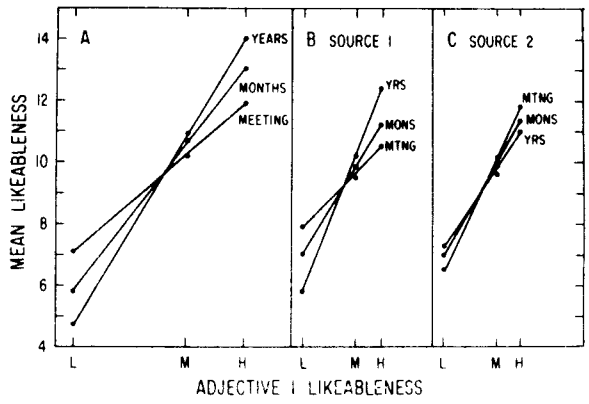


**Figure 3. (A) Mean rating of likeableness for single source-adjective combinations, plotted as a function of adjective value (L = low, M = medium, H = high value) with a separate curve for each source (the source has known the target person for either *one meeting, 3 months, or 3 years).* (B) Mean likeableness for source-adjective pairs, averaged over communications provided by the second source, plotted as in Panel A. (C) Mean likeableness as a function of the adjective provided by the first source with a separate curve for each level of the second source. (Note that the first adjective has *less* effect when the second source has greater credibility.) (Experiment II.)**

Figure 3C provides the evidence that discriminates between the two averaging models. Figure 3C plots mean ratings of two adjective combinations as a function of Adjective 1, with a separate curve for each level of Source 2. According to the constant-weight model (Equation 5), the effect of the first adjective should be independent of the second source. According to Equation 6, however, the relative weight (the slopes) of the information should be *inversely* related to the absolute weight of the other information. As can be seen from the figure, the data support Equation 6, since the effect of Adjective 1 (the slope) is inversely related to the length of acquaintance of Source 2. This interaction is statistically significant, $F(4,196) = 29.72$; similarly, the Source 1 by Adjective 2 interaction is also significant, $F(4,196) = 12.54$; $MS_e = 2.71$ and 2.02, respectively. Equation 6 also predicts that the slopes in Figure 3C should show less variation than the slopes of Figure 3B. This follows since $w_1/(w_0 + w_1 + w_2)$ will show greater variation as a function of a given variation of $w_1$ than it will for the same variation of $w_2$. These results show that the relative weight of a piece of information depends not only on the credibility of the source that provided the information, but also on the credibility of the other source.

Figure 4 shows the mean ratings for the larger, 3 by 3 by 3 by 3, design. The nine points in each panel

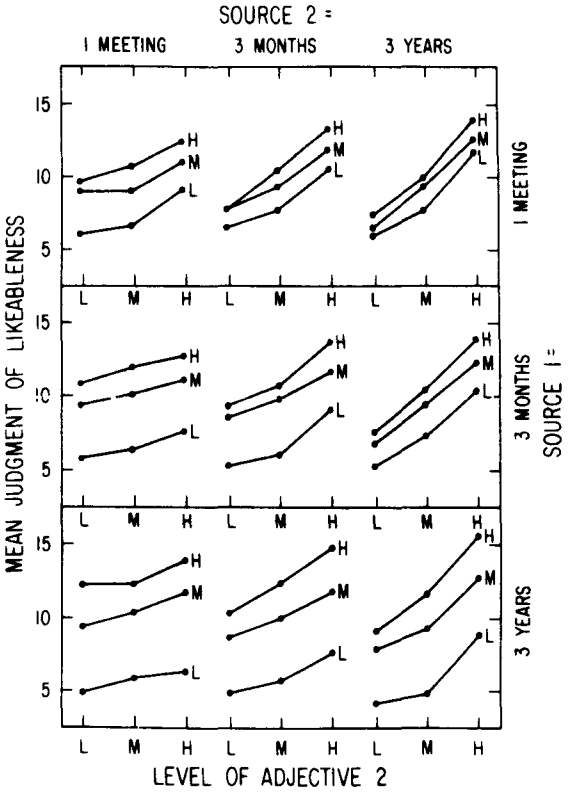


**Figure 4. Mean ratings of likeableness as a function of the adjective contributed by the second source with a separate curve for each adjective contributed by the first source. Each row of panels represents a different level of Source 1; each column of panels represents a different level of Source 2 (Experiment II).**

represent the same nine adjective combinations. Adjective 2 is plotted on the abscissa, and separate curves represent different values for Adjective 1. The slopes of the curves reflect the weight of Adjective 2; distances between the curves reflect the weight of Adjective 1. Each row of panels has a different level of Source 1; the first row represents *one meeting*, the second represents *3 months*, and the third represents *3 years*. Each column of panels has a different Source 2. Since increases in slopes reflect increases in weight ($w_2$), it can be seen that as one proceeds from the left panel to the right, the weight of Source 2 increases. Similarly, as Source 1 increases in length of acquaintance, the distances between the curves (reflecting $w_1$) increase. The need for relative weighting (Equation 6) can be seen in the figure as follows: as the slopes increase, the distances between the curves decrease. For example, the first row of panels shows that as the credibility of Source 2 increases, with the slopes ($w_2$) increasing, the distances between the curves decrease. This follows from Equation 6, since the relative weight of Adjective 1 is $w_1/(w_0 + w_1 + w_2)$; hence, relative weight of one piece of information is predicted to vary inversely with the absolute weight of the other information.

The need for a postulated initial impression ($w_0s_0$) can be seen most easily by studying the three panels in which the two sources are equal, i.e., the downward diagonal of Figure 4. As both sources increase in length of acquaintance, relative weight of the initial impression $[w_0/(w_0 + w_1 + w_2)]$ decreases. Since the value of $s_0$ would be near the center of the scale, Equation 6 correctly predicts that the curves should increase in both slope and spread as the length of acquaintance of both sources is increased.

The nonparallelism of the curves cannot be accounted for by any of the models without elaboration. As written, all of the models predict parallelism. The curves in each panel show a divergent Adjective 1 by Adjective 2 interaction which is statistically significant, $F(4,196) = 10.02$; $MS_e = 3.36$. The divergent interaction has also been obtained with unmodified adjectives and is not attributable to the rating scale (Birnbaum, 1974). There is also some evidence for small higher order interactions, in which the divergent Adjective 1 by Adjective 2 interaction is greater when the sources are more credible. The divergent interaction could be interpreted in terms of Equation 6, by postulating that the weight of an adjective depends upon scale value or upon the stimulus configuration, with the lower valued item receiving greater weight (Birnbaum, 1974).

## DISCUSSION

It is interesting that a small set of simple assumptions can give a nice account of some rather complicated data obtained in two different judgment situations. In Experiment I, the credibility of a source of information about used cars depends on his mechanical expertise. In Experiment II, credibility of a source of personality information depends on the length of the source's acquaintance with the person to be judged. Differences in source credibility are represented by differences in absolute weights (w) that amplify (multiply) the value of the information the sources provide. Each piece of information is represented by a point on a value continuum (s). For used cars, the estimates have value on a monetary dimension; for impression formation, the adjectives can be represented by values on a likeableness continuum. The data of both experiments are consistent with the hypothesis that source credibility serves as an amplifier of information (Figures 1, 2, 3A, 3B).

The assumption that the subject weights each piece of information and strikes a balance (average) leads to an important prediction supported by the data of both experiments: when two sources provide information, the effect of one communication is directly related to the weight of its source (Figures 2A and 3B) and *inversely* related to the weight of the other source (Figures 2B and 3C). Consequently, these data are

inconsistent with Equations 1, 4, and 5 and the model of Rosenbaum and Levin (1968, 1969), which imply that the effect of one piece of information is independent of the credibility of the other source. The data are consistent with the relative-weight model which posits that the relative weight of information provided by one source is inversely related to the number and weights of other sources. Relative weighting thus provides a simple description of an interesting contextual effect.

The experiments of Rosenbaum and Levin (1968, 1969) were not designed to test Equation 6, and consequently neither report has the appropriate cells in the experimental design to allow the present analyses. By combining data from the two studies (a questionable procedure), a graph similar to Figure 4 can be made. The combined data of Rosenbaum and Levin had the important feature of Figure 4 that favors relative weighting: the greater the weight of a source, the less the impact of the other source.

Wyer (1974) presented subjects with pairs of adjectives provided by different sources. In 40 of 48 comparisons where the credibility of the source providing *less* extreme information was increased, the extremity of the response also increased. It is important to note that the averaging model can account for this result. For example, if $s_0 = 11$, $s_1 = 15$, $s_2 = 18$, $w_0 = 2$, $w_1 = 1$, and $w_2 = 1$, then $R = 13.75$. Increasing the weight of the less extreme information ($w_1 = 2$) makes the judgment more extreme ($R = 14$). Therefore, this directional effect cannot be used to test between adding and averaging models. Wyer also found no significant effect of the ratio of scale values on this effect. However, because of the small sample, limited experimental design, and the use of different adjectives in different cells of the design (thus increasing the noise in the data), failure to obtain a significant effect may be attributed to lack of power. Hence, Wyer's (1974) experiment titled a "case against averaging" should be considered nondiagnostic.

It is interesting that a similar discrepancy from the averaging model is obtained in both studies. Figures 2C and 4 show divergent interactions consistent with those obtained in previous studies of impression formation and morality judgment (Birnbaum, 1972, 1973, 1974; Riskey & Birnbaum, 1974): given one piece of information of low or unfavorable value, the other piece of information has less effect. The model could be elaborated to account for this interaction by allowing differential or configural weighting of lower valued items (Birnbaum, 1974).

Mathematical models of attitude formation provide a useful framework for the discussion of everyday sources. In the present studies, the source was presumed to affect the weight parameter only. Since these sources differed with respect to mechanical expertise or length of acquaintance, they can be thought of as differing in *reliability*. Real-life sources may differ not only with respect to reliability but also with respect to *bias*. A biased source has an ax to grind; for example, Egypt would tend to underestimate her own aircraft losses and exaggerate the number of enemy planes shot down. There may also be configural effects; for example, if Egypt reported losses that exceeded Israeli claims, the report might have increased credibility. Further study of the cognitive algebra of bias and reliability seems a promising direction for the experimental analysis of social judgment.

## REFERENCES


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). New York: Academic Press, 1974.

BIRNBAUM, M. H. Morality judgments: Tests of an averaging model. *Journal of Experimental Psychology*, 1972, **93**, 35-42.

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 Monograph*, 1974, **102**, 543-561.

PETERSON, C. R., & BEACH, L. R. Man as an intuitive statistician. *Psychological Bulletin*, 1967, **68**, 29-46.

RISKEY, D. R., & BIRNBAUM, M. H. Compensatory effects in moral judgment: Two rights don't make up for a wrong. *Journal of Experimental Psychology*, 1974, **103**, 171-173.

ROSENBAUM, M. E., & LEVIN, I. P. Impression formation as a function of source credibility and order of presentation of contradictory information. *Journal of Personality and Social Psychology*, 1968, **10**, 167-174.

ROSENBAUM, M. E., & LEVIN, I. P. Impression formation as a function of source credibility and the polarity of information. *Journal of Personality and Social Psychology*, 1969, **12**, 34-37.

WYER, R. S. *Cognitive organization and change An information-processing approach*. Potomac, Md: Lawrence Erlbaum, 1974.


(Received for publication May 23, 1975;
revision received September 26, 1975.)