ORGANIZATIONAL BEHAVIOR AND HUMAN DECISION PROCESSES **51**, 313–343 (1992)

Distributional Theories of Impression Formation

BARBARA A. MELLERS

University of California, Berkeley

VIRGINIA RICHARDS

University of Pennsylvania

AND

MICHAEL H. BIRNBAUM

California State University, Fullerton

This paper examines theories of impression formation which represent stimulus information as distributions on subjective dimensions rather than points along a continuum. These theories explain why unfavorable information often overrides favorable information. Subjects were asked to imagine hypothetical persons described by single adjectives or adjective combinations and to estimate the probability that each person would have various degrees of likeableness. Three models were considered to describe how the likeableness distribution for the adjective combination depends on the distributions of the single adjectives. All three models assume that the mean of the distribution for the adjective combination is described by the equal probability criterion which, for symmetric distributions, implies a weighted average of the single adjective means with weights inversely proportional to their standard deviations. The models can be distinguished on the basis of the standard deviations of the adjective combinations. Estimated standard deviations of the single unfavorable adjectives were smaller than those for the single favorable adjectives. Furthermore, the standard deviation associated with an adjective combination tended to fall between the standard deviations of the single adjectives, consistent with the horizontal averaging model.

INTRODUCTION

A common finding in research on evaluative judgment is that unfavorable information overrides favorable information. For example, a person with one very unfavorable trait will be rated as dislikeable, despite having favorable traits (Birnbaum, 1974, 1982; Fiske, 1980; Hodges, 1974; Ka-

Requests for reprints should be sent to Barbara Mellers, Department of Psychology, University of California, Berkeley, CA 94720. The authors thank Julie Shaffer for helpful discussions, Shi-jie Chang and Lisa Ordonez for assistance with data analyses, and two anonymous reviewers for comments and suggestions. This research was supported by an NSF grant to the first author (BNS-8451368).

0749-5978/92 $3.00

nouse & Hansen, 1972; Ronis & Lipinski, 1985; Schmidt & Levin, 1972). Similarly, a person who has committed a very bad deed will be rated as immoral, despite having done a number of very moral deeds (Birnbaum, 1972a, 1972b; Riskey & Birnbaum, 1973).

This differential effect of stimulus information often results in a divergent interaction between one factor and another. In other words, two favorable adjectives are rated as likeable, two unfavorable adjectives are rated as dislikeable, and the discrepant combinations (a favorable and an unfavorable adjective) are rated as dislikeable, only slightly more likeable than two unfavorable adjectives. Such results are inconsistent as parallel-averaging models (Anderson, 1974) but could be accounted for by a variety of alternative explanations (Birnbaum, 1974).

One theory of the divergent interaction is that subjects average the information but use the response scale in a nonlinear fashion. However, when proper tests are performed, the evidence is not consistent with this hypothesis (Birnbaum, 1974, 1982). Another account that has received considerably more attention is that adjectives are weighted differentially. In particular, less favorable evidence receives greater weight than favorable or neutral evidence.

Ronis and Lipinski (1985) note that, in the impression formation literature, two general classes of weighting theories have been proposed. In one class of theories, the weight associated with an adjective is thought to be related to its subjective value. For example, the congruity model (Osgood & Tannenbaum, 1955) assumes that more extreme information is weighted more heavily. Second, the novelty model (Fiske, 1980) asserts that rare, novel, or otherwise distinctive information is weighted more heavily. Third, the range model (Birnbaum, 1974) assumes that the weight of an adjective depends on the rank order of its subjective value within the particular adjective combination. In impression formation, lower-valued information receives more weight than the higher-valued information.

The other class of theories assume the weight associated with an adjective depends on its ability to reduce uncertainty (Wyer, 1973, 1974; Birnbaum, 1970, 1972a, 1972b; Ronis & Lipinski, 1985). This paper considers the power of uncertainty models to explain why lower-valued adjectives might be given greater weight than higher-valued adjectives. This paper extends those models to describe the uncertainty of adjective combinations.

Concept Identification Model

Wyer (1973, 1974) proposed a concept identification model in which the relationship between a trait and a category is thought to be probabilistic. Each stimulus is represented as a distribution of values over a set of equally spaced response categories. According to the concept identification model, the judged likeableness of a target person based on a single adjective is the expected value of the adjective distribution over all categories. Wyer (1973) reported that the expected values of judged distributions are found to be positively correlated with single likeableness ratings.

The concept identification model also asserts that the dispersion of the probability distribution reflects the subject's uncertainty about the rating. Using a measure of uncertainty from information theory, Wyer reported a positive correlation between judged uncertainty and estimated uncertainty based on the adjective distribution and a curvilinear relationship between estimated uncertainty and estimated likeableness. More extreme likeableness ratings in either direction were associated with greater certainty.

The concept identification model also makes predictions about the distribution of likeableness ratings based on two adjectives. The distribution of the adjective combination is assumed to be described by the overlap in the single adjective distributions. For each category value, the distribution of the adjective combination is assumed to be the smaller of the two probabilities, normalized to sum to 1.0. If no overlap occurs, one of two simplifying assumptions is made. Either the more ambiguous information is assumed to be discounted or the single adjective means are averaged.

According to the concept identification model, the likeableness rating of an adjective pair can be approximated as the expected value of the overlap. Wyer (1973) found that ratings of the likeableness of adjective pairs correlated highly with the expected values of the overlap in the single adjective distributions. He pointed out that if negative traits have smaller variances, the model predicts that the likeableness value associated with a favorable and an unfavorable adjective combination lies closer to the unfavorable adjective than the favorable one. Thus, the model can, in principle, account for the divergent interaction found with adjective combinations.

Equal Probability Criterion

A similar line of ideas was concurrently developed by Birnbaum (1970, 1972a, 1972b, 1974) who proposed a model in which each stimulus is represented by a latent distribution on a subjective likeableness continuum. This distribution represents the set of possible impressions one might have when presented with the stimulus. In this view, people are not only seen as learning a single value for each word, but are also theorized to retain information about variability, or uncertainty. Birnbaum's (1972a) treatment differed from Wyer's (1973) in that Wyer's theory operates on the observed scale, whereas Birnbaum postulated an underlying subjective continuum. Birnbaum proposed several rules to describe the judgment of someone described by two pieces of information, including a rule that Mellers (1986) referred to as the equal probability criterion.

Birnbaum's (1972a, 1972b) distributional model is related to but different from Thurstone's (1927) theory of comparative judgment. Thurstone proposed that on any single occasion, a stimulus generates a sensation within the subject. Repeated occasions of the same stimulus elicit different sensations, and sensations for a given stimulus are represented as a distribution over occasions. In contrast, Birnbaum theorized that the subject has access to the entire distribution associated with the stimulus on any single occasion. That is, the subject could report a distribution of values associated with a stimulus on any given trial.

In Birnbaum's framework, the mean of the distribution represents the average sensation associated with the stimulus, and the variance reflects uncertainty. For example, imagine a person described as "intelligent." Intelligent people may often be likeable, but some clearly are not. This adjective may evoke a distribution of likeableness values with a reasonably high mean and a wide variance that reflects a relatively large degree of uncertainty. On the other hand, imagine a person described as "cruel." This adjective might have a distribution with a low mean and a narrow variance, since cruel is presumably more diagnostic than intelligent.

Birnbaum (1972a, 1972b) proposed the equal probability criterion as a model of how the judged likeableness of a person described by two single adjectives can be predicted from the adjective distributions. When two adjectives describing a hypothetical person are presented for judgment, the subject is assumed to choose a likeableness response, Y^*, such that the probability that the target individual is *more* likeable given the less favorable adjective is equal to the probability that the person is *less* likeable given the more favorable adjective. This notion can be expressed as

$$P(Y > Y^*|X_1 = x_1) = P(Y < Y^*|X_2 = x_2), \tag{1}$$

where Y is the random variable, likeableness; x_1 is the value of the first stimulus, X_1; x_2 is the value of the second stimulus, X_2.

When making a likeableness judgment based on two adjectives, there are two types of "errors." One could rate the target person as either too likeable given the negative information or too dislikeable given the positive information. The equal probability criterion asserts that the subject selects the value that makes the probabilities of these two errors identical. It is easy to imagine situations, however, in which the costs of one error might be larger than the other and subjects might shift the criterion accordingly.

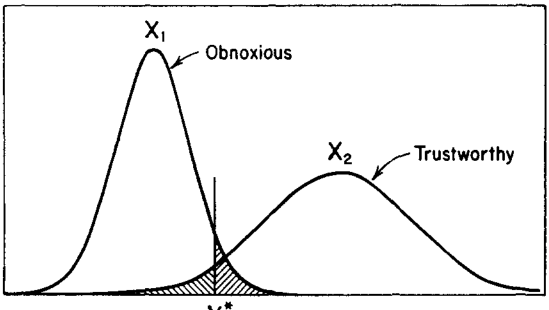

FIG. 1. Illustration of the equal probability model. Stimulus X_1 has a lower mean and a smaller variance than stimulus X_2. Y^* is that value for which the shaded areas in the tails of the distributions are equal.

Figure 1 illustrates the equal probability criterion with two distributions. X_1 represents the trait "obnoxious" and X_2 represents the trait "trustworthy." The horizontal axis depicts subjective likeableness, and the vertical axis represents probability density. The trait obnoxious has a lower mean and a smaller variance than the trait trustworthy. The combined likeableness of both adjectives, Y^*, is that value on the abscissa such that the shaded area in the right tail (given the distribution of obnoxious) equals the shaded area in the left tail (given the distribution of trustworthy). The response to the adjective pair, Y^*, falls closer to the stimulus with the smaller variance, which in this case is obnoxious.

If the distributions are symmetric (as in Fig. 1), the standard scores in the two distributions associated with Y^* are equal but opposite in sign (they sum to zero). Y^* can then be expressed as

$$Y^* = (s_1/\sigma_1 + s_2/\sigma_2)/(1/\sigma_1 + 1/\sigma_2), \tag{2}$$

where σ_1 and σ_2 are the standard deviations of distributions, X_1 and X_2, having means, s_1 and s_2. Equations (2) is an averaging model in which scale values are the means of the distributions and weights are the inverses of the standard deviations. With more than two stimuli with symmetric distributions, the sum of the standard scores is still zero, and Y* will be a simple extension of Eq. (2), namely $Y^* = (\Sigma\, s_i/\sigma_i)/(\Sigma\, 1/\sigma_i)$.

The above analysis shows that the equal probability interpretation provides the basis for an averaging model with differential weights. As Mellers (1986) noted, Birnbaum's equal probability criterion can be either more general or more specific than the differential weight averaging model (Anderson, 1971). For example, when the distributions are asymmetric, the equal probability criterion does not necessarily imply a differential weight averaging model. It is more specific than differential weighting because it requires a particular theory connecting the distributions to the integrated impressions and would be refuted by evidence, for example, that items with greater variance have greater weight than items with smaller variance.

Birnbaum (1972a, 1974) pointed out that distributional theories could explain why overall likeableness ratings are low when either adjective is unfavorable, if unfavorable adjectives have smaller variances than favorable adjectives. In such cases, the lower-valued adjective would have greater weight. If adjective variance increases with the favorableness of the adjective, a person described by two neutral traits could be judged as more likeable than a person described by a favorable and an unfavorable trait.

Birnbaum (1972a, 1974) investigated the idea that unfavorable adjectives have smaller dispersions in two ways. First, he estimated scale values and dispersions using successive interval techniques (Torgerson, 1958) and found that scale values were positively correlated with dispersions. However, Birnbaum (1974) cautioned that because these dispersions were produced by variability among subjects, they may not be the best estimate of subjective distributions generated by a single subject. Birnbaum (1972a, 1974) also asked subjects to rate the range of likeableness values associated with a given stimulus. These values were highly correlated with mean ratings of likeableness. Similar results were found for moral judgments (Birnbaum, 1972b).

Overview

The present paper has two objectives. First, it tests the premise that unfavorable adjectives have smaller variances than favorable ones using a technique that may have advantages over those previously employed. Second, the paper investigates extensions of the equal probability criterion that predict the distribution of the combination from the distributions of the individual components. If an unfavorable adjective has a small variance and a favorable adjective has a larger one, what is the distribution associated with a favorable and unfavorable adjective combination?

Measuring Adjective Distributions

In the present experiments, subjects were told to imagine 100 people described by either a single adjective or a pair of adjectives, and they were asked, "Of those 100 people, how many would you dislike very much, dislike, feel neutral toward, etc.?"

Figure 2 illustrates hypothetical responses to the single adjectives and the adjective pair obnoxious and trustworthy. At the top left, a histogram representing a likeableness distribution for the adjective "malicious" is

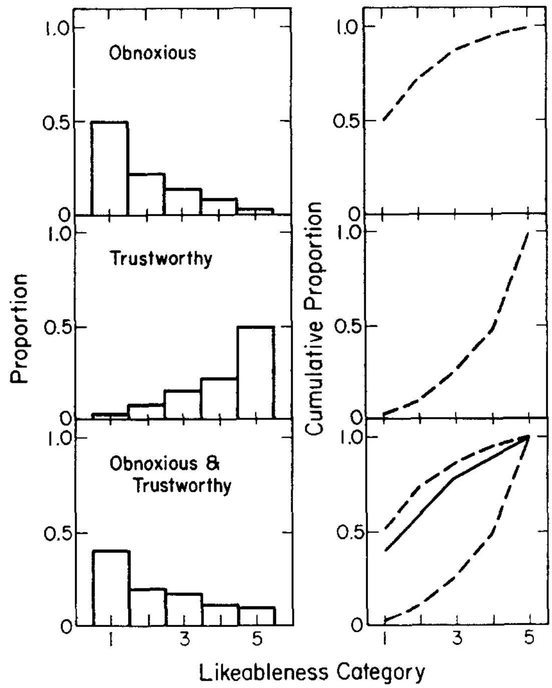

FIG. 2. Hypothetical likeableness distributions for the adjectives Obnoxious (top), Trustworthy (middle), and Obnoxious and Trustworthy (bottom). Cumulative proportions are shown on the right.

shown. Likeableness categories on the abscissa range from "dislike very much" (Category 1) to "like very much" (Category 5). In this case, 50% of malicious people would be "disliked very much," 23% would be "disliked," 17% would be "neutral," 7% would be "liked," and only 3% would be "liked very much." Cumulative proportions are shown in the upper right panel.

The center panels show hypothetical responses to the adjective trustworthy. Fifty percent of people described as trustworthy are liked very much and only 3% are disliked very much. The cumulative distribution is shown on the right. The lower panels depict the distribution of likeableness of persons who would be described as both obnoxious and trustworthy. The solid line on the right represents the cumulative distribution for the combination of traits; dashed lines represent the distributions of likeableness of persons described by each individual trait. The next section describes how different theories would predict the distribution of the adjective combination (solid line) from the distributions of the two single adjectives (dashed lines).

The Distribution of the Adjective Combination

Three models that predict the distribution of the adjective pair will now be considered. The first model implies that the variability or the uncertainty in the combination is typically *larger* than the variances of the single components. The second model asserts that the uncertainty associated with the combination is typically *bounded* by the variances of the single components. The third model implies that the uncertainty of the combination is typically *smaller* than the variances of the single adjectives.

All of the models share three initial assumptions. First, single adjective distributions are assumed to be described by logistic functions with different means and standard deviations. The logistic is a bell-shaped distribution that is similar to the normal distribution but is more convenient to express mathematically. The cumulative distribution for a single adjective can be written as

$$F_1(x) = 1/(1 + e^{-(x-s_1)/\sigma_1}), \tag{3}$$

where s_1 is the mean of the distribution for a single adjective and σ_1 is a dispersion parameter that is related to the standard deviation by a constant of proportionality. With these continuous distributions, there will always be some overlap on the subjective continuum for the distributions of likeableness given the individual adjectives. This approach differs from the discrete distributions assumed by the concept identification model (Wyer, 1973). The concept identification model takes the distributions at face value and requires special assumptions when no overlap occurs. Values of x in Eq. (3) represent points along the subjective continuum. When Eq. (3) is treated as a model and fit to data with a finite number of likeableness categories, values of x are estimated parameters that represent boundaries between one category and the next (category limens).

Second, the present models allow a distribution for the initial impression. Research on impression formation has shown that two good traits are typically judged as more likeable than one good trait, and two bad traits are less likeable than a bad one. This "set size effect" led to the theory that an initial impression (or subjective value based on no specific information) is averaged into the overall judgment (Anderson, 1967). The natural extension of this idea to the present approach treats the initial impression as a distribution of values. This initial distribution can be thought of as the subject's prior beliefs about the probability of liking people given no information about them.

Third, all three models assert that the mean of the distribution of the adjective combination is determined by the equal probability criterion. That is, the sum of the standard scores associated with the single adjective distributions and the distribution of the initial impression is zero. This assumption, in conjunction with the other two, implies that the mean of the distribution for the adjective pair is a weighted average of the means for the single adjectives and the initial impression with weights that are the inverses of the standard deviations.

To highlight the differences among the models, they will first be illustrated under simplified conditions in which two single adjective distributions have identical standard deviations, and the initial impression will have zero weight. In a later section, the models will be presented in a more general form.

Vertical averaging. One way that subjects could combine the two single adjective distributions is to "pool" them into one larger distribution, analogous to a statistician combining the observations from two samples to form a larger sample. If two samples have the same number of observations, then the pooled distribution will have a mean equal to the average of the separate means. The variance of the combination will typically be larger than the variance of either single distribution, since the pooled variance includes both within-group and between-group variability.

Psychologically, this model implies that each adjective is an independent source of information about the distribution of the combination. The difference between the means of the single adjectives is also evidence of variability over and beyond the variability in each single distribution.Thus, when a person is described by two very different traits, vertical averaging implies that there is even more uncertainty in likeableness than the uncertainties in persons described by a single adjective.

This model can be written

$$F_c(x) = (F_1(x) + F_2(x))/2, \tag{4}$$

where F_c is the cumulative distribution for adjective combination at likeableness level, x; F_1 and F_2 are the cumulative distributions for the two single adjectives. Graphically, this model implies that the cumulative distribution for the adjective combination is the vertical average of the cumulative distributions associated with the two single adjectives.

Figure 3A illustrates this model. Dashed lines show curves for two single distributions; the solid line shows the pooled distribution, which is the vertical average of the two separate distribution functions. Choosing three points on the abscissa, x, y, and z, corresponding to the first adjective, the second adjective, and the adjective pair, respectively, with the same cumulative probabilities ($F_1(x) = F_2(y) = F_c(z)$), the derivative (or slope) of the function for the adjective combination will typically be less than the derivatives of the single adjectives ($F_c'(z) < F_1'(x)$ and $F_2'(y)$).

Horizontal averaging. Another possibility is that subjects behave as though they combine likeableness values associated with the two single

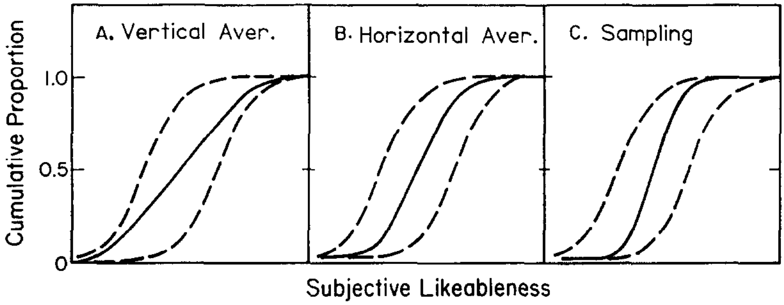

FIG. 3. Illustration to show how the distribution for the adjective combinations (solid lines) depends on the two single adjective distributions (dashed lines) for the three models. Single adjective distributions are assumed to differ in means but to have identical variances.

adjective distributions for each level of probability. This theory can be written in terms of the cumulative distribution functions by stating that for adjective levels, x and y, selected such that $F_1(x) = F_2(y)$,

$$F_c((x + y)/2) = F_1(x) = F_2(y). \tag{5}$$

Graphically, the model implies that the combined distribution is the horizontal average of the two separate cumulative distributions.

This process of combining individual probability distributions to obtain a group distribution has been referred to as "vincentizing" (Ratcliff, 1979). Thomas and Ross (1980) have shown that for certain distributions (e.g., normal and logistic), when the distributions are combined by vincentizing, the combination belongs to the same family as the individual distributions and has parameters that are the average of the individual parameters. If the single adjective distributions are logistic distributions, the combined function is also a logistic with mean and standard deviation equal to the averages of the means and standard deviations for the single adjectives, respectively.

Psychologically, this model implies that when a person is described by two very different traits, the variability or uncertainty associated with likeableness of the adjective combination is a weighted average of the variabilities associated with the single adjectives. Adjectives with greater uncertainty receive less weight. In this case, the uncertainty associated with the adjective combination is typically greater than the most informative adjective but less than the least informative adjective.

Predictions for the horizontal averaging model are shown in Fig. 3B. If the two single distributions have identical variances, the combined distribution has the same variance as either single distribution. In this case, for any three points on the abscissa, x, y, and z, with the same cumulative probabilities ($F_1(x) = F_2(y) = F_c(z)$), the derivatives of the slopes are identical for all three curves.

Sampling. Still another possibility is that subjects act as though they are using a strategy that can be represented by the sampling distribution of the mean. If the two individual distributions were normally distributed and independent, the distribution of the sample mean would be normally distributed with a mean equal to the mean of the means and a standard deviation equal to the standard error of the mean.

According to this model, subjects are seen as taking evidence to reduce their uncertainty about an underlying parameter in a fashion analogous to a statistician's sampling distribution of the mean. The variance or the uncertainty of the combination is typically smaller than the variances of the single adjective distributions. This model implies that one can decrease uncertainty by combining adjectives which are themselves inherently uncertain. The sampling model can be contrasted with the vertical averaging model in that for the sampling model, a wealth of fuzzy information leads to subjective certainty, whereas in the vertical averaging model, when the evidence disagrees, the variability becomes greater.

Predictions for this model are approximated by logistic functions and are shown in Fig. 3C. For any three points on the abscissa, x, y, and z, with the same cumulative probabilities ($F_1(x) = F_2(y) = F_c(z)$), the derivative (slope) of the curve for the adjective combination is greater than the derivatives of the curves for the two single adjective distributions ($F_c'(z) > F_1'(x)$ and $F_2'(y)$).

Generalizations of the Models

All three models can be generalized to allow for unequal adjective variances and an initial impression. The vertical averaging model can be written

$$F_c(x) = (1/\sigma_0 F_0(x) + (1/\sigma_1 F_1(x) + (1/\sigma_2 F_2(x))/ (1/\sigma_0 + 1/\sigma_1 + 1/\sigma_2), \tag{6}$$

where $F_0(x)$ and σ_0 are the cumulative distribution and the dispersion parameter associated with the initial impression. The horizontal averaging model can be expressed in terms of the cumulative distribution functions by stating that if $F_0(x) = F_1(y) = F_2(z) = k$ (for likeableness values x, y, and z), then the cumulative distribution for the adjective pair is

$$F_c((x/\sigma_0 + y/\sigma_1 + z/\sigma_2)/ (1/\sigma_0 + 1/\sigma_1 + 1/\sigma_2)) = k. \tag{7}$$

The generalized sampling model implies that the mean of the distribution for the adjective pair is a weighted average of the single adjective means with weights equal to the inverses of the standard deviations. The standard deviation of the combination is the weighted standard error of the mean.

Figure 4 illustrates the models when the unfavorable trait (dashed line on the left) has a lower mean and a smaller variance than the favorable trait (dashed line on the right). In this case, all three models imply that the curve for the adjective combination (solid line) will be closer to the unfavorable trait on the left, since unfavorable information receives greater weight. The models can be distinguished on the basis of the slopes associated with the distribution of the adjective combination. Appendix A provides the equations used to describe these models.

Figure 5 illustrates the generalized models once again in terms of their probability density functions. Vertical averaging is in the left-hand columns, horizontal averaging is in the center columns, and sampling is in the right-hand columns. Upper panels show two hypothetical adjective distributions and a distribution for the initial impression. The unfavorable adjective (S_L) has a lower mean and a smaller variance than the favorable adjective (S_H). The distribution of the initial impression falls between the two adjective distributions and has the largest variance (S_0).

Center panels show hypothetical distributions for single adjectives after incorporating the initial impression. According to vertical averaging (A2), these distributions have larger standard deviations than the individual distributions. Horizontal averaging implies that the standard deviations of these distributions are bounded by the standard deviations of the individual distributions (B2). Sampling implies that these standard deviations are smaller than those of the original distributions (C2).

The bottom three panels show distributions of likeableness for persons described by a favorable *and* an unfavorable trait. Once again, the vertical averaging model implies that there is more variability in this distribution

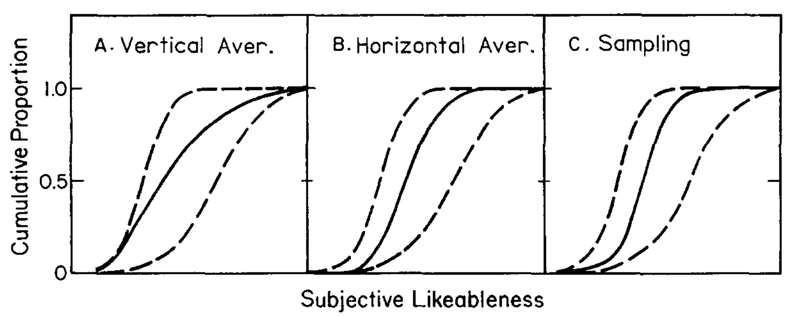

FIG. 4. Illustration of the three models when single adjective distributions are assumed to differ in both means and standard deviations and a distribution of the initial impression is included.

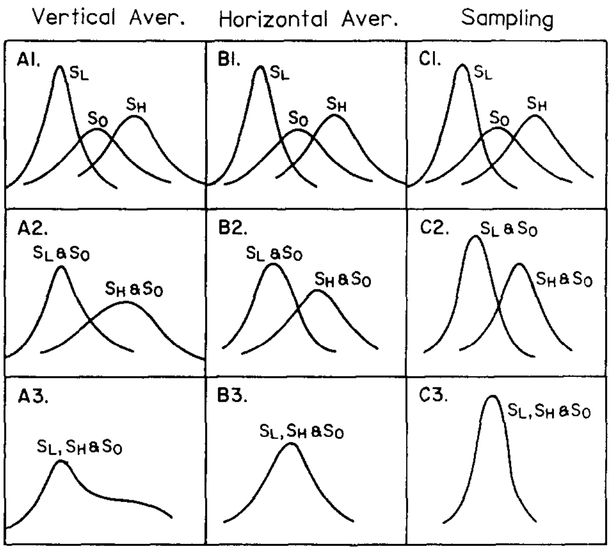

FIG. 5. Illustration of the three models using probability density functions rather than cumulative density functions. Upper panels show hypothetical distributions for a very unfavorable adjective (S_L), an initial impression (S_0), and a very favorable adjective (S_H). The center panels show distributions based on a single adjective (combinations of the original adjective distribution and the distribution of the initial impression); the lower panels show distributions based on the adjective combination (a combination of all three distributions).

than in the single adjective distributions (A3). Horizontal averaging implies that the variability is a weighted average (B3). Sampling implies that there is less variability than in the single adjective distributions (C3).

METHOD

There were two experiments with different subjects in each. In the first experiment, subjects read descriptions of persons described by single adjectives or adjective pairs. Then they judged the probability that each person so-described would have various degrees of likeableness. In the second experiment, subjects read from one to six adjectives and judged the probability that such people would have different degrees of likeableness.

Instructions

Instructions read (in part) as follows: "Imagine a person described as 'intelligent.' You may like some people called 'intelligent' more than others. . . . To express your opinion, you will be asked, of 100 people,

each described by an adjective (or pair of adjectives), how many would fall into each of the categories below:
1 = Dislike very much
2 = Dislike
3 = Neutral
4 = Like
5 = Like very much"
Subjects were instructed to distribute 100 points to represent the percentages of people of each description who would fall into each category.

Stimuli and Designs

In the first experiment, the stimuli were composed of six sets of adjectives used by Birnbaum (1974), listed in Table 1. The five levels in each set (A or B) are designated by the labels VL (very low), L (low), M (medium), H (high), and VH (very high). The adjectives had been selected from Anderson's (1968) list of 555 common personality traits. Within each set, the five levels of likeableness were separated by approximately 1.28 on scale from 0 to 6.

Adjectives were presented either in pairs or alone. Each replicate had 35 trials (A × B, A alone, and B alone), made up from a 5 × 5 Adjective A by Adjective B factorial design for each replicate in Table 1.

In the second experiment, there were 27 trials. Seventeen trials consisted of single adjectives; 6 trials had 1 VL adjective, 5 trials had 1 M adjective, and 6 trials had 1 VH adjective. Seven additional trials were homogeneous sets: 3 VL, 5 VL, 6 VL, 5 M, 3 VH, 5 VH, and 6 VH. The remaining three trials were heterogeneous sets: 1 VH and 5 VL, 3 VH and 3 VL, 5 VH and 1 VL. Adjectives were a subset of those used in Experiment 1.

There were two versions of the second experiment. Both versions included 21 of the same trials (6 trials with 1 VL; 6 trials with 1 VH; 5 trials with 1 M, 6 VL, 5 M, and 6 VH). The remaining five trials used different adjectives for 3 VL, 5 VL, 3 VH, 5 VH, and the heterogeneous sets.

Procedure

In Experiment 1, each subject was given two warm-up trials (2 VL and 2 VH) and a set of 35 test trials presented on a single page in random order. Single adjectives and adjective pairs were interspersed together. Subjects were told to check to make sure their responses summed to 100 after every trial. The experiment was run in groups of subjects ranging in size from one to five and took approximately half an hour.

The same procedure was used in Experiment 2. There were two warm-up trials (VL and VH). Single adjectives and adjective combinations were

TABLE 1
ADJECTIVE REPLICATES

	Replicate 1		Replicate 2		Replicate 3	
	A	B	A	B	A	B
VL	Phony	Mean	Deceitful	Cruel	Malicious	Obnoxious
L	Squeamish	Listless	Untidy	Angry	Unaccommodating	Noisy
M	Blunt	Solemn	Impulsive	Unpredictable	Changeable	Shy
H	Informal	Lighthearted	Studious	Refined	Practical	Self confident
VH	Sincere	Trustworthy	Honest	Dependable	Loyal	Understanding

mixed together in a random order. Half of the subjects were randomly assigned each version. The experiment was run in a large group and took approximately 20 min to complete.

Participants

The 135 undergraduates in Experiment 1 received credit in a lower-division psychology course. There were 88 subjects from the University of Illinois, Urbana–Champaign, with about 30 different subjects in each adjective replicate. In addition, there were 47 subjects from the University of California at Berkeley, with about 15 different subjects in each replicate. Students received credit in a lower division psychology course for their participation. Differences between these two groups of subjects were trivial, and the data were pooled. The 59 undergraduates in Experiment 2 were volunteers from an upper-division psychology course at UCB.

RESULTS

Experiment 1: Distribution of the Adjective Combination

The points in Fig. 6 show cumulative proportions averaged over subjects. Since all three replicates showed a similar overall pattern, the data are averaged over replicates and adjective sets as well. Each of the 15 cells presents data for the adjective pairs (solid points) and the corresponding single adjectives (open points) plotted as a function of likeableness categories where 1 is dislike very much, and 5 is like very much. The dashed and solid lines are the predictions of the horizontal averaging model and will be discussed in more detail later.

Homogeneous combinations. The left-most panels in each row are homogeneous adjective pairs. The open points represent the distribution for the single adjective, and the solid points are the distribution of the combination of adjectives. The bottom left panel shows the distributions for one and for two very unfavorable traits (VL and 2 VL). On the average, 46% of the people described by one very unfavorable trait would be disliked very much. 28% would be disliked, and the remaining 26% would fall in the other three categories. Persons described by two very unfavorable adjectives are even less likeable. According to the subject's judgments, 56% of the people described by two very unfavorable traits would be disliked very much, 28% would be disliked, and the remaining 16% would fall in the other three categories.

The top right panel presents data for one and for two very favorable adjectives (VH and 2 VH). The open points show that with one very favorable trait, 2.5% would be disliked very much, 35% would be liked, and 35% would be liked very much. Persons described by two very fa-

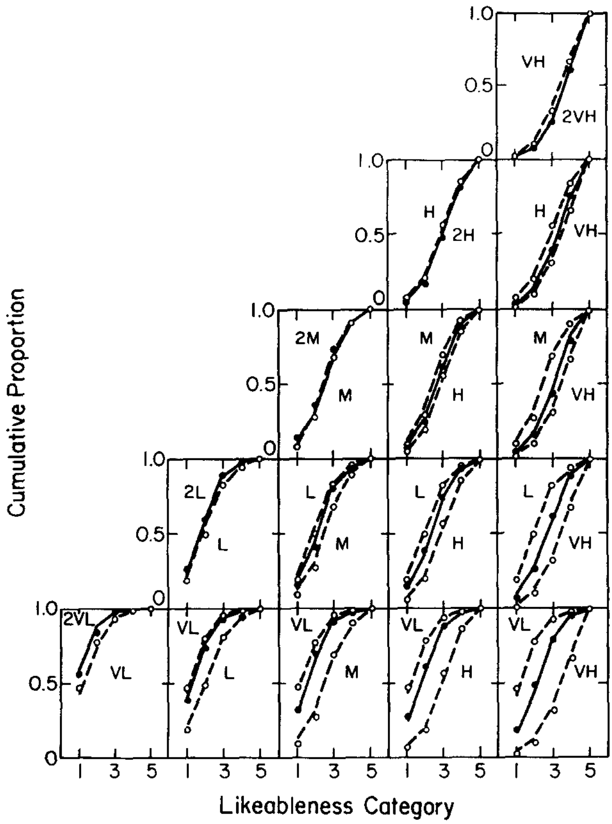

FIG. 6. Cumulative proportions averaged over subjects, replicates, and adjective (row–column). Open and solid points represent the single adjectives and adjective pairs, respectively. Dashed and solid lines are the predictions of the horizontal averaging model. VL, L, M, H, and VH refer to likeableness levels for the adjectives (very low, low, medium, high, and very high).

vorable traits are judged as more probable to be liked than persons described by one. The solid points show that 2% of persons with two VH traits would be disliked very much, 36% would be liked, and 40% would be liked very much.

The relation between the open points and the solid points is called the set size effect. In the bottom left panel, solid points lie to the left of the open points indicating that two very unfavorable traits (2 VL) are more dislikeable that one unfavorable trait (VL). On the other hand, two very favorable adjectives are better than one: open points are to the right of the solid points. All three of the models can account for the set size effect when a distribution of the initial impression is allowed. This effect cannot be accounted for by the concept identification model.

Heterogenous combinations. Off-diagonal panels display data for heterogeneous adjective combinations together with the two related single adjectives. For example, the lower, right-hand panel shows results for the most favorable (VH) and least favorable (VL) adjectives. Open points represent the single adjective distributions; solid points represent the distribution for the adjective pair.

There are two important points to notice about the heterogeneous combinations. First, in all of the off-diagonal panels, the distribution for the pair (solid points) lies between the distributions for the two separate adjectives (open points), consistent with the idea that the distribution of the combination is some type of average of the two component distributions. Second, the distribution for the adjective pair tends to lie closer to the less favorable adjective (on the left), with a few exceptions. In two of the panels (VH and H; VH and M) the points for the adjective pair lie closer to the more favorable adjective (VH) than the less favorable one.

In sum, the adjective distributions exhibit many of the same effects that have been found with single ratings—a set size effect for homogeneous adjectives, an averaging effect for heterogeneous adjectives, and a negativity effect or a tendency for unfavorable adjectives to override favorable adjectives.

Test of the concept identification model. According to the concept identification model (Wyer, 1973), means and standard deviations associated with single adjectives and adjective pairs can be directly computed from the data. Predictions for the distributions associated with the adjective pairs can be derived from the single adjective distributions. This model was fit to the data using the single adjective distributions as parameters (20 average responses). Predictions for the distribution of the adjective pairs were calculated from the overlap.

Data used in the analysis were the average cumulative proportions shown in Fig. 6, excluding the last cumulative proportion (which is necessarily 1.0). There were 15 adjective combinations (with four cumulative proportions for each) and 5 single adjectives (with four cumulative proportions for each). This model fit three times worse than the horizontal averaging model that will be discussed later. In addition, there were a number of systematic deviations. Predicted slopes for the heterogenous adjective combinations were systematically too steep. In other words, predicted standard deviations of the adjective pairs generated from the overlap were too small.

In sum, the concept identification model could not account for the set size effect with the simplifying assumptions that Wyer (1973) made to generate predictions. If these simplifying assumptions are modified in such a way that the number of individuals in each scale category who are believed to possess both attributes is assumed to increase with the extremity of the scale, then the model could, in principle, account for set size effects (Wyer, personal communication). However, there does not seem to be any psychological basis for such an assumption.

Test of the three extensions of the equal probability criterion. Unlike the concept identification model, the present extensions of Birnbaum's equal probability criterion assert that (1) the underlying adjective distributions are continuous, latent logistic functions and (2) the category boundaries may not be equally spaced. These assumptions imply that the standard deviations cannot be directly computed from the data, but must be estimated as parameters of the models that are fit to data.

All three models were fit to the data by means of computer programs that found parameter estimates to minimize the sum of the squared deviations between observations and predictions of the model. All of the programs used Chandler's (1969) subroutine STEPIT to accomplish the function minimization.

Each of the models required 14 estimated parameters to account for the 80 mean responses. With five categories, there are four category boundaries to consider (values of x in Eq. (3)). These boundaries are not assumed to vary and have constant positions with respect to all of the adjectives. They are estimated rather than fixed at 1.5, 2.5, etc., to avoid any assumptions about their relation to subjective likeableness, e.g., linearity. The first was arbitrarily fixed to 1.0, and the other three boundaries were estimated. Five means and five spread parameters (values of s_i and σ_i in Eq. (3)) were also estimated for the five levels of adjectives. In addition, a distribution for the initial impression was estimated with its standard deviation arbitrarily fixed to 1.

The horizontal averaging model fit the data better than the other two models. Residual variance between data and predictions from this model was less than 1% of the systematic variance in the means. The proportion of variance unaccounted for in the sampling model was twice as large, and for the vertical averaging model, deviations were five times as large as the horizontal averaging model. Fits of the three models to data for each adjective replicate showed the same pattern; horizontal averaging accounted for the data best in all three replicates with less than 1% of the systematic variance remaining.

Predictions for the horizontal averaging model are plotted in Fig. 6 together with the data (open and solid points). Predicted curves for single adjectives (dashed lines) and adjective pairs (solid lines) lie close to the data points; deviations from the model do not appear to be systematic for either the single adjectives or the adjective pairs.

Figure 7 compares the predictions of the three models for one of the panels in Fig. 6 (VH and VL). Open and solid points depict data; dashed and solid lines show predictions. Vertical averaging implies that the stan-

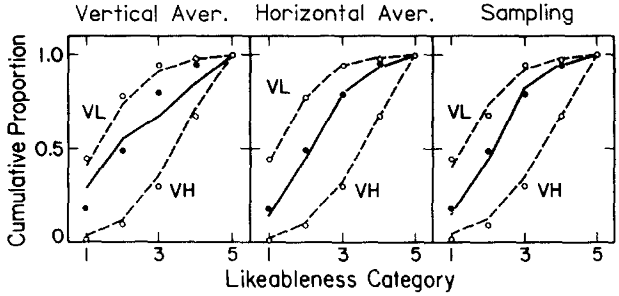

FIG. 7. Data and predictions for distributions with one VL (very low) adjective on VH (very high) adjective, and both adjectives together. Data are open and solid points for single adjectives and adjective combinations, respectively.

dard deviation of a heterogeneous combination of adjectives is *larger* than the standard deviation of the single adjectives. The best-fit solution (left panel) has a slope for the distribution of the adjective pair that is systematically too flat. The sampling model predicts that the standard deviation of the adjective combination is *smaller* than the two single adjective distributions. The best-fit solution has a slope for the adjective combination that is slightly too steep (right panel). The horizontal averaging model implies that the standard deviation of the adjective combination should fall *between* the standard deviations of the individual distributions. This model gives the best account of the data (Fig. 7, center panel).

Figure 8 presents estimated distributions for the single adjectives according to the horizontal averaging model. Unfavorable adjectives have both lower means and smaller variances. The estimated means for the five levels of likeableness are −0.35, 1.44, 2.72, 3.77, and 5.66. Estimated standard deviations are 0.71, 0.81, 0.91, 1.01, and 1.09. The mean and standard deviation for the initial impression are 3.43 and 1.00, respectively. Estimated category boundaries are 1.0, 2.27, 3.74, and 5.20, suggesting that category size increases slightly with likeableness. That is to say, finer judgmental distinctions are made at the lower end of the subjective continuum.

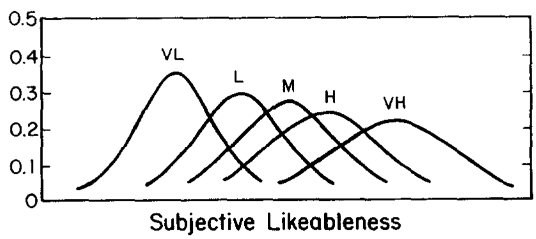

FIG. 8. Distributions for single adjectives according to the horizontal averaging model.

TABLE 2
ESTIMATED MEANS AND DISPERSION PARAMETERS FOR SINGLE ADJECTIVE DISTRIBUTIONS FROM HORIZONTAL AVERAGING MODEL

	Replicate	VL	L	M	H	VH
1	Mean	−0.31	1.01	2.99	3.83	5.25
2	Mean	0.01	2.03	2.66	4.19	6.35
3	Mean	−0.91	1.34	2.34	3.41	5.64
1	Dispersion parameter	0.67	0.76	0.67	0.90	0.90
2	Dispersion parameter	0.71	0.74	1.00	1.13	1.35
3	Dispersion parameter	0.83	0.94	1.14	1.02	1.10

Table 2 shows the estimated means and standard deviations from the horizontal averaging model fit separately to each replicate with different adjectives and different subjects. For all three replicates, means and standard deviations increase with adjective likeableness. The means of the distributions for the initial impressions were 3.02, 3.55, and 3.84 for the three replicates, and the standard deviations were fixed to 1.0.

Individual subject analyses. Because averaged data may not represent the individual subject results, all three models were fit to individual subject data. The horizontal averaging model fit better (had a smaller average squared error) than the other two models for 63% of the subjects. The vertical averaging model and the sampling models fit the data best for 21 and 16% of the subjects, respectively. Percentage of variance unaccounted for by the horizontal averaging model at the individual subject level ranged from 1 to 18%, with a median value of 3%. In addition, estimated standard deviations showed the pattern predicted by the equal probability criterion. For the majority of subjects, the estimated standard deviation of the VL adjective was smaller than that of the VH adjective.

Although the horizontal averaging model fit the data better than the sampling or the vertical averaging models, the sampling model did not do too poorly (see Fig. 7). To provide a more strenuous test between sampling and horizontal averaging, Experiment 2 uses distributions with more than two adjectives.

Experiment 2: Distributions with More than Two Adjectives

Homogeneous sets. Figure 9 shows distributions of homogenous adjective combinations varying in size for VL adjectives (in the upper panels) and VH adjectives (in the lower panels). Open points are mean responses, and dashed lines are predictions of the horizontal averaging model, which will be discussed later. Notice that in the upper panels, a larger proportion of people are disliked very much when they are described by six VL traits than when they are described by one VL trait

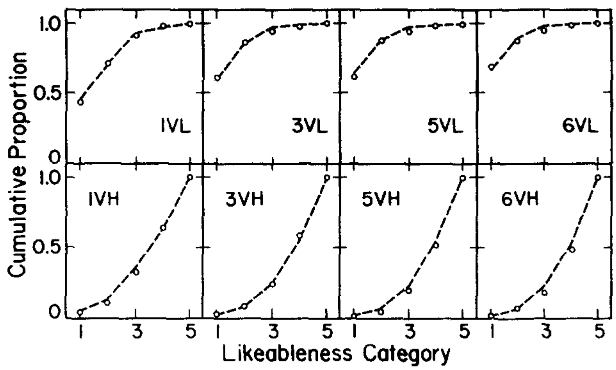

FIG. 9. Mean responses in Experiment 2 with 1, 3, 5, or 6 VL (very low) or VH (very high) adjectives. Open points are data; dashed lines are predictions of the horizontal averaging model.

(44% versus 69%). Similarly, in the lower panels, a larger proportion of people are liked very much when they are described by six VH traits versus one VH trait (35% versus 51%). The concept identification model is unable to account for these set size effects.

Fit of the models. All three models were fit to the mean responses. Distributions for different single adjectives of the same level were averaged together, making a total of 52 mean responses (3 single adjectives, VL, M, and VH, and 10 adjective combinations with 4 responses for each). All three of the models required 10 estimated parameters: three means (values of s_i in Eq. (3) for VL, M, and VH), three dispersion parameters (values of σ_i), three category boundaries (values of x), with the smallest arbitrarily fixed to 1.0, and a mean for the initial impression. The dispersion parameter for the initial impression was fixed to 1.0. Since the single adjective distributions were based on more responses than the distributions of adjective combinations, squared residuals were weighted according to the number of single adjective distributions of the same level that were averaged together (6 for VL, 5 for M, and 6 for VH).

The horizontal averaging model fit the data the best, with 1.4% of the variance unaccounted for. The sum of squared deviations for the vertical averaging model was twice as large, and it was three times as large for the sampling model. Estimated parameters for the horizontal averaging model are shown in Table 3. Lower-valued adjectives have smaller means and smaller dispersion parameters, as in Experiment 1. Estimated category boundaries were 1.0, 1.80, 2.91, and 3.97, showing the same pattern as in Experiment 1, with slightly larger spaces between the categories as likeableness increases.

Predictions for the horizontal averaging model are shown as dashed lines in Fig. 9. This figure shows that the model can account for the VL

TABLE 3
ESTIMATED PARAMETERS FROM THE HORIZONTAL AVERAGING MODEL EXPERIMENT 2

	VL	M	VH
Mean	0.49	2.45	4.01
Dispersion parameter	0.49	0.61	0.77

and VH adjective distributions of sizes ranging from one to six. Deviations between the predictions of the model and the data do not appear to be systematic.

Figure 10 illustrates the differences among the three models for two distributions based on three homogeneous adjectives and a heterogeneous combination of six adjectives. Open points represent homogeneous distributions with either 3 VL or 3 VH adjectives. Solid points depict the heterogeneous distribution with 3 VL and 3 VH adjectives. Predictions from the three models are shown as dashed and solid lines with the vertical averaging model on the left, the horizontal averaging model in the center, and the sampling model on the right. The best-fit solution for the heterogeneous adjective combination (solid line) is too flat for the vertical averaging model and too steep for the sampling model (left and right panels, respectively). However, deviations between the data and the predictions of the horizontal averaging model do not appear to be large or systematic.

Individual subject analyses. Each of the three models was fit to the individual subject data. In Experiment 2, more subjects gave responses in which all 100 points were assigned to a single category than in Experiment 1. Only 2% of all the responses were single category responses in Experiment 1, but 11% were point estimates in Experiment 2. Subjects who tended to give point estimates (19 of 59) were not used in the analyses because the distributional models require variability. For the remaining 40

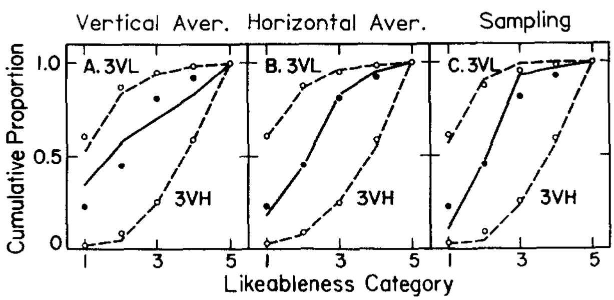

FIG. 10. Data and predictions for distributions with 3 VL (very low) adjectives, 3 VH (very high) adjectives, and all 6 adjectives. Data are open and solid points for homogeneous adjective distributions and heterogeneous adjective combinations, respectively.

subjects, 48, 37, and 15% were best fit by the horizontal averaging model, the sampling model, and the vertical averaging model, respectively.

DISCUSSION

The present results are encouraging for a distributional approach to the representation of stimuli. First, the patterns of responses are consistent with previous experiments that used ratings of likeableness. Distributional responses show the set size effect for homogeneous sets, the combination of adjectives for heterogeneous sets, and the interaction between adjectives with unfavorable adjectives overriding favorable ones. Second, the data are internally consistent and show systematic trends that can be modeled. In other words, it is possible to predict the distribution of the adjective combination from the distributions of the single adjectives using the horizontal averaging model. Third, the predicted relation between distributional means and standard deviations is obtained; lower means tend to have smaller standard deviations.

Equal Probability Criterion

All three models considered were based on the assumption that the equal probability criterion describes the mean of the distribution for the adjective combination. When the distributions are symmetric, this criterion implies a special form of the differential weight averaging model. The weight of a stimulus is assumed to be inversely related to its standard deviation. Evidence consistent with the equal probability interpretation of the divergent interaction between adjectives in likeableness ratings is the finding that lower-valued adjectives have smaller dispersions. In both experiments, unfavorable adjectives had lower means and smaller standard deviations, consistent with this explanation.

This differential weight account of impression formation differs from the configural weight account proposed earlier by Birnbaum (1982). Tests between differential weight and configural weight averaging have been discussed by Birnbaum and Stegner (1979), and future work should attempt to distinguish between these representations.[1]

[1] The configural weight model of impression formation can be derived from the assumptions that (1) subjects select an overall likeableness value that minimizes the sum of the squared errors and (2) the loss function associated with each single adjective is asymmetric. Asymmetric loss functions imply that errors in one direction are more costly than errors in the other direction. For example, overestimates of likeableness might be more costly than underestimates. See Birnbaum, Coffey, Mellers, and Weiss (1992) for a discussion of the configural weight averaging model. Configural weighting and the equal probability criterion are related, but the two models have distinct interpretations of the deviations from parallelism observed in impression formation. Equal probability deals with the probability of over- or underestimating likeableness; configural weighting can be understood in terms of the cost, rather than the probability, of errors. Both factors may be at work.

Manipulating Variability

The equal probability model criterion implies that the overall response based on two or more stimuli has different forms, depending on the distributions of the stimuli. When the stimulus distributions are symmetric and the standard deviations are identical across levels of a factor, the equal probability criterion implies that the overall response can be described by a relative weight averaging model. Scale values are the means of the distributions, and weights are the inverses of the standard deviations.

Birnbaum (1976) tested this special case of the model in an intuitive numerical prediction task. Subjects made predictions of a numerical criterion based on two numerical cues. The cues were constructed so that the criterion variance was identical within all levels of each cue, but differed for different cues. Birnbaum found that cues with less variability at each cue value (and higher correlation to the criterion) received greater weight in a relative weight averaging process.

Mellers (1986) investigated intuitive numerical predictions with cues that had different amounts of variability at each level of the factor in order to distinguish between the generic relative weight averaging model and the more restricted equal probability criterion. Cue means and cue variability were unconfounded; cues with low and high means had either small or large variance.

In her experiment, the equal probability criterion took the form of a restricted differential weight averaging model (Eq. (2)). Results were consistent with the equal probability criterion; the effect of a cue was inversely related to its standard deviation. Thus, the equal probability criterion has also been supported in numerical cue prediction tasks where the stimulus distributions have been manipulated rather than measured, as in Birnbaum (1972a, 1972b, 1974) and the present study.

Distribution of the Adjective Combination

In both Experiments 1 and 2, the data were best represented by an extension of the equal probability criterion, namely the horizontal averaging model. According to this model, both the mean and the standard deviation of the adjective combination are a weighted average of the single adjective means and the single adjective standard deviations, respectively, with weights inversely related to standard deviations. This model implies that the uncertainty in an adjective combination composed of a favorable and an unfavorable adjective is a weighted average of the uncertainties associated with the single adjectives.

Effects on the Adjective Distributions

It is interesting to consider some of the factors that might affect the adjective distributions. First, distributions might vary depending on the context or the purpose of the evaluation. For example, likeableness distributions associated with someone described as "Loud and Comical" might change depending on whether that person was a roommate, co-worker, colleague, parent, etc. Second, adjective modifiers might also influence the distributions in predictable ways, as has been found with single adjectives (Cliff, 1957). The distribution associated with a very obnoxious person might have even less variability than that of an obnoxious person. Finally, individual differences may also be accounted for in terms of different distributions. The distribution of likeableness for someone described as "Fastidious" might have a higher mean for a meticulous person than a lax one. Individual differences in attitudes toward in-groups and out-groups might also be captured by changing distributions. In fact, Linville, Fischer, and Salovey (1989) have argued that greater familiarity with a particular group leads to greater perceived variability in the traits that describe the group members.

Some Analogies

Cue diagnosticity model. The cue diagnosticity model proposed by Skowronski and Carlson (1987, 1989) assumes people categorize other individuals according to cues that are available. Furthermore, cues contribute to categorization in a probabilistic fashion. Cues are said to be more diagnostic when they lead to higher perceived probabilities that a person belongs to one category and lower perceived probabilities the person belongs to other categories. If negative behaviors are less characteristic of likeable persons than positive behaviors are of dislikeable persons, then according to this approach, negative cues are more diagnostic than positive cues.

The distributional models can be thought of in the same spirit as the cue diagnosticity model. According to the cue diagnosticity model, negative information is more diagnostic than positive information because it leads to a higher perceived probability that a person belongs to one category and not another. One way of representing these perceived probabilities is in terms of distributions associated with the adjectives. When adjectives with lower means have smaller variance, it is less likely that someone described by an unfavorable trait would be judged very likeable than it is for someone described by a favorable trait to be judged very dislikeable.

Skowronski and Carlson (1987, 1989) argue that in some cases, favorable information may be more diagnostic than unfavorable information. With ability categories, poor performance may be attributed to many factors (e.g., fatigue or lack of motivation), whereas good performance is more diagnostic of ability. Skowronski and Carlson (1987) show that people are judged to have high ability if they are described as performing well on two occasions or if they performed well on one occasion and poorly on another.

Skowronski and Carlson argue that the cue diagnosticity model can account for these results because for ability judgments, good performance is more diagnostic than poor performance. The distributional approach could also describe the results if the distributions associated with good performance were found to have less variability than the distributions associated with poor performance.

Fuzzy logic. Massaro (1987); Oden (1977); Wallsten, Budescu, Rapoport, Zwick, and Forsyth (1986); and others have used principles from fuzzy set theory to help explain how subjects assign items to categories. In fuzzy logic, propositions are not necessarily entirely true or false. Instead, they are depicted as having continuous values, ranging from 0 to 1.0. Truth values represent the degree to which the item matches the category and are analogous to, but should be kept distinct from, probabilities. To say that a penguin is a bird to degree 0.3, does not not imply that there is a 0.3 probability that a penguin is a bird.

The distributional approach has some similarities to the concept of the membership function in fuzzy set theory. A membership function is a rule that assigns to each element in the universe of discourse a number in the closed [0,1] interval. If the category is well defined, such as the set of men named John, the membership function has only a few values in the interval (e.g., 0 or 1). If the category is not well defined, such as the set of bald men, the membership function can take on many values in the [0,1] interval. Although both approaches use functions to represent stimuli, the present paper represents categories as intervals along a subjective continuum, whereas membership functions typically apply to a single category.

General recognition theory. Ashby and Perrin (1988) recently proposed a distributional representation referred to as the general recognition theory to describe recognition and similarity judgments. On any given trial, the perceptual effect of a stimulus is assumed to be random point in a multidimensional space. Over trials, stimuli are described by multivariate normal distributions. As the overlap between a pair of distributions increases, the similarity (and the confusability) between stimuli increases.

Ashby and Perrin (1988) derive connections between the general recognition theory and Euclidean distance models, including INDSCAL (Carroll & Chang, 1970). INDSCAL is a multidimensional theory that allows individuals to differ in the weights they attach to the underlying dimensions. Ashby and Perrin (1988) note that INDSCAL can be thought of as a special case of the general recognition theory in which weights are inversely related to the standard deviations of the stimuli along the dimensions. Thus, the variability of the stimuli along a dimension is inversely related to the importance or attention placed on that dimension.

This special case of the general recognition theory is more closely related to Thurstone's formulation than that of Birnbaum's in that the subjective distribution is constructed over trials, and the subject is not assumed to have access to the entire distribution on any single trial. However, both the distributional approach and this special case of the general recognition theory imply that the weight attached to a stimulus is inversely related to its variability. According to the general recognition theory, variability reflects changes in the percept associated with a stimulus over trials. According to the distributional approach in the present paper, variability reflects the uncertainly associated with the stimulus on a given trial.

Conclusion

The present paper argues that, in some cases, stimuli might be better understood in terms of different subjective distributions along a variety of dimensions rather than as point estimates. In impression formation, the variability associated with trait adjectives is positively correlated with the means—unfavorable traits have lower means and smaller variances than favorable traits. For the majority of subjects, the variability in an adjective combination can be well described by the horizontal averaging model. According to this model, the uncertainty associated with an adjective combination is a weighted average of the uncertainties of the component adjectives, with weights inversely proportional to the standard deviations.

APPENDIX A

Vertical Averaging

The cumulative distribution for a single adjective is expressed as

$$F_1(L_k) = \frac{1/\sigma_0}{1/\sigma_1 + 1/\sigma_0}\left(\frac{1}{1 + e^{-(L_k - s_0)/\sigma_0}}\right) + \frac{1/\sigma_1}{1/\sigma_1 + 1/\sigma_0}\left(\frac{1}{1 + e^{-(L_k - s_1)/\sigma_1}}\right), \tag{1A}$$

where $F_1(L_k)$ is the cumulative proportion for adjective 1 in category k or lower; σ_1 and σ_0 are the spread parameters for adjective 1 and the initial impression; s_1 and s_0 are the means of the distribution for adjective 1 and the initial impression; and L_k is the limen between the kth and the $(k+1)$th boundaries. The cumulative proportion for the adjective pair is

$$F_c(L_k) = \frac{1/\sigma_0}{1/\sigma_0 + 1/\sigma_1 + 1/\sigma_2}\left(\frac{1}{1 + e^{-(L_k - s_0)/\sigma_0}}\right) + \frac{1/\sigma_1}{1/\sigma_0 + 1/\sigma_1 + 1/\sigma_2}\left(\frac{1}{1 + e^{-(L_k - s_1)/\sigma_1}}\right) + \frac{1/\sigma_2}{1/\sigma_0 + 1/\sigma_1 + 1/\sigma_2}\left(\frac{1}{1 + e^{-(L_k - s_2)/\sigma_2}}\right), \tag{2A}$$

where the symbols are as defined above. For this theory, the mean of the distribution for the adjective pair is

$$\frac{s_0/\sigma_0 + s_1/\sigma_1 + s_2/\sigma_2}{1/\sigma_0 + 1/\sigma_1 + 1/\sigma_2} \tag{3A}$$

and the standard deviation is

$$\frac{1}{1/\sigma_0 + 1/\sigma_1 + 1/\sigma_2} \cdot \left[\sigma_0 + \sigma_1 + \sigma_2 + \frac{(s_0 - \mu)^2}{\sigma_0} + \frac{(s_1 - \mu)^2}{\sigma_1} + \frac{(s_2 - \mu)^2}{\sigma_2}\right]^{1/2}, \tag{4A}$$

where μ is defined in Eq. (3A).

Horizontal Averaging

The cumulative distribution for a single adjective is described by Eq. (3) in the text where s_1 is replaced with

$$\frac{s_0/\sigma_0 + s_1/\sigma_1}{1/\sigma_0 + 1/\sigma_1} \tag{5A}$$

and σ_1 is replaced with

$$\frac{2}{1/\sigma_0 + 1/\sigma_1}. \tag{6A}$$

For the adjective pair, s_1 is replaced with Eq. (3A) and σ_1 is replaced with

$$\frac{3}{1/\sigma_0 + 1/\sigma_1 + 1/\sigma_2}. \tag{7A}$$

Equation (3A) describes the mean of the distribution for the adjective pair, and Eq. (7A) is the dispersion parameter (related to the standard deviation by a constant of proportionality).

Sampling

The expression for a single adjective can be described by Eq. (3), where s_1 is replaced with Eq. (5A), and σ_1 becomes

$$\frac{2}{1/\sigma_0 + 1/\sigma_1}. \tag{8A}$$

For the adjective pair, s_1 is Eq. (3A), and σ_1 is

$$\frac{3}{1/\sigma_0 + 1/\sigma_1 + 1/\sigma_2}. \tag{9A}$$

According to this model, the mean of the distribution is described by Eq. (3A), and the dispersion parameter is Eq. (9A).

REFERENCES

Anderson, N. H. (1965). Averaging vs. adding as a stimulus-comparison rule in impression formation. *Journal of Experimental Psychology,* **70,** 394–410.

Anderson, N. H. (1967). Averaging model analysis of the set size effect. *Journal of Experimental Psychology,* **75,** 158–165.

Anderson, N. H. (1968). Likeableness ratings of 555 personality-trait words. *Journal of Personality and Social Psychology,* **9,** 272–279.

Anderson, N. H. (1971). Information integration and attitude change. *Psychological Review,* **78,** 171–296.

Anderson, N. H. (1974). 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). San Francisco: Freeman.

Ashby, F. G., & Perrin, N. A. (1988). A unified theory of similarity and recognition. *Psychological Review,* **95,** 124–150.

Birnbaum, M. H. (1970). *Impression formation: Averaging versus range model.* Paper presented at the Mathematical Psychology Meeting, Miami.

Birnbaum, M. H. (1972a). *The nonadditivity of impressions.* Unpublished doctoral dissertation, Microfilms, University of California, Los Angeles.

Birnbaum, M. H. (1972b). Morality judgments: Tests of an averaging model. *Journal of Experimental Psychology,* **93,** 35–42.

Birnbaum, M. H. (1973). Morality judgment: Test of an averaging model with differential weights. *Journal of Experimental Psychology,* **99,** 395–399.

Birnbaum, M. H. (1974). The nonadditivity of personality impressions. *Journal of Experimental Psychology* (Monograph), **102,** 543–561.

Birnbaum, M. H. (1976). Intuitive numerical prediction. *American Journal of Psychology,* **89,** 417–429.

Birnbaum, M. H. (1982). Controversies in psychological measurement. In B. Wegener (Ed.), *Social attitudes and psychophysical judgment.* Hillsdale, NJ: Erlbaum.

Birnbaum, M. H., Coffey, R., Mellers, B. A., & Weiss, R. (1992).

Carroll, J. D., & Chang, J. J. (1970). Analysis of individual differences in multidimensional scaling via an N-Way generalization of 'Eckart-Young' decomposition. *Psychometrika,* **35,** 238–319.

Chandler, J. D. (1969). Subroutine STEPIT-Finds local minima of a smooth function of several parameters. *Behavioral Science*, **14**, 81–82.

Cliff, N. (1957). Adverbs as multipliers. *Psychological Review*, **66**, 27–44.

Fiske, S. T. (1980). Attention and weight in person perception: The impact of negative and extreme behavior. *Journal of Personality and Social Psychology*, **38**, 889–906.

Hodges, B. H. (1974). Effect of valence on relative weighting in impression formation. *Journal of Personality and Social Psychology*, **30**, 378–381.

Kanouse, D. E., & Hanson, L. R. (1972). Negativity in evaluations. In E. E. Jones, D. E. Kanouse, H. H. Kelley, R. E. Nisbett, S. Valins, & B. Weiner (Eds.), *Attribution: Perceiving the causes of behavior*. Morristown, NJ: General Learning Press.

Linville, P., Fischer, G., & Salovey, P. (1989). Perceived distributions of the characteristics of in-group and out-group members: Empirical evidence and a computer simulation. *Journal of Personality and Social Psychology*, **57**, 165–168.

Massaro, D. W. (1987). *Speech perception by ear and eye: A paradigm for psychological inquiry*. Hillsdale: NJ: Erlbaum.

Mellers, B. A. (1986). Test of a distributional theory of intuitive numerical prediction. *Organizational Behavior and Human Decision Processes*, **38**, 279–294.

Oden, G. C. (1977). Integration of fuzzy logical information. *Journal of Experimental Psychology: Human Perception and Performance*, **3**, 565–575.

Osgood, C. E., & Tannenbaum, P. H. (1955). The principle of congruity in the prediction of attitude change. *Psychological Review*, **62**, 42–55.

Ratcliff, R. (1979). Group reaction time distributions and an analysis of distribution statistics. *Psychological Bulletin*, **86**, 446–461.

Riskey, D., & Birnbaum, M. H. (1973). Compensatory effects in moral judgment: Two wrongs don't make up for a right. *Journal of Experimental Psychology*, **103**, 171–173.

Ronis, D. L., & Lipinski, E. R. (1985). Value and uncertainty as weighting factors in impression formation. *Journal of Personality and Social Psychology*, **21**, 47–60.

Schmidt, C., & Levin, I. (1972). Test of an averaging model of person preference: Effect of context. *Journal of Personality and Social Psychology*, **23**, 277–282.

Skowronski, J. J., & Carlston, D. E. (1987). Social judgment and social memory: The role of cue diagnosticity in negativity, positivity, and extremity biases. *Journal of Personality and Social Psychology*, **52**, 689–699.

Skowronski, J. J., & Carlston, D. E. (1989). Negativity and extremity biases in impression formation: A review of explanations. *Psychological Bulletin*, **105**, 131–142.

Thomas, E., & Ross, B. (1980). On appropriate procedures for combining probability distributions within the same family. *Journal of Mathematical Psychology*, **21**, 136–152.

Thurstone, L. L. (1927). A law of comparative judgment. *Psychological Review*, **34**, 368–389.

Torgerson, W. S. (1958). *Theory and methods of scaling*. New York: Wiley.

Wallsten, T. S., Budescu, D. V., Rapoport, A., Zwick, R., & Forsyth, B. (1986). Measuring the vague meanings of probability terms. *Journal of Experimental Psychology: General*, **115**, 348–365.

Wyer, R. S. (1973). Category ratings as 'subjective expected value': Implications for attitude formation and change. *Psychological Review*, **80**, 446–467.

Wyer, R. S. (1974). *Cognitive organization and change: An information processing approach*. Potomac, MD: Erlbaum.

RECEIVED February 15, 1990