

A Theory of Comparative Response Times and "Difference" Judgments

MICHAEL H. BIRNBAUM

California State University, Fullerton, California 92634

AND

JR-WEN JOU

University of Illinois, Urbana-Champaign, Illinois 61820

In two experiments, subjects performed three tasks: First, they learned associations between names of hypothetical persons and adjectives that described them. Second, they judged "differences" in likeableness between pairs of names. Third, they pressed one of two keys to indicate which person of each pair was more (or less) likeable. Response times showed the traditional distance effect, end effect, and semantic congruity effect. A simple model was developed to describe response times and "difference" ratings, using a single scale of subjective magnitude. The second experiment applied this model to measure the likeableness of persons described by two or four adjectives. This experiment indicated that adjectives do not combine by a parallel-averaging model. Instead, "difference" ratings and comparative response times were consistent with Birnbaum's configural-weight theory, in which the weight that a stimulus receives depends on its relation to the other stimuli to be combined. In the case of impression formation, the less favorable information receives greater weight. Results are consistent with the hypothesis that ratings of "differences" and choice response times are mediated by the same scale of subjective value. © 1990 Academic Press, Inc.

The time required to compare two stimuli on a given dimension depends on the subjective difference between the stimuli on that dimension. The farther apart the stimuli are, the smaller the time required to decide which is greater. For example, as Dashiell (1937, p. 57) put it, "it should take a person less time to choose between something he likes very much and something he likes very little than for him to choose between two things he likes equally well." This phenomenon, sometimes called the "distance effect," has been observed in a large number of experiments involving a variety of continua (Banks, 1977; Buckley & Gillman, 1974; Dashiell, 1937; Link, 1978, in press; Moyer, 1973; Moyer & Dumais, 1978; Marks, 1972; Potts, Banks, Kosslyn, Moyer, Riley, & Smith, 1978).

This research was facilitated by a Hughes Faculty Fellowship. Requests for reprints may be sent to Michael H. Birnbaum, Department of Psychology, California State University, Fullerton, CA 92634.

The original aim of the present research was to utilize the distance effect in comparative response times to derive a scale of subjective magnitude that could be compared with scales obtained in judgment studies. In particular, we planned to use reaction times to provide a scale-free test of the parallel-averaging models of impression formation (Birnbaum, 1974, 1982). In the course of this research, however, it became apparent that theories of response time did not give a complete account of various aspects of response time data, and therefore it was necessary to develop a theory that would describe response time phenomena and also connect response times to judgments. Therefore, our first experiment was designed to clarify the relationships between judgments and response times. Our second experiment utilized the theory developed in the first experiment to accomplish our original goal of utilizing response times to provide a convergent test of theories developed by Birnbaum (1974) for the combination of information from different sources.

PHENOMENA & THEORIES: COMPARATIVE RESPONSE TIMES

Distance Effect

Initially, we intended to fit comparative response times to the following model:

$$T_{ij} = M[|s_i - s_j|] \quad (1)$$

where T_{ij} is the predicted response time to compare stimuli i and j ; s_i and s_j are subjective magnitudes of the stimuli; and M is a strictly decreasing, monotonic function. This model implies a distance effect, because the time to compare two stimuli becomes smaller as the distance between the stimuli increases. This model was discussed by Buckley and Gillman (1974), and Potts et al. (1978), among others, though these authors noticed deviations from Equation 1, especially for comparisons involving the extreme stimuli.

If Equation 1 were correct, then it should be possible to find values of s and a monotonic function, M , to reproduce the rank order of observed response times. In principle, rank order information concerning the times would permit estimation of a scale, s , that would be unique to an interval scale. However, comparisons involving extreme stimuli often deviate from the best-fit solution to Equation 1. Buckley and Gillman (1974) used a nonmetric, multidimensional scaling analysis to show that response times could not be fit in one dimension (as in Eq. 1), but instead required two dimensions. Buckley and Gillman interpreted the second dimension as the consequence of a random walk process. Appendix A illustrates how response time data can violate Equation 1, even though the model allows the function M and the scale values, s_i , to be estimated from the data.

End Effect

The term "end effect" refers to the fact that response times are faster to stimuli near the ends of the series to be compared (Banks, 1977). For example, when comparing integers from 1 to 9, subjects are faster at saying that "2 is more than 1" or "9 is more than 8" than to decide that "5 is more than 4." Moyer and Dumais (1978) use the term "serial position" effect for the same phenomenon.

The end effect or serial position effect has not always been recognized as a threat to Equation 1 for three reasons: First, because the end stimuli are farther on the average from the other stimuli, the end effect (serial position effect) is occasionally interpreted as a consequence of the distance effect, although the two phenomena are distinct (see Moyer and Dumais, 1978, p. 128). Second, when the stimuli have not been scaled independently, the fact that two end stimuli can be compared faster than two stimuli in the middle of the series might be interpreted by Equation 1 to mean that end stimuli are farther subjectively from the others, unless the deviations are of sufficient magnitude to violate Equation 1 (Appendix A). Third, when the task is to compare stimuli from a fixed set, the subject might use a different strategy for end stimuli. For example, when the subject knows that the stimuli are integers from the set from 1 to 9, the subject could respond that "9 is larger" without attending to the other stimulus (Potts et al., 1978), because the subject has learned that 9 is the largest stimulus in the set. These three interpretations allowed investigators to retain Equation 1 in spite of deviations.

Semantic Congruity Effect

A third major phenomenon of comparative response time is that the rank order of response times depends on the polarity of the judgment in the instructions (Shipley, Cottin, & Hadsell, 1945; Shipley, Norris, & Roberts, 1946). Shipley et al. (1945, 1946) found that the time to choose which of two colors is *more* pleasant is faster for two preferred colors than for two unpreferred ones; however, the order of times for these two comparisons is reversed when the task was to choose the *less* pleasant. This phenomenon is called the "semantic congruity effect." Semantic congruity effects have been replicated with a variety of continua, and considerable attention has been given to their explanation (Banks, 1977; Holyoak, 1978; Holyoak & Mah, 1982; Marks, 1972; Moyer & Dumais, 1978; Petrusic & Baranski, 1989; Potts et al., 1978). For example, when comparing digits, subjects are faster at deciding "9 is more than 8" than "2 is more than 1"; however, they are faster at deciding that "1 is less than 2" than "8 is less than 9."

As Holyoak and Mah (1982) noted, Equation 1 would interpret such

changes of rank order as changes in the spacing of the scale values, depending on the semantic wording of the comparison task ("more" vs "less"). Holyoak and Mah (1982) attempted to connect this phenomenon with "difference" ratings, but they concluded that congruity effects obtained in ratings are too small to account for the response time phenomena. Similarly, Marks (1972) proposed a theory in which stimulus discriminability could account for these changes; however, Banks (1977) presented arguments and evidence that pose difficulties for the perceptual, discriminability account of the semantic congruity effect. Indeed, if Equation 1 requires different scales, s , for different ways of asking the same question, one can wonder whether either scale represents a meaningful metric for subjective value.

Random Walks

Link (1978; in press) developed a coherent theory of choice response times and choice proportions in terms of a random walk model. Random walk models have proven useful for connecting phenomena in recognition memory (Ratcliff, 1978), signal detection, and other experimental tasks (Luce, 1986), and they had been proposed for choice reaction time by Buckley and Gillman (1974) and Moyer and Dumais (1978), among others. The model represents choice as the outcome of a unidimensional random walk between two boundaries. In each interval of time, differences between the stimuli are accumulated to drive the walk toward one or the other boundary. When the cumulated evidence reaches one or the other boundary, the subject decides on the corresponding response. The separation of the boundaries is assumed to depend on instructions emphasizing speed or accuracy. In Link's (in press) model, the starting point for the random walk is assumed to depend on the subject's bias (tendency, or expectation) to use one or the other response. By placing the starting point closer to the response expected, the subject can respond on the basis of less evidence (faster) than by placing the starting point midway between the boundaries. The rate of evidence accumulation in the random walk is assumed to depend on the separation of the stimuli.

COHERENT THEORY OF RESPONSE TIMES AND "DIFFERENCE" JUDGMENTS

The present theory was developed from examination of the results of Experiment 1, combined with insights from previous developments cited above. The theory consists of a set of equations to describe the distance effect, end effect, semantic congruity effect, and "difference" ratings, using a single scale of subjective magnitude. These equations are compatible with an extension of random walk theory, and the random walk analogies are discussed in greater detail in Appendix B.

In our experiments, subjects judged the "differences" in likeableness between pairs of persons. Having memorized that *Carl is malicious* and *Bill is loyal* for example, subjects were asked how much more they would like *Bill* than *Carl*. In addition, subjects generated two response times by pressing a key to indicate which person was "more" likeable, and (in a separate block of trials), they indicated which person was "less" likeable. There are therefore three arrays of data to be explained by the theory: judgments of "differences," "more" response times, and "less" response times.

The key equations can be written as follows:

$$D_{ij} = a|s_i - s_j| + b \quad (2)$$

$$TM_{ij} = m_0 + m_i m_j |s_i - s_j| \quad (3)$$

$$TL_{ij} = l_0 + l_i l_j |s_i - s_j| \quad (4)$$

where D_{ij} is the predicted "difference" in likeableness between person i and person j ; TM_{ij} is the predicted response time to choose which is "more" likeable; TL_{ij} is the predicted response time to choose which is "less" likeable; a and b are linear constants describing the linear relationship between "difference" ratings and subjective differences; m_0 and l_0 are additive constants for response time, representing the time required to read the stimuli and to press the key; and m_i , m_j , l_i , and l_j are bias parameters that are proposed to account for both end effects and the semantic congruity effect. Note that the same subjective scale of stimulus magnitude (s_i) is involved in all three equations.

A version of Equation 1 is a special case of Equations 3 and 4 in which the values of m and l are constants. In that case, both types of response times and "difference" ratings would be monotonically related to each other. However, when the bias parameters are not constants, they can produce end effects and semantic congruity effects that would otherwise violate Equation 1. To account for the end effect, the values of m and l are smaller for end stimuli than for middle stimuli; to account for the congruity effect, bias values of m are relatively lower for large stimuli than small stimuli compared with the relative values of l . Multiplication of the bias values implies that the time to compare two stimuli will be faster if *either* stimulus is biased to yield a fast time. The time required to process a given stimulus difference will be long when both stimuli are relatively unbiased (e.g., when both stimuli are near the middle of the series).

Equations 2-4 are compatible with an extension of Link's (1978, in press) random walk model. The additional supposition is that the starting point for the random walk is the product of bias parameters. Using the product asserts that when either one of the stimuli is near the end of the continuum, the subject begins the random walk closer to the correct

choice boundary, which allows a faster response with less accumulation of evidence from the stimulus difference. If the product of bias parameters ($I_i I_j$ or $m_i m_j$ in Equations 3 and 4) corresponds to the starting point of the walk, then Equations 3 and 4 give the average times for correct responses in a random walk model (See Appendix B for further detail). In conjunction with Equation 2, the theory assumes that subjects access the same subjective scale when judging "differences" as the one that mediates the accumulation of evidence in choice response time.

EXPERIMENT 1

Method

Overview of Three Tasks

Each subject performed three tasks: learning, judgment, and choice response time. The learning task was a prerequisite to the other tasks: subjects learned to associate personality trait adjectives with names. For example, they learned pairings such as Fred: *malicious and cruel*, and Bill: *loyal*. In the judgment task, subjects were asked to rate "differences" in likeableness between the persons (e.g., how much more would you like Bill than Fred?). In the response time tasks, pairs of names appeared simultaneously on the terminal screen and subjects pressed one of two keys to indicate which of the two people they liked "more", or which one they liked "less". The order of the judgment and response time tasks was counterbalanced across subjects.

Subjects

Thirty University of Illinois undergraduates, enrolled in Introductory Psychology, took part in the experiment for extra credit.

Learning Task

Materials. The learning list consisted of seven, single-syllable, common, four-letter male first names, each beginning with a different first letter. These were Bill, Carl, Fred, John, Mike, Paul, and Rich. Personality trait adjectives were associated with the seven names. Five of the adjectives had previously been studied by Birnbaum (1974): *loyal, practical, changeable, unaccommodating, and malicious*. Each of these adjectives was randomly associated with one name. The remaining two names were randomly associated with two very extreme personality traits. The positive extreme was *honest and understanding*, and the negative extreme was *malicious and cruel*. Two different mappings of the personalities onto the names were used, with one being the reverse of the other.

Procedure. Subjects memorized the association of each name with a personality. The learning criterion was that they be able to give the personality traits without error for two successive rounds when cued by the experimenter with the names. This part of the experiment averaged about 25 min.

Response Time Task

Design. The seven different names generated 21 distinct comparisons between two names. Each comparison was presented to the subjects once with the two names printed in a certain left-right order, and once with the names reversed. Every pair was also presented with a "like more" question, and with a "like less" question. To discourage fast guessing, the time interval between the termination of the question word and the onset of the pair of

names had two values, 300 ms, or 1000 ms. This waiting time interval had virtually no effect, and it is ignored in subsequent analyses.

The variables of name order in the pair (2 orders), of like "more" or "less" question (2 questions), and of the waiting time interval (2 intervals) were all completely crossed. Thus, the 21 basic pairs were multiplied into a set of $21 \times 2 \times 2 \times 2 = 168$ trials.

Procedure. The subjects were told that they would be presented with pairs of names on the screen. On the basis of what they knew about each person's personality, they were to choose the one that they "liked more" or the one that they "liked less," depending on the question. For half of the subjects, the first half of the series of trials (84 randomly ordered trials) was presented with the "more" question, and the second half of the series with the "less" question, and vice versa for the other subjects. The first 11 trials of each half-series were practice trials.

At the beginning of each trial, a "+" sign was displayed at the center of an LSI-11 computer terminal screen. When ready, the subject could press the space bar to cause the sign to disappear and the first question word ("more" or "less") to be displayed for 500 ms. After the waiting interval (300 or 1000 ms), the first pair of names appeared on the screen. They were asked to rest their left and right middle fingers on the two response keys, and to rest their thumbs on the space bar. They pressed the left side key or right side key spatially corresponding to the position of the name that they wanted to choose. When either response key was pressed, the stimulus was terminated, the "+" sign reappeared on the center of the screen, and subjects could press the space bar to start the next trial. They were instructed to respond as rapidly and accurately as they could.

When the first half of the series was finished, there was an instruction displayed on the screen telling the subjects that from the next trial on, the question word would be switched and that they should respond accordingly. After this instruction, the next 11 trials were practice trials with the new task.

Response latency was measured from the onset of the pair of names to the response that terminated the display of the stimulus. The entire response time task took an average of about 25 min.

Judgment Task

The judgment task consisted of 49 pairs of names generated from a 7×7 , First Person by Second Person, factorial design, in which the same 7 names could appear in each position to be compared with each other and themselves. The 49 trials were presented in random order with the constraint that the same name would not occur on two successive trials.

Each pair was presented in the format of a subtraction problem, e.g., Rich - Carl = ?.

The subjects rated "differences" in likeableness between two persons, using integers from -9 to 9. A rating of 9 meant that the "first person was very much more likeable than the second person." A difference of zero meant that the two persons being compared were "equal in likeableness." Negative numbers were to be used when the first person was less likeable than the second.

The judgment task required about 40 min on the average to complete. The entire experiment took about 1½ h.

Results

Figure 1 shows mean "difference" ratings as a function of the estimated scale value of the more likeable person with a separate curve for each level of the less likeable person, averaged over the two presentation orders. For example, the highest set of judgments (triangles) show the

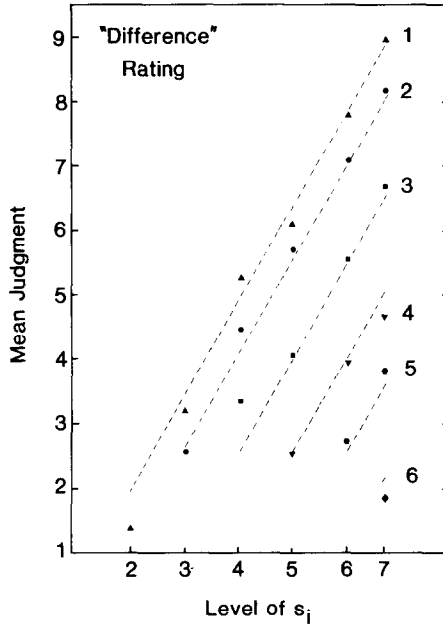


FIG. 1. "Difference" ratings plotted against estimated likeableness of preferred person with a separate type of symbol for each level of the less preferred person. (Levels of likeableness refer to personality traits in Table 1. Dashed curves show predictions.)

mean judgments of "difference" in likeableness of the persons described by levels 2 through 7, compared with the person who was described as *malicious and cruel* (level 1). As the level of the preferred person increases from 2 to 7 (abscissa), the "difference" rating increases. The *malicious* person (level 2 on the abscissa) is rated only slightly better than the *malicious and cruel* person (level 1). The "difference" increases as the description of the preferred person changes to *unaccommodating* (level 3), *changeable* (level 4), *practical* (level 5), *loyal* (level 6), and *honest and understanding* (level 7). Each set of symbols shows the "differences" of the same people compared with a different standard.

Equation 2 implies that curves connecting data symbols in Fig. 1 should be parallel (i.e., the vertical distance between two data curves should be the same for any value on the abscissa). Dashed curves show predictions of a simultaneous fit of Equations 2-4 to all of the data. Because the levels are spaced on the abscissa according to the estimated scale value, the predictions are not only parallel, but also linear (Model fitting is discussed in a later section). Results are compatible with previous studies of "difference" judgments (Birnbbaum, 1974, Experiment 3; 1978, 1980, 1982), in that the empirical curves are approximately parallel, but judg-

ments of small “differences” deviate slightly from parallelism. Exaggerations or underestimations of small “differences” may be attributed to slight nonlinearities in the judgment function that converts subjective differences to numerical responses (Birnbaum, 1980, 1982). Equation 2 approximates the judgment function with a linear function.

Figures 2, 3, and 4 display three phenomena of comparative response time, as illustrated by Banks (1977), for the present results. Symbols show data and dashed lines show predictions of Equations 2–4, which will be discussed below. Figure 2 shows the distance effect: Response times are shown, averaged over pairs of stimuli that are separated by the same number of steps. (For this 7×7 design, there are six times as many observations for a one-step difference as for a six-step difference). Responses are faster for greater ordinal differences, consistent with the typical finding (Banks, 1977; Moyer, 1973; Moyer & Dumais, 1978). For example, Fig. 2 shows that the average time to compare the two most different people (level 1 vs 7) is nearly a second faster than the average time to compare two people who differ by one step.

Fig. 3 shows the end effect: Response times are averaged (over all other factors) for each stimulus and plotted against the stimulus value. Response times are faster on the average for stimuli near either end of the likeableness continuum than for stimuli near the center (levels 3, 4, 5), consistent with previous results for other continua (see Banks, 1977; Moyer & Dumais, 1978).

Fig. 4 shows the semantic congruity effect by plotting response times for comparisons of adjacent stimuli (e.g., levels 1 vs 2, 2 vs 3, etc.), with a separate symbol for each type of comparison. Response times for “more” comparisons (data are shown as squares; predictions are dashed curve) are relatively faster at the high (likeable) end of the scale, whereas response times for “less” comparisons (solid circles and solid curve) are relatively faster at the low (dislikeable) end. For example, the squares show that it takes less time to decide that level 7 is *more likeable* than level 6 than that level 2 is *more likeable* than level 1; however, it takes less

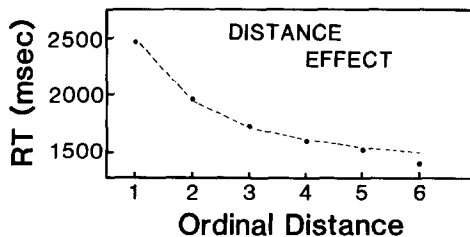


FIG. 2. Distance effect. Response times are faster for comparison of stimuli that are separated by greater differences.

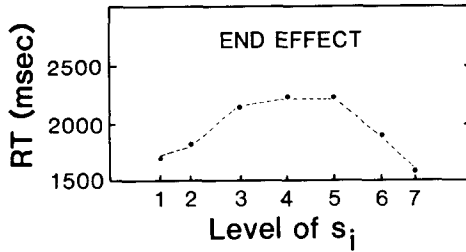


FIG. 3. End effect. Response times are faster to comparisons involving stimuli nearer to the ends of the series.

time to decide that level 1 is *less likeable* than level 2 than to decide that level 6 is *less likeable* than level 7. In summary, the rank order of the times depends on the type of comparison. The three patterns of data in Figs. 2, 3, and 4 are quite compatible with findings of previous research (Banks, 1977; Moyer & Dumais, 1978; Potts et al., 1978), using other continua.

Figure 5 presents a more detailed look at response times. Mean response times for "more" and for "less" comparisons are plotted separately in the right and left panels as a function of the more likeable person with a separate curve for each level of the less likeable person, as in Fig. 1. Note that response times are smaller for pairs that are farther apart, for pairs involving an extreme level, and for pairs that occupy a position on the scale congruent with the direction of the comparison instruction. Note also that the points in Fig. 1 and the two panels of Fig. 5 reveal three different rank orders for the same stimulus comparisons. These three different rank orders are inconsistent with Equations 1 and 2, which would

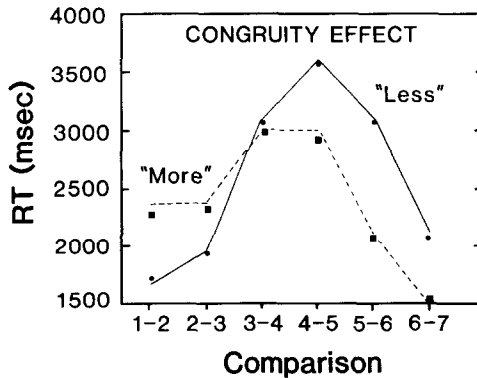


FIG. 4. Semantic congruity effect, illustrated for comparisons of adjacent stimuli. Response times are relatively faster for pairs that are near the end of the scale that matches the comparison word.

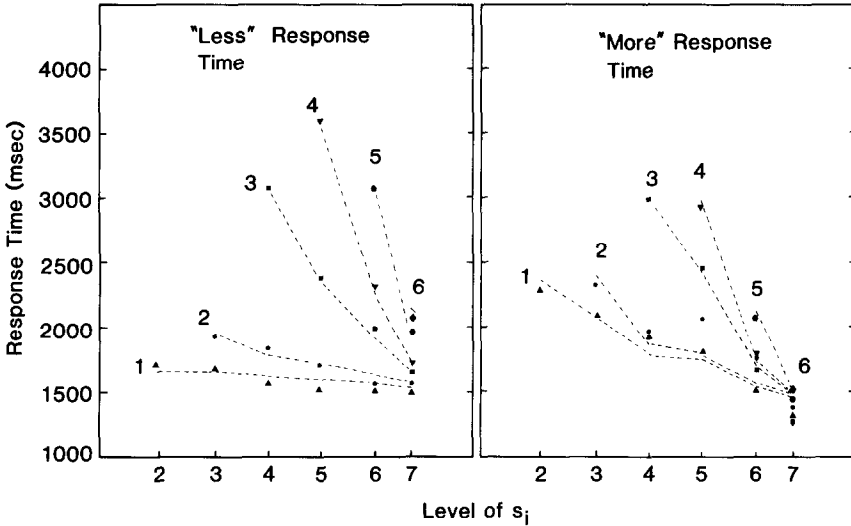


FIG. 5. Mean response times for comparative judgments plotted as in Fig. 1, with "more" times and "less" times shown in right and left panels, respectively.

imply that all three arrays should be monotonically related. However, the fit of Equations 2, 3, and 4 shows that the three sets of data can be reconciled by a single scale of subjective value.

Model Fitting

To fit the model requires the estimation of a single scale of likeableness (s_i) for the seven personalities to simultaneously fit three arrays of data. To accomplish this goal, Birnbaum's (1980) method for simultaneously fitting several sets of data was modified to allow the analysis of response time and judgment data. First, a loss function was defined for each array of data as follows:

$$L_k = \frac{\sum \sum (X_{ijk} - \hat{X}_{ijk})^2}{\sum \sum (X_{ijk} - \bar{X}_k)^2} \quad (5)$$

where L_k is the value of the loss function for data array k (levels of k distinguish the three tasks); X_{ijk} is the data value for cell (i,j) in data array k ; \hat{X}_{ijk} is the corresponding prediction (from Equation 2, 3, or 4); \bar{X}_k is the grand mean response for the k th array; and the summation is over all cells in the k th array for the present experiment. Each value of L_k represents the proportion of variance in that array that deviates from the predictions of the model ($1 - L_k$ is the squared correlation between predictions and data for array k).

For all three arrays ("difference" ratings, "more" times, and "less" times), L_k is a function of the scale values (s_i), as shown in Equations 2-4.

TABLE 1
Estimated Scale Values and Bias Parameters (Experiment 1)

Level	Description	Scale value s_i	"Less" bias l_i	"More" bias m_i
1	Malicious and Cruel	0.92	8.0	30.3
2	Malicious	1.92	16.5	29.4
3	Unaccommodating	3.69	46.5	55.8
4	Changeable	5.38	57.2	45.8
5	Practical	7.09	62.3	57.7
6	Loyal	8.80	43.4	19.3
7	Honest and Understanding	10.00	17.3	0.464

Note. Values of a and b in Equation 2 are .862 and 1.09, respectively. Values of l_0 and m_0 are 1521 and 1469 ms, respectively. Values of s_i are the same in all three equations. (Eqs. 2-4).

In addition, each equation contains some unique parameters (e.g., a and b appear only in Equation 2 for the "difference" array). A special computer program was written to select parameter values so as to minimize the sum of the three values of L_k , utilizing the subroutine STEPIT (Chandler, 1969). The three values of L_k are .02, .06, and .01 for "differences," "more" times, and "less" times, respectively. The best-fit values of the parameters are given in Table 1. The estimated scale values of the five single adjectives are very nearly equally spaced, consistent with previous scaling based on rating methods (Birnbaum, 1974; 1982).

The predicted values, shown as dashed or solid curves in Figures 1-5, appear to give a good account of the various features of the data, shown as points in the figures. Figures 2, 3, and 4 show that the model gives a good approximation of the three major phenomena of response times as well as the more detailed account in Fig. 5. "Difference" ratings are also reasonably well described by the same scale values, as shown in Fig. 1.

EXPERIMENT 2

The success of Equations 2-4 to represent response times and "difference" judgments encourages the use of this model to define a scale of likeableness. Experiment 2 applies this model to our original purpose, to test the parallel-averaging model of adjective combination in impression formation (Anderson, 1974; Birnbaum, 1974). According to the parallel-averaging model, the integrated impression created by a set of adjectives is the average of the components:

$$\Psi_{ij} = \frac{w_0 s_0 + w_1 s_i + w_2 t_j}{w_0 + w_1 + w_2} \quad (6)$$

where Ψ_{ij} is the integrated impression; w_0 , w_1 , and w_2 are weights; s_i and t_j are the scale values of the adjectives; s_0 is the scale value of the initial impression. According to Equation 6, the difference in likeableness due to a change in one adjective should be independent of the value of the other adjective (Birnbaum, 1974). For example, Equation 6 implies that the difference in likeableness between Mike who is *loyal and understanding* vs Paul who is *loyal and obnoxious* should be equal to the difference between Fred who is *malicious and understanding* vs John who is *malicious and obnoxious*. However, Birnbaum (1974) found that when one trait is bad, the difference due to the other trait is smaller than when the trait is good, contrary to Equation 6.

Birnbaum's (1974, 1982) configural-weight theory, on the other hand, predicts that the judged difference between Mike and Paul will be greater than the judged difference between Fred and John even though the personality difference is *understanding vs obnoxious* in each case. Birnbaum's (1974, 1982) configural-weight theory assumes that the effective weight depends not only on the order of presentation, but also the rank order of the stimulus in the set of stimuli presented on each trial. Impression formation and moral judgment appear consistent with the theory that weight is taken from the higher valued stimulus and given to the lower-valued stimulus (see Birnbaum, 1982; Birnbaum and Stegner, 1979, 1981). Further information on configural weighting is given in Appendix C.

If these configural effects in likeableness are truly psychological, and if response times depend on metric differences in likeableness as in the model of Equations 2–4, then Birnbaum's (1974) theory would also imply that the time to choose which is more (or less) likeable should be faster for the greater difference, when the times are adjusted for bias according to Equations 3 and 4. Experiment 2 therefore investigates a combination of theories including Birnbaum's (1974, 1982) theory of impression formation in conjunction with the model (Eqs. 2–4) of response times and "difference" judgments.

Method

The procedures of Experiment 2 resemble those of Experiment 1, with the exceptions that there were 11 names to learn, and each person was described by either two or four adjectives.

Subjects

Twenty-two University of Illinois undergraduates enrolled in Introductory Psychology participated in the experiment for extra credit.

Learning Task

Eleven male names were used, seven of which were the same as in the first experiment, and the additional four names were Dave, Glen, Nick, and Stan. Nine of the 11 names were

assigned to personalities described by two adjectives. The nine two-adjective persons were constructed from a factorial combination of each of three adjectives (*malicious, changeable, or loyal*) with each of three other adjectives (*obnoxious, shy, or understanding*). The remaining two (of the 11) combinations were extremes; one was described by four very favorable personality trait adjectives (*trustworthy, dependable, sincere, and honest*), and the other by four very unfavorable ones (*phony, deceitful, mean, and cruel*).

To facilitate the learning, each name was printed on one side of an index card, and the associated personality traits were printed on the opposite side. Subjects drilled themselves by looking at the name and trying to recall the traits, and turning over the card to check. Subjects learned the association of names with personality traits by quizzing themselves and checking their answers using the drill cards, until they reached the criterion of two errorless rounds.

Reaction Time Task

The nine names that were associated with two-word personality descriptions were used to generate pairs of names for comparison. As in Experiment 1, each name was combined with each of the other names to yield a total of 36 distinct pairs. Each pair was presented once with the two names in one left-right order, and once with the names reversed. Each of these was presented once with the "more" question and once with the "less" question. Hence, each distinct comparison was presented four times, yielding 144 trials [$36 \text{ pairs} \times 2 \text{ (two name orders)} \times 2 \text{ ("more" or "less")} = 144$]. The waiting interval between the termination of the question word and the onset of the two names was fixed to 750 ms.

In addition to the comparisons between names with two-word personality traits, for the first 9 subjects, each of the two extreme names (with four-word personality traits) was used to randomly combine with each other and 4 of the 9 two-word trait names. But for the last 14 subjects, a complete set of pairing of the extremes with the other names was used.

There were 10 practice trials preceding each block of "more" or "less" trials. Other details of the response time procedure were the same as in Experiment 1.

Judgment Task

There were $9 \times 9 = 81$ likeableness "difference" judgments for two-word personality trait names, and 16 trials involving end-anchors for the first 9 subjects. The remaining 14 subjects received a full set of end-anchor pairings with the other items. The judgment procedures were the same as in Experiment 1. The whole experiment, including the judgment task, required about 1 hr and 45 min.

Results

The model of Equations 2-4 was fit to the three arrays of data of Experiment 2, using the same procedures as in Experiment 1. The estimated scale values for the two-adjective combinations are given in Table 2, and they are plotted in the lower panel of Fig. 6 in the same fashion as Birnbaum's (1974) values (upper panel) to facilitate comparison. Birnbaum (1974) obtained mean ratings of the likeableness of persons described by the same pairs of adjectives from 100 subjects. The values estimated from "difference" ratings and from "more" and "less" response times in Experiment 2 for the same adjective pairs are nearly linearly related. Note that both sets of curves in Fig. 6 show similar, divergent interactions.

TABLE 2
Estimated Scale Values and Bias Parameters (Experiment 2)

Level	Description	Scale value s_i	"Less" bias l_i	"More" bias m_i	Birnbaum (1974) Rating
2	Malicious and Obnoxious	2.69	22.2	40.8	1.39
3	Malicious and Shy	4.28	25.5	24.4	2.48
4	Malicious and Understanding	4.74	39.2	47.1	3.20
5	Changeable and Obnoxious	3.75	44.4	50.3	2.56
6	Changeable and Shy	6.53	40.0	35.1	5.17
7	Changeable and Understanding	8.58	60.0	39.0	6.60
8	Loyal and Obnoxious	5.72	43.0	57.0	3.64
9	Loyal and Shy	7.21	46.5	50.4	6.65
10	Loyal and Understanding	9.68	20.9	8.7	7.99

Note. Values of a and b in Equation 2 are .828 and 1.60, respectively. Scale values for the four-adjective extreme sets were 1.28 and 10.0, respectively. Values of l_0 and m_0 are 2257 and 1695 ms, respectively. Values of s_i are the same in all three equations. (Eqs. 2-4).

The values in Table 2 and Fig. 6 show the divergence predicted by Birnbaum's (1974, 1982) configural-weight model. For example, the difference in scale values between *loyal and understanding* and *loyal and obnoxious* is $9.68 - 5.72 = 3.96$, which is almost twice the difference between *malicious and understanding* and *malicious and obnoxious*, $4.74 - 2.69 = 2.05$. These results are inconsistent with the parallel-averaging model (Eq. 6), but are compatible with Birnbaum's (1974) results and configural-weight theory.

Three different empirical results are mutually consistent: The present ratings of "differences" and the relative magnitudes of choice response times are compatible with Birnbaum's (1974) results. For example, the rated "difference" between

loyal and understanding – *loyal and obnoxious*

is rated as a "difference" of 4.48 and the times to respond "less" and "more" are 2.46 and 1.82 s, respectively. However, the comparison between

malicious and understanding – *malicious and obnoxious*

is 3.00, the time to indicate which is "less" likeable is 3.23 s, and the time to indicate which is "more" likeable is 2.96 s. Thus, when the first trait is "loyal," the difference due to *understanding vs obnoxious* is greater and both response times are less (roughly a second faster) than when the first trait is "*malicious*."

Similarly, the rated "difference" between

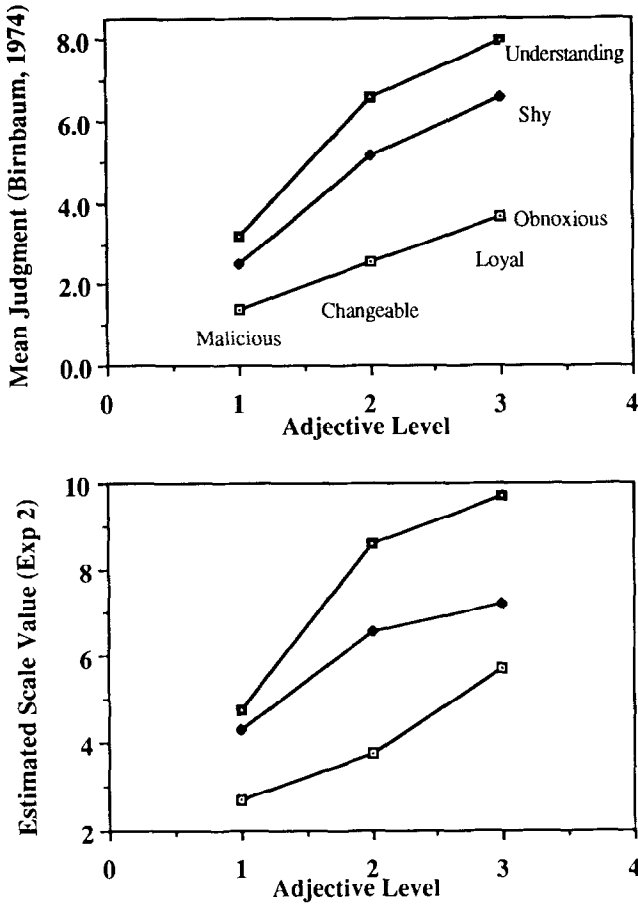


FIG. 6. (Top) Mean ratings of likeableness of persons described by two adjectives (from Birnbaum, 1974, Experiment 1), plotted as a function of adjective A with a separate curve for each level of adjective B. (Bottom) Estimated scale values from model of response times and "difference" judgments, plotted in the same fashion for comparison.

loyal and understanding – malicious and understanding

is rated 5.89; the "less" time is 2.31, and the "more" time is 1.80. However, the "difference" between

loyal and obnoxious – malicious and obnoxious

is rated 4.43; the "less" time is 3.32, and the "more" time is 2.46. Hence the difference between *loyal* vs *malicious* is greater when the second trait is *understanding* (and the response times are less) than when the second trait is *obnoxious*.

Greater separations in Birnbaum's (1974) ratings correspond to greater "difference" judgments and smaller average times to choose. Although response times are not perfectly correlated with subjective distances due to end effects and congruity effects (as in Experiment 1), the estimated scale of subjective value accounts for response times and "difference" judgments, and it is consistent with the interaction obtained by Birnbaum (1974).

In sum, when these data are fit to a model that accounts for "difference" ratings, the distance effect, the end effect, and the semantic congruity effect, the estimated parameter values for the adjective combinations are consistent with Birnbaum's (1974, 1982) findings. Therefore, these data indicate that we can account for "difference" ratings and response times by using the configural-weight averaging model of impression formation, but these data refute the parallelism models of impression formation.

Discussion

The present results encourage attempts to develop a coherent theory of judgment and response times. Equations 2, 3, and 4 can describe difference ratings and the three major phenomena of comparative response time: the distance effect, the end effect, and the semantic congruity effect. The second experiment also shows a compatibility of comparative response times and "difference" judgments for the analysis of information integration. In particular, the parallel-averaging models of impression formation are violated by ratings of likeableness (Birnbaum, 1974), by "difference" ratings, and by response times in a consistent fashion. The configural-weight theory of impression formation remains compatible with the major features of the data.

It is important to emphasize that the present theory is not simply a set of separate equations postulated to account for unrelated phenomena, but that Equations 2-4 connect three arrays of data by the same scale of subjective magnitude (s_i). As noted in the introduction, one interpretation of the semantic congruity effect has been the idea that the subjective magnitudes change, depending on the direction of comparison. However, the present results show that when the theory distinguishes response bias from stimulus difference, it may be able to account for the changing rank order of the response times using a single scale of value, preserving the principle of scale convergence (Birnbaum, 1974, 1980, 1982; Birnbaum & Veit, 1974). The fact that the scale values that describe response times also predict judgments is consistent with the proposition that response times reflect an internal psychophysics (Moyer, 1973; Moyer & Dumais, 1978) and further encourages the construction of coherent theory.

Choice Response Time

Equations 3 and 4 are compatible with a random walk model of comparison. In Equations 3 and 4, bias is represented as the product of bias parameters which are free to vary with the stimulus value and the polarity of the instructions. The bias parameters in Tables 1 and 2 have some consistent properties that encourage the development of theories to account for them. To illustrate how one might theorize about these parameters, consider a 7×7 design such as employed in Experiment 1. We can examine the probability, associated with each stimulus, that the stimulus in the other position will be greater. These probabilities are $P_i = 1, .83, .67, .50, .33, .17,$ and 0 for the stimulus ranks, $i = 1$ to 7, respectively. The uncertainties, defined as $U_i = (P_i)(1 - P_i)$, are then $U_i = 0, .14, .22, .25, .22, .14,$ and 0 for the seven levels. The values of m_i and l_i in Table 1 resemble these values of U_i . This line of reasoning suggests that if the experimental design is altered, the value of m and l would be affected. For example, if the subject's task were to compare integers from 1 to 9, the integer 5 should have a relatively large value of m . However, if the subjects were instructed that the stimuli would only include the integers from 1 to 5, then the value of m for 5 would be very small.

Link (in press) proposed techniques to unconfound stimulus uncertainty and stimulus magnitude, using two variations of experimental procedure and design. First, one can present the two stimuli sequentially instead of simultaneously. With this change in procedure, it is reasonable to theorize that the bias should be completely due to the first stimulus. Second, Link introduced a positive correlation between successive stimuli, designed to unconfound the first stimulus from the difference between the first and second stimuli. These manipulations should change the relationship between the stimulus magnitude and the bias parameter, and Link reported evidence consistent with this prediction. On the basis of analogy with Link's work, it is reasonable to theorize that the bias parameters m_i and l_i in Tables 1 and 2 should not be regarded as measures of the stimuli per se, but as indices that depend on the context provided by experimental design.

Random walk theory yields predictions for choice proportions as well as response times (Link, 1978, in press; Luce, 1986). A worthwhile subject for future investigations would be to explore the relationships between comparative judgments (e.g., "ratios" and "differences") as in Birnbaum (1980, 1982), choice proportions, and choice response times in speeded psychophysical comparison tasks. When there are sufficient errors, the predictions of random walk models for response time for each response can also be examined. As Luce (1986) has noted, there are a

number of alternative versions of random walk models and a variety of implications that can be investigated to test them against their rivals.

Impression Formation

The present data are consistent with the hypothesis that when one adjective is unfavorable, other adjectives have less impact. Configural weight theory provides a simple interpretation, and it remains consistent with a variety of data involving combination of evidence from different sources (Birnbaum & Stegner, 1979, 1981; Birnbaum, 1982; Birnbaum & Mellers, 1983) and psychophysical integration (Birnbaum, 1982). The parallel-averaging models lead to scales that disagree with scale values obtained from "difference" judgments involving the same adjectives (Birnbaum, 1974, Experiment 3) and they cannot explain judgments of "differences" between people who are described by two adjectives (as in Birnbaum, 1974, Experiment 4 and Experiment 2 of the present paper).

There is a connection between configural-weight theory and the dual bilinear representation (Narens & Luce, 1986). The dual bilinear representation is equivalent to a simple configural weight model, the range model (Birnbaum, 1974). Narens and Luce note that the dual bilinear representation arises as the only version (of a fairly general class of models) that leads to interval scales. The interval scale is an advantage of the range model over other possible explanations of the nonparallelism.

Nonconfigural averaging models also fail to explain why enough good deeds cannot eventually overcome bad deeds in the judgment of morality (Birnbaum, 1973; Riskey & Birnbaum, 1974). Furthermore, changes in the judge's point of view can alter the rank order of judgments in a fashion consistent with the predictions of the configural-weight model Birnbaum & Stegner (1979). These arguments, together with the present data, lead to the conclusion that impression formation can be better represented by a configural-weight model in which the lower valued items receive more weight than by parallel averaging or nonconfigural theories.

In conclusion, the present data add further evidence against parallel-averaging models of impression formation, but form a consistent pattern of evidence that encourages development of coherent theory. In this coherent theory, subjective values can be measured on an analog scale, and the same scale can be used to predict ratings of combinations, "difference" ratings, and the times required to decide which is "more" or "less."

APPENDIX A

Equation 1 might seem to be a difficult theory to disprove, since it allows one to estimate scale values from the data and it permits any

monotonic function, M , to convert subjective differences into response times. However, the theory is testable. Table 3 illustrates hypothetical response times that would permit rejection of Equation 1. Each numerical entry represents the value of T_{ij} and each algebraic entry clarifies the interpretation under Equation 1. The letters, a, b, c, \dots , refer to successive differences in subjective values ($a = s_2 - s_1; b = s_3 - s_2; c = s_4 - s_3$; etc; note that $a + b = s_3 - s_1$). Because M is strictly decreasing, $T_{31} > T_{62}$ implies $a + b < b + c + d + e$; therefore, $a < c + d + e$; however $T_{21} < T_{63}$ implies $a > c + d + e$, which contradicts the comparison of T_{31} and T_{62} , according to Equation 1. Table 3 contains many similar contradictions that refute Equation 1, even without prior knowledge of the subjective values.

Information about the spacing of the stimuli would impose further constraints and yield further tests of Equation 1. For example, suppose it were known that the subjective values of the stimuli were equally spaced. The data of Table 3 would violate the assumption of equal spacing because $T_{21} < T_{32} < T_{43} < T_{54}$, whereas Equation 1 would require equal differences to produce equal times. Accordingly, in the present study, we also obtained ratings of "differences" to have an independent scaling of

TABLE 3
Hypothetical Response Times and Theoretical Interpretation of Equation 1.

Stimulus level (i)	Stimulus level (j)					
	1	2	3	4	5	6
2	1700 (a)					
3	1750 ($a + b$)	2500 (b)				
4	1700 ($a + b + c$)	2100 ($b + c$)	4500 (c)			
5	1650 ($a + b + c + d$)	1900 ($b + c + d$)	3000 ($c + d$)	5100 (d)		
6	1580 ($a + b + c + d + e$)	1700 ($b + c + d + e$)	2167 ($c + d + e$)	2700 ($d + e$)	3900 (e)	
7	1533 ($a + b + c + d + e + f$)	1580 ($b + c + d + e + f$)	1750 ($c + d + e + f$)	1900 ($d + e + f$)	2100 ($e + f$)	2300 (f)

Note. These values violate the interpretation of Equation 1. They were calculated from Equation 4, using successive integers from 1 to 7 as the scale values (s_j); the values of l_i were 10, 20, 50, 60, 60, 40, 20; and $l_0 = 1500$ (similar to the estimated parameters in Table 1). Letters in parentheses represent successive differences in scale value; i.e., $a = s_2 - s_1; b = s_3 - s_2; c = s_4 - s_3$, etc. According to Equation 1, response times are inversely related to absolute differences.

subjective distances (Birnbaum, 1978, 1980, 1982). By the principle of scale convergence, both arrays should be reflecting the same underlying psychological values. If Equation 1 were correct and if "difference" judgments are a monotonic function of subjective differences, then response times and "difference" ratings should be monotonically related. Although the data of Table 3 refute Equation 1, they were calculated from the present theory, using equally spaced values.

APPENDIX B

The theory of Equations 2-4 is compatible with random walk models. As has been noted by many authors, there are always many sets of assumptions from which a given conclusion can be deduced; therefore, by Ockham's razor, no particular theory should be given special credence until its new and wider implications have been tested. The Equations of Equations 2-4 constitute a testable theory in their own right. Nevertheless, it is worthwhile to spell out the connections of the present theory to random walk models that have predictive power for a variety of other implications (Link, 1975, 1978, in press; Luce, 1986; Moyer & Dumais, 1978).

The basic analogy states that the process of choosing between two stimuli is like a random walk in one dimension between two boundaries. The analogy to a gambler provides a useful illustration. Consider a gambler who *starts* with \$60 and plans to bet on a series of comparisons between cards chosen randomly (with replacement) from two decks. If the gambler's card (chosen from the gambler's deck) is higher, the gambler wins an amount that is proportional to the difference in the cards; if the opponent's card is higher, the gambler loses in proportion to the difference. For example, if the bet is \$1 per unit, and if the gambler drew a 10 and the opponent drew a 4, then the gambler would win \$6. The gambler's *fortune* starts at \$60, and goes up or down every time as the luck of the draw produces increments or decrements to the gambler's wealth. The gambler has determined to *quit ahead* if his cumulated fortune ever reaches or exceeds \$100 and to *quit behind* if his fortune ever reaches \$10 or less. If the gambler started with higher stakes, (e.g., \$80 instead of \$60) the probability of quitting ahead would be higher, the average number of steps to quit ahead would be less, and the average number of steps to quit behind would be greater. If the cards were dealt to give a better hand to the gambler than the opponent, then the probability of quitting ahead would also be increased and the average time to do so would be decreased. If the gambler were to bet less on each trial or to place the boundaries for quitting (ahead or behind) farther apart, the process would take longer on the average. The gambler's situation can be represented as a random walk, a process that seems analogous to the

process by which subjects decide which of two stimuli has "more" or "less" of an attribute.

In this analogy, the stimuli are represented as distributions of subjective values, analogous to the decks of cards held by the gambler and the opponent. The greater the difference between the distributions (d -prime), the more likely one stimulus will exceed the other on each comparison. The subject is presumed to sample the stimuli in each increment of time and to accumulate the difference between the stimuli, $A-B$. If and when the difference reaches the criterion to say "A is more than B" the subject responds; if and when the difference reaches the other boundary, the subject responds that "B is more than A." Reaching the criteria to decide is analogous to the gambler's fortune reaching the criteria to quit. Deciding A or B is analogous to the gambler quitting ahead or behind. The time for the subject to decide which stimulus is greater is analogous to the number of steps until the gambler quits. The greater the difference between the stimuli ($A-B$) the faster the evidence will accumulate leading to the A response; if $A-B$ is negative, the evidence will accumulate toward the B response. The rate of evidence accumulation is analogous to the difference between the gambler's deck and the opponent's deck. Instructions to the subject to be accurate rather than fast should cause the subject to place the boundaries farther apart, analogous to the gambler placing his quitting points farther apart.

The starting point, or bias parameter, is set by the subject according to beliefs concerning whether A or B will be the correct response on the next trial. The starting point is analogous to the gambler's initial stakes. By starting the walk closer to one boundary or the other, the time required to reach that response can be decreased. Sequential effects and contextual effects should affect the bias parameter. The gambling analogy is useful for discussing the bias parameter although the gambler may have different tendencies from the subject. By choosing a starting point nearer to the criterion to quit ahead, the gambler would reduce the time to quit ahead and increase the probability of quitting ahead, but would decrease the amount won. Therefore, if the gambler could look at the gambler's own deck and realize that the cards were high, the gambler might choose to place the bias parameter far from quitting ahead. On the other hand, the subject desires to respond as quickly as possible; therefore, if the subject knows that one of the stimuli is extreme, the bias parameter can be set to allow a faster response that will be likely to be correct.

In the present model, the starting point for the random walk is represented as the product of bias parameters for each stimulus. In the random walk interpretation, the values of the parameters and the multiplicative representation depend on the use of simultaneous presentation of the stimuli (rather than sequential) and the use of a fixed stimulus set. Simply

put, the bias measure indicates how much evidence will be required to decide for a given stimulus; larger values indicate that more evidence is required.

The random walk model of comparative response time requires no special assumptions concerning the dimensions to be compared. Therefore, the model should take the same *form* for continua such as sizes of animals, magnitudes of numbers, or likeableness of people, but the values of the parameters will depend on the choice of stimuli.

Figure 7 illustrates how a random walk model, combined with certain assumptions, can lead to Equations 3 and 4. The height of the point in the figure represents the accumulated evidence after a given amount of time. In the figure, the boundaries represent the two responses, and are assumed to be fixed for this experiment. When the accumulated evidence reaches a boundary, the subject decides on a response. The top boundary represents the correct response in these cases. The average rate of evidence accumulation (slope in Figure 7) is assumed to be proportional to

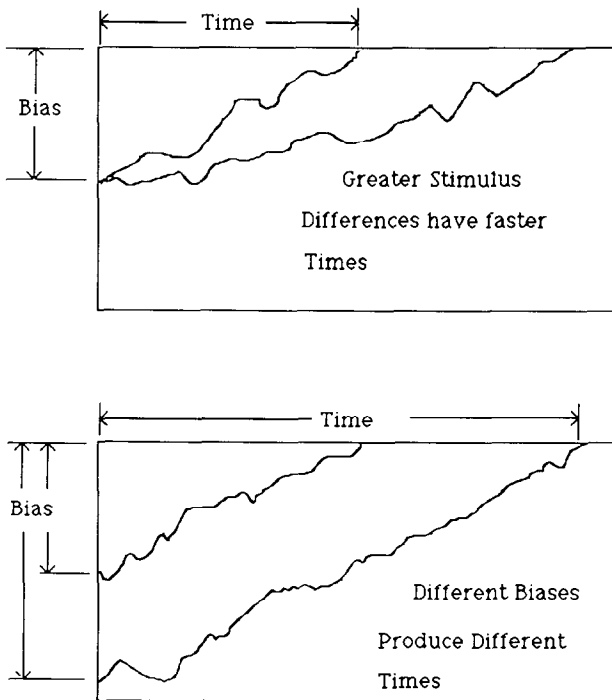


FIG. 7. Illustration of random walk model. (Top) two decisions involving equal bias, but unequal stimulus differences. (Bottom) two decisions involving equal stimulus difference but different biases. Bias and stimulus difference affect response times in different ways in this model.

the absolute difference in scale value ($|s_i - s_j|$). The starting point of the walk (the initial height) determines the subject's bias, which is a measure of how much evidence will be required to be accumulated to reach either response.

The top panel of Figure 7 shows two walks for pairs of stimuli that are equally biased, but vary in stimulus difference. Note that the greater stimulus difference corresponds to a faster accumulation of evidence (steeper slope), which yields a shorter response time. The lower panel illustrates random walks for two comparisons with the same stimulus difference, but different biases. Note that the evidence accumulation rates (slopes) are equal, but because one begins closer to the correct response, the response times are different. These cases illustrate how random walk theory requires a decomposition of bias from stimulus difference: equal differences can produce unequal times and equal times can be produced by unequal differences.

To further clarify the model, it is helpful to consider the following example. Suppose the subject's task is to press the left or right button to indicate whether the left or right person is "more" likeable. Suppose Carl (*malicious*) appeared in the left position. The subject could bias the starting point closer to the right button, because it is a good bet that the person on the right would be more likeable than Carl. Similarly, if Bill (*loyal*) appeared in the right position, then the subject could also bias the walk to begin closer to the right response, because Bill would probably be more likeable than whoever might appear in the left position. These bias effects of the stimuli do not involve comparisons of the stimuli with each other, but depend on the relative positions of the stimuli in the set of stimuli. The product of bias parameters in Equations 3 and 4 indicates that the subject can produce a faster starting point if *either* stimulus allows it. The actual comparison of stimuli proceeds by accumulating differences that lead to one or the other boundary.

Equations 3 and 4 can be understood as follows: (a) *evidence* accumulated = *rate* of accumulation \times *time*; (b) the amount of accumulated *evidence* required depends on the *bias*, as shown in Fig. 7, therefore:

$$\text{time} = (\text{bias})/(\text{rate}) \quad (7)$$

(c) If *bias* (the amount of evidence required) in Equation 7 is the product of the bias parameters for the two stimuli ($m_i m_j$ or $l_i l_j$), and if (d) *rate* of evidence accumulation is the absolute difference in scale value ($|s_i - s_j|$), then Equation 7 implies that Equations 3 and 4 describe the *time* to reach the correct boundary, where the additive constants in Equations 3 and 4 represent the times required for perceptual and motor components that are required on all trials.

In sum, a random walk model is compatible with the theory of Equa-

tions 2-4. In this model, the subject's decision time is represented as a random walk, and additional time is required to read the names and actually respond. The subject sets a starting point based on the relative positions of the stimuli in the set of stimuli to be compared. The starting point has the property that the amount of evidence required will be great only if neither stimulus allows good prediction of the correct response; however, when either stimulus permits a fast time, the time will be relatively fast. This multiplicative representation of bias is the novel aspect of the present model to explain semantic congruity effects and end (serial position) effects. If the average rate of evidence accumulation is proportional to the difference between the stimuli, then random walk theory implies Equations 3 and 4 for mean times to make correct choices.

APPENDIX C

In the parallel-averaging model, Equation 6, the weights w_0 , w_1 , and w_2 are independent of the scale values and independent of the configuration of scale values. Configural-weight theory, on the other hand, allows the weight of a stimulus to depend on the rank order of its scale value among the stimuli to be integrated (Birnbbaum, 1974, 1982; Birnbbaum & Stegner, 1979). Table 4 illustrates how configural-weighting modifies Equation 6 to allow certain interactions. In Table 4, the weight of the lowest value within each configuration (of s_0 , s_i , and t_j) is 3.0 and the other two weights are 1.0. The values of s_i and s_j ; t_j are listed in the table; the value of s_0 is 5.0.

The predicted impression for two low-valued adjectives ($s_1 = 0$, $t_1 = 1$) is $(1.5 + 3.0 + 1.1)/(1 + 3 + 1) = 6/5 = 1.2$. However, the two high valued adjectives ($s_3 = 11$ and $t_3 = 12$) have the value $(3.5 + 1.11 + 1.12)/(3 + 1 + 1) = 7.6$, because s_0 (5) is the lowest value and receives the highest weight. Similarly, the (1,3) and (3,1) predictions are $(1.5 + 3.0 + 1.12)/5 = 3.4$ and $(1.5 + 1.11 + 3.1)/5 = 3.8$, respectively. Therefore, the predicted difference between *loyal and understanding vs loyal and*

TABLE 4
Hypothetical Values of Adjective Combinations to Illustrate Configural-Weighting

Adjective <i>i</i>	Value of s_i	Adjective <i>j</i>		
		<i>Obnoxious</i>	<i>Shy</i>	<i>Understanding</i>
		Value of t_j	6	12
<i>Malicious</i>	0	1.2	2.2	3.4
<i>Changeable</i>	5	2.6	5.2	6.4
<i>Loyal</i>	11	3.8	6.4	7.6

Note. In this example, $s_0 = 5$. Entries are calculated using Equation 6, except weight of lowest scale value (s_i , t_j , or s_0) has a weight of 3.0 and the other two terms have weights of 1.0 and 1.0, respectively.

obnoxious is $7.6 - 3.8 = 3.8$, which is greater than the predicted difference between *malicious and understanding* vs *malicious and obnoxious*, which is $3.4 - 1.2 = 2.2$. In sum, configural weighting accounts for the divergent interaction pattern observed by Birnbaum (1974) by assigning greater weight to the lowest valued stimulus in the set of stimuli to be integrated on each trial.

REFERENCES

- 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. 2). San Francisco: Freeman.
- Banks, W. P. (1977). Encoding and processing of symbolic information in comparative judgments. In G. H. Bower (Ed.) *The psychology of learning and motivation* (Vol. 11), New York: Academic Press.
- Banks, W. P., White, H., Sturgill, W., & Mermelstein, R. (1983). Semantic congruity and expectancy in symbolic judgments. *Journal of Experimental Psychology: Human Perception and Performance*, 9, 560-582.
- Birnbaum, M. H. (1974). The nonadditivity of personality impressions. *Journal of Experimental Psychology Monograph*, 102, 543-561.
- Birnbaum, M. H. (1978). Differences and ratios in psychological measurement. In N. J. Castellan and F. Restle (Eds.) *Cognitive Theory* (Vol. 3, pp. 33-74). Hillsdale, NJ: Erlbaum.
- Birnbaum, M. H. (1980). A comparison of two theories of "ratio" and "difference" judgments. *Journal of Experimental Psychology: General*, 109, 304-319.
- Birnbaum, M. H. (1982). Controversies in psychological measurement. In B. Wegener (Ed.) *Social attitudes and psychophysical measurement* (Pp. 401-485). Hillsdale, NJ: Erlbaum.
- Birnbaum, M. H., & Mellers, B. A. (1983). Bayesian inference: Combining base rates with opinions of sources who vary in credibility. *Journal of Personality and Social Psychology*, 45, 792-804.
- Birnbaum, M. H., & Stegner, S. E. (1979). Source credibility in social judgment: Bias, expertise, and the judge's point of view. *Journal of Personality and Social Psychology*, 37, 48-74.
- Birnbaum, M. H., & Stegner, S. E. (1981). Measuring the importance of cues in judgment for individuals: Subjective theories of IQ as a function of heredity and environment. *Journal of Experimental Social Psychology*, 17, 159-182.
- Birnbaum, M. H., & Veit, C. T. (1974). Scale-free tests of an additive model for the size-weight illusion. *Perception & Psychophysics*, 16, 276-282.
- Buckley, P. B., & Gillman, C. B. (1974). Comparisons of digits and dot patterns. *Journal of Experimental Psychology*, 103, 1131-1136.
- Dashiell, J. F. (1937). Affective value-distances as a determinant of esthetic judgment-times. *American Journal of Psychology*, 50, 57-67.
- Hagerty, M., & Birnbaum, M. H. (1978). Nonmetric tests of ratio vs. subtractive theories of stimulus comparison. *Perception & Psychophysics*, 24, 121-129.
- Holyoak, K. J. (1978). Comparative judgments with numerical reference points. *Cognitive Psychology*, 10, 203-243.
- Holyoak, K. J., & Mah, W. A. (1982). Cognitive reference points in judgments of symbolic magnitude. *Cognitive Psychology*, 14, 328-352.

- Link, S. W. (1975). The relative judgment theory of choice reaction time. *Journal of Mathematical Psychology*, 12, 114-135.
- Link, S. W. (1978). The relative judgment theory analysis of response time deadline experiments. In N. J. Castellan & F. Restle (Eds.) *Cognitive Theory* (Vol. 3). Hillsdale, NJ: Erlbaum.
- Link, S. W. (in press). The psychological design of psychological experiments. In H.-G. Geissler (Ed.) *Proceedings of the Fechner Symposium*. Göttingen, Federal Republic of Germany: Hogrefe Verlag.
- Luce, R. D. (1986). *Response times*. New York: Oxford Univ. Press.
- Marks, D. F. (1972). Relative judgment. A phenomenon and a theory. *Perception & Psychophysics*, 11, 156-160.
- Moyer, R. S. (1973). Comparing objects in memory: Evidence suggesting an internal psychophysics. *Perception & Psychophysics*, 13, 180-184.
- Moyer, R. S., & Dumais, S. T. (1978). Mental comparison. In G. H. Bower (Ed.), *The psychology of learning and motivation* (Vol. 12, pp. 117-155).
- Narens, L., & Luce R. D. (1986). Measurement: The theory of numerical assignments. *Psychological Bulletin*, 99, 166-180.
- Petrusic, W. M., & Baranski, J. V. (1989). Semantic congruity effects in perceptual comparisons. *Perception & Psychophysics*, 45, 439-452.
- Potts, G. R., Banks, W. P., Kosslyn, S. M., Moyer, R. S., Riley, C. A., & Smith, K. H. (1978). Encoding and retrieval in comparative judgments. In N. J. Castellan & F. Restle (Eds.), *Cognitive theory* (Vol. 3). Hillsdale, NJ: Erlbaum.
- Ratcliff, R. (1978). A theory of memory retrieval. *Psychological Review*, 85, 59-108.
- Riskey, D. R., & Birnbaum, M. H. (1974). Compensatory effects in moral judgment: Two rights don't make up for a wrong. *Journal of Experimental Psychology*, 103, 171-173.
- Shipley, W. C., Cottin, J. I., & Hadsell, K. C. (1945). Affective distance and other factors determining reaction time in judgments of color preference. *Journal of Experimental Psychology*, 35, 206-215.
- Shipley, W. C., Norris, E. D., & Roberts, M. L. (1946). The effect of changed polarity of set on decision time of affective judgments. *Journal of Experimental Psychology*, 36, 237-243.
- (Accepted July 19, 1989)