Relative judgment: A phenomenon and a theory*

DAVID F. MARKS
University of Otago, Dunedin, New Zealand

An experiment is described in which it was demonstrated that relative judgments of the more probable of two statements are quicker if the statements are both probable rather than improbable. For judgments of the less probable, the reverse result was obtained. This phenomenon is discussed in relation to various theories of judgment and choice. A theory is presented which assumes that judgment involves a relation between a stimulus and a word and follows Thurstone's notion that stimuli differ in their discriminial dispersions. This new theory is shown to be consistent with recent results in psycholinguistics.

De Soto, London, and Handel (1965) describe the following pertinent anecdote: "Mickey Mantle tells of an exhibition baseball game in which he and Willie Mays, both superstars, were playing. Unhappily, the two superstars were having a bad day. Finally a leather-lunged fan shouted at Mantle, 'I came to see which of you guys was better, you or Mays. Instead, I'm seeing which is worse!' Mantle knew he had been insulted. But wherein was the insult?" The two statements, "A is better than B" and "B is worse than A," are, of course, logically equivalent. But as De Soto et al point out, "better" connotes general goodness, whereas "worse" connotes general badness: "The words convey more than strictly relational or comparative meaning; they say something about absolute positions [p. 513]."

Relative judgment—its time, confidence, accuracy, difficulty, etc.—is an old problem in psychology (e.g., see Guilford, 1954; Johnson, 1955). However, little has been found out about the question raised by the anecdote of the angry young baseball fan: how is the relative judgment of a pair of stimuli related to the absolute judgments of those stimuli? It is the purpose of the present paper to describe empirical findings concerning this relationship and to present a theory to account for it.

Shipley, Coffin, and Hadsell (1945) and Shipley, Norris, and Roberts (1946) found that Ss discriminate more rapidly between preferred colors when asked to choose the more pleasant color and between unpreferred colors when asked to choose the less pleasant color. They

*This article is based on part of a dissertation submitted to the Department of Psychology, University of Sheffield, England, in partial fulfillment of the requirements for the PhD degree. The author would like to express his appreciation to Mr. J. K. Clarkson for his encouragement and assistance.

called this phenomenon a "polarity of set factor" and suggested that it might exemplify a general principle governing the speed of many kinds of comparative judgment. Audley and Wallis (1964) and Wallis and Audley (1964) conducted experiments on relative judgments of brightness and pitch to look for this phenomenon and obtained positive evidence. A set of results is shown in Fig. 1a. For stimuli of relatively high average luminance, the brighter stimulus is chosen more quickly than the darker. For stimuli of relatively low average luminance, however, the darker stimulus is chosen more quickly.

For obvious reasons, Audley and Wallis (1964) refer to this phenomenon as the "cross-over effect," but not all the data collected by Audley and Wallis presented exactly this picture. The more common result was a "funnel effect" in which the average response time for one instruction, "choose darker," was greater than that for the opposite instruction, "choose brighter." A typical funnel effect is represented in Fig. 1b. The essential characteristic of both cross-over and funnel effects is that judgment time depends on an interaction between the position of the stimuli along the dimension being judged and the instruction set which is employed. In the case of the funnel effect, this interaction is overlaid by a difference in average speed of response for the two instruction conditions.

In the terminology of Johnson (1955), the experiments referred to have studied either "perceptual" stimulus characteristics (brightness and pitch) or an "affective" characteristic (color preference). The purpose of the following experiment was to investigate the cross-over phenomenon for an "abstract" stimulus characteristic, that of subjective probability. If the cross-over, or funnel, effect could be obtained for such judgments, a fundamental principle could be assumed.

METHOD Stimuli

The stimulus material consisted of two parts: incomplete statements (e.g., "It will snow in Sheffield next . . .") and various pairs of alternative statement endings (e.g., "January, February"). Each part of a stimulus was written on a card. S was presented with an incomplete statement and then with a pair of endings. He could be instructed to choose either the more likely alternative or the less likely alternative. Some attempt was made to keep discriminability constant by pairing endings which are equally "distant" from one another, e.g., for the above incomplete statement, by always pairing months which are adjacent in the calendar. Table 1 gives the complete set of stimuli which were used.

These stimuli were chosen after a pilot experiment in which several kinds of statement were tried.¹ As shown in Table 1, five incomplete statements were used, one for practice trials (P) and four for experimental trials. For each incomplete statement, eight pairs of endings were selected to permit a grouping of two pairs at each of four probability levels. The endings are categorized into the four levels on the basis of median rank along the probability scale, as ranked by the Ss after the main experiment. It can be seen from Table 1 that the pairs of endings for Statement S at Levels 1, 2, 3, and 4 were the same as the pairs of endings for Statement H at Levels 4, 3, 2, and 1, respectively. The endings for D and E complemented each other in the same way. This was an important feature of the experimental design, since the cross-over effect could be investigated not only for each statement, but also for each pair of endings taken separately, as each pair occurred at two different levels of probability.

Fig. 1. Typical results on the effect of opposite instruction sets on the time of relative judgments: (a) the cross-over effect; (b) the funnel effect. These results were obtained by Audley & Wallis (1964, Experiments 2 and 4).

Table 1
Stimuli Used in the Probability Experiment

	Incomplete Statements	Pairs of Endings							
		Level 1 (Low)		Level 2		Level 3		Level 4 (High)	
P	The distance (in miles) from Sheffield to Penzance is**	200 230	220 210	260 230	240 250	340 310	320 330	340 350	360 330
S	It will snow in Sheffield next	June July	July Aug.	May Aug.	June Sept.	March Dec.	Feb.* Nov.	Feb. Jan.	Jan. Dec.
H	The hottest month in the next 12 months in Sheffield will be	Feb. Jan.	Jan. Dec.	March Dec.	Feb. Nov.	May Aug.	June Sept.	June July	July Aug.
D	The Duke of Edinburgh is aged**	67 66	65 68	59 62	61 60	55 54	53 56	47 50	49 48
E	Elizabeth the Queen Mother is aged**	47 50	49 48	55 54	53 56	59 62	61 60	67 66	65 68

*The names of the months were displayed in unabbreviated form to the Ss.

**None of the Ss knew the actual values for Statements P, D, and E, which were 347, 48, and 68, respectively.

Apparatus

A Dodge-type tachistoscope was used for presenting stimuli. S looked through an aperture 30 cm from the display where an incomplete statement was on view. A keypress by E initiated a trial which comprised a delay of 1 sec, followed by a 1-sec warning tone, and then, in place of the associated incomplete statement, the exposure of a card showing a pair of endings. A clock was started at the onset of the exposure of the endings and stopped when S pressed either of two response keys, which were placed conveniently for the left and right hands.

Design and Procedure

As shown in Table 1, two pairs of endings occurred at each probability level for each incomplete statement. One pair of endings was presented under one instruction set—"choose the more likely ending" (M set)—and the other pair under the opposite instruction set—"choose the less likely ending" (L set). For half the presentations, S was given the M set and for half, the L set. Eight Ss (the LM group) gave 20 judgments under the L set, followed by 20 under the M set, and eight (the ML group) gave judgments under sets in the reverse order. Trials were arranged in blocks of four, and each block contained presentations in random order of four pairs of endings at different probability levels for a given incomplete statement. Before each block of trials, a card showing an incomplete statement and appropriate instruction set was placed in the display and S was allowed as much time as he wished to contemplate this

(usually 5-10 sec). Four subgroups within each of the two major groups (LM and ML) were formed, and, across these subgroups, the order of occurrence of the four incomplete statements associated with different blocks of trials gave a standard form of the 4 by 4 Latin square.

Subjects

The Ss were 16 male university students, aged 18-28 years.

RESULTS

The average response times at the four probability levels for the two

instruction sets are shown in Fig. 2. The results were clear-cut. In analysis of variance, the Probability Level by Instruction Set interaction was significant— $F(3,42) = 6.47$, $p < .01$. Judgments of the more probable of two statements were quicker [$t(15) = 3.54$, $p < .01$] if the statements were both probable (Level 4) rather than improbable (Level 1). Judgments of the less probable of two statements, on the other hand, tended to be faster for improbable statements than for probable ones, although this result was not significant [$t(15) = 1.97$, $p < .10$].

As shown in Fig. 2, the average curves for the M and L sets exhibit a funnel effect. This result is similar, therefore, to that obtained by Audley and Wallis (1964) in their Experiments 1, 2, and 3 and by Wallis and Audley (1964) in their Experiment A1. The mean difference between the M and L instruction sets was 0.17 sec and was significant— $F(1,14) = 10.82$, $p < .01$.

Data are also plotted separately in Fig. 2 for the two orders of instruction, LM and ML. It can be seen that judgments under the L set were subject to a large practice effect, although the mean difference of 0.22 sec for the two groups under this set was not significant— $t(14) = 1.81$, $p < .10$. Judgments under the M set, on the other hand, showed no practice effect, the mean time for Group ML actually being marginally less than for Group LM. Times for judgments of "more probable" were, therefore, relatively close to asymptote at the start of the experiment.

A second analysis was conducted in terms of individual pairs of statement endings. This utilized the fact that in

Fig. 2. Mean judgment times for the instructions "choose more likely" (M set) and "choose less likely" (L set) for four levels of probability.

Fig. 3. Mean judgment times for various pairs of statement endings.

this experiment, a physically constant stimulus mediated different subjective probabilities (see Table 1). Although it would have been possible to analyze response times for each individual pair of endings, there was large variability in these data, and it was decided to pool response times across the two pairs of endings at each level, giving in each case a mean of 16 values, one from each S. Four graphs were obtained for Statements D and E and four for Statements S and H. These are presented in Fig. 3.

To illustrate the logic of this analysis, consider Fig. 3e. This graph gives the mean judgment times for the two pairs of stimuli, "June, July" and "July, August," as they occurred for Statements S and H at Levels 1 and 4, respectively. As can be seen in Fig. 3e,

these endings produced a funnel effect which appeared as a significant interaction in analysis of variance— $F(1,14) = 5.41$, $p < .05$. The only other significant interaction was that of Fig. 3h, which indicates a full cross-over effect— $F(1,14) = 10.35$, $p < .01$. Of the remaining six cases, Figs. 3a, b, c, and f yielded funnel effects, Fig. 3g gave parallel lines, and Fig. 3d gave a funnel effect of the opposite kind to that predicted. Hence, only one of these eight fine-grain analyses gave a result which was inconsistent with the general trend portrayed by Fig. 2.

DISCUSSION

The time of relative judgments of probability, like those of color preference (Shipley et al, 1945, 1946), brightness (Audley & Wallis, 1964),

and pitch (Wallis & Audley, 1964), depends on a relationship between the position of the stimuli along the attribute being judged and the response set employed to judge them. Audley and Wallis (1964) suggested an explanation of this pervasive phenomenon in terms of a theory of choice originally presented by Audley (1960). Relative judgment is viewed in terms of a stream of implicit responses which are competing to become overt, as in a kind of race. A different class of implicit response is associated with each possible overt response. An overt response occurs after the occurrence of an uninterrupted run of implicit responses of a given kind. In order to account for the typical psychophysical ogive obtained for single-stimulus judgments of, say, brightness, Audley and Wallis point out that the implicit response rate for the response "bright" must be assumed to increase monotonically with increasing stimulus intensity. The implicit response rate for the response "dark" is assumed to decrease with increasing stimulus intensity. These assumptions are described in Fig. 4.

Either by assuming that the S uses the implicit response system associated with the instruction or by assuming that the S uses the dominant response and then translates his response to match the instruction if required, the system given in Fig. 4 will produce the cross-over effect. The main problem for this ingenious, if elaborate, model is the funnel effect observed by Audley and Wallis in four published experiments of their own and in the probability experiment described here.

A second theory which has been put forward to account for the cross-over effect is Helson's (1964) adaptation-level theory. Arthur's (1965) attempt to explain this phenomenon in terms of assimilation and contrast effects or in terms of the ad hoc assumption of two adaptation levels, one for each instruction set, has been criticized by Audley and Wallis (1965) for its ability to predict both the cross-over effect and its exact opposite, a result which, thus far, has been unreported. The essential feature of the cross-over and funnel phenomena is the difference in judgment time obtained with opposite instruction sets for a given pair of focal stimuli in a given context. Adaptation-level theory (Helson, 1964) states: "Intensity of response is a function of distance from, or ratio of stimulation to, prevailing level [p. 63]." Exactly what can be predicted from adaptation-level theory about the time of relative judgments is uncertain. Its inability to make direct and unambiguous predictions concerning the cross-over effect

Fig. 4. Rates of generation of implicit "darker" and "brighter" responses as a function of the relative intensity of the judged stimuli (cf. Audley, 1960; Audley & Wallis, 1964).

represents a serious weakness.

In the apparent absence of any other satisfactory theory, an explanation of the cross-over and funnel effects is suggested in terms of a new theory of judgment. Judgment is viewed as the affirmation of a proposition, e.g., "A is deep red," "B is quite loud," "C is extremely improbable," etc. A judgmental response is therefore assumed to involve a relation between a stimulus and a word which is appropriately modified so that the response accords as far as possible with the conventional usage of words. As Helson (1964) points out, any stimulus dimension can be characterized by two words which define its extremities, e.g., "dark, bright," "loud, soft," "probable, improbable," and exactly what stimulus one is prepared to label, say, "loud" greatly depends on the context. Essential to Helson's position, despite his emphasis on the poles which define scales, is the assumption that judgments are relative to a neutral point located between the poles.

Let us suppose that, instead of being relative to a neutral point between the poles, judgments are actually relative to one of the poles. In this case, eliciting a judgmental response may be conceived of in three stages: (1) choice of pole, P (e.g., "heavy"); (2) assignment of an ideal stimulus value, I_p , which the judge assumes the pole label accurately represents; (3) determination of the judgment, i.e., its value, time, accuracy, etc. This third stage is conceived of in terms of a special case of Thurstone's (1927) "discriminal dispersion" theory. Each stimulus perceived by the S is represented internally by a normal distribution of so-called discriminational processes, and the standard deviation of such a distribution is known as the

"discriminal dispersion." The present model makes the assumption that the discriminational dispersion of a stimulus is proportional to its distance from I_p —the ideal polar stimulus value. Figure 5 presents this assumption pictorially.

Figure 5 shows two pairs of stimuli which are represented by single values along the stimulus continuum (horizontal line at top). There are, of course, many dimensions along which these stimuli could be judged, but suppose we have a dimension defined by the polar labels "P, Q." Let us now consider the two instruction sets corresponding to these two polar labels: "Choose P-er" and "Choose Q-er." The assumption is made that the discriminational dispersion of each stimulus along the psychological scale is dependent upon which pole value it is related to. It is as if in measuring the length of a piece of string the variability of your measurements depended on which end of your ruler you measure from, while the mean value remained constant. To take a specific example, a light source of relatively low intensity would, it is assumed, produce a small discriminational dispersion if judged for relative darkness and a large dispersion if judged for relative brightness.

The only remaining stage of this theory is the Thurstonian assumption that the speed of a relative judgment is directly proportional to the discriminational difference of two stimuli and inversely proportional to some monotonically increasing function of the dispersion of the distribution of discriminational differences, e.g.,

$$\text{speed}_{j,k,p} = C \frac{\mu_j - \mu_k}{\sigma_{j,p}^2 + \sigma_{k,p}^2}$$

where μ_j and μ_k are the mean values of stimuli S_j and S_k , $\sigma_{j,p}$ and $\sigma_{k,p}$ are the discriminational dispersions of S_j and S_k for the response set P, and C is a constant.

Thurstone (1927, p. 387) suggested that practice has the effect of reducing the dispersions of the stimuli being judged. This ties in rather nicely with the present approach, since one is assuming that the closer a stimulus is to the ideal polar value, I_p , or, in other words, the more compatible a stimulus is with the polar label, the narrower its dispersion. The more compatible stimulus-word relationships are precisely those which receive most practice in ordinary language behavior.

The funnel effect does not represent a problem for the present theory, as it would seem to for that of Audley (1960), which predicts the full cross-over effect. For example, the finding of the present study that judgments of "less probable" take

Fig. 5. Discriminal dispersions of two pairs of stimuli in relation to two opposite polar labels (P and Q).

longer than judgments of "more probable" (see Fig. 2) could quite plausibly be related to the relative frequency with which these two judgment sets are employed in daily life. This notion is corroborated by the tendency for times for the M set to show a smaller practice effect than those for the L set (Fig. 2). Although nonsignificant, this result provides some evidence that Ss were relatively more practiced at responding to the more probable of two stimuli prior to the experiment.

Of crucial importance in comparing the Audley model with that of Fig. 5 are error data. Audley and Wallis (1964, p. 72) predict an association between greater accuracy and slower responses, while the model of Fig. 5 predicts the reverse result. The data of Kellogg (1931) and unpublished results of the writer on brightness discriminations show greater accuracy to be associated with faster responses and, hence, favor the discriminational dispersion model.

The theory proposed assumes judgment to be a linguistically mediated process. Clearly, a linguistic theory of judgment should be consistent with existing psycholinguistic data. This proves to be the case. Of particular relevance are two linguistic processes described by Clark (1969a) which he terms "congruence" and "lexical markedness." Congruence concerns the relationship between a question and the information which is searched to provide an answer. Clark (1969a, b) has demonstrated that solutions to three-term series problems take less time when the information to be searched and the question to be answered are congruent in their base strings than when they are incongruent. To interpret relative judgment in terms of congruence, consider the presentation of two stimuli, P_1 and P_2 , both of high probability, but where the probability

of P_1 exceeds that of P_2 . Suppose S is instructed to choose the more probable stimulus. In linguistic terms, the task can be conceptualized in the following stages: The stimuli provide the information ("P₁ is probable++"; "P₂ is probable"). The instruction requires the selection of a P so that ("P is probable++"), that is, it requires a search for the term with the most-plussed "probable." When coded linguistically in this way, the form of the instruction is congruent with the information in the stimuli, and the response should therefore be relatively fast. If, on the other hand, P₁ and P₂ are of low probability and coded in the form ("P₁ is improbable"; "P₂ is improbable++"), the instruction form (choose "P is probable++") and the stimulus information will be incongruent. The extra time required by a futile search for the congruent item and for a reformulation of the instruction from ("P is probable++") to congruent ("P is improbable—") will produce a relatively slow response. Incongruence becomes more likely as the stimuli change from a high to a low level of probability, and the mean judgment time for the instruction set "choose more probable" is therefore predicted to increase accordingly. Since congruence for opposite instruction sets occurs at opposite ends of the stimulus continuum, we have another explanation of the cross-over effect.

But what about the funnel effect? This can be neatly accounted for in Clark's (1969a) theory by the principle of lexical marking. Antonymous adjectives like "good" and "bad" and "probable" and "improbable" may be categorized as "unmarked" or "marked." This dichotomy is based on two criteria. First, the unmarked member of such pairs can be neutralized while the marked member cannot. A speaker who asks, "How probable is event X?" will be satisfied whether he is told that X is probable or improbable. On the other hand, the speaker who asks, "How improbable is X?" has already

judged X to be improbable and is asking about the extent of X's improbability. Second, the unmarked member of a pair is used to name the full scale, e.g., it is the "probability" scale, not the "improbability" scale. This is because the marked term tends to refer to only one-half of the full scale. The unmarked adjective, therefore, has two senses, which Clark (1969a, p. 389) terms "nominal" and "contrastive," and the marked adjective only one (contrastive). According to Clark's principle of lexical marking, a nominal sense takes up less storage space and is therefore stored and retrieved more quickly than its corresponding contrastive sense. There is considerable evidence to support this principle (e.g., Clark, 1969a, b; Clark & Card, 1969), and, in applying it to relative judgment, faster responses would be predicted for unmarked instruction sets like "choose more probable" than for marked instruction sets like "choose less probable" (equivalent to "choose more improbable"). It should be noted that, on average, an unmarked word (with two senses) occurs more often than a marked word (with only one sense). The principle of lexical marking is therefore consistent with the explanation of the funnel effect in terms of practice effects proposed above.

In sum, Fig. 5 can be regarded as one possible formal representation, in terms of discriminial dispersions, of Clark's (1969a) principles of congruence and lexical marking. While the discriminial dispersion theory and Clark's (1969a) linguistic theory are not inconsistent, neither are they identical. For example, Clark's theory implies that relative judgment is a serial multistage process, while the discriminial dispersion theory gives a single-stage model. Further data are needed to determine which theory is the better.

REFERENCES

- ARTHUR, A. Z. 'Cross-over' effect and subjective intensity. *British Journal of Psychology*, 1965, 56, 477-478.

- AUDLEY, R. J. A stochastic model for individual choice behavior. *Psychological Review*, 1960, 67, 1-15.
- AUDLEY, R. J., & WALLIS, C. P. Response instructions and the speed of relative judgments. I. Some experiments on brightness discrimination. *British Journal of Psychology*, 1964, 55, 59-73.
- AUDLEY, R. J., & WALLIS, C. P. 'Cross-over' effect: A reply to Arthur. *British Journal of Psychology*, 1965, 56, 479-482.
- CLARK, H. H. Linguistic processes in deductive reasoning. *Psychological Review*, 1969a, 76, 387-404.
- CLARK, H. H. Influence of language on solving three-term series problems. *Journal of Experimental Psychology*, 1969b, 82, 205-215.
- CLARK, H. H., & CARD, S. K. Role of semantics in remembering comparative sentences. *Journal of Experimental Psychology*, 1969, 82, 545-553.
- DE SOTO, C. B., LONDON, M., & HANDEL, S. Social reasoning and spatial paralogic. *Journal of Personality & Social Psychology*, 1965, 2, 513-521.
- GUILFORD, J. P. *Psychometric methods*. New York: McGraw-Hill, 1954.
- HELSON, H. *Adaptation-level theory*. New York: Harper & Row, 1964.
- JOHNSON, D. M. *The psychology of thought and judgment*. New York: Harper, 1955.
- KELLOGG, W. N. The time of judgment in psychometric measures. *American Journal of Psychology*, 1931, 43, 65-86.
- SHIPLEY, W. C., COFFIN, J. I., & HADSELL, K. C. Reaction time in judgment of color preference. *Journal of Experimental Psychology*, 1945, 35, 206-215.
- SHIPLEY, W. C., NORRIS, E. D., & ROBERTS, M. L. The effect of changed polarity of set on decision time of affective judgments. *Journal of Experimental Psychology*, 1946, 36, 237-243.
- THURSTONE, L. L. Psychophysical analysis. *American Journal of Psychology*, 1927, 38, 368-389.
- WALLIS, C. R., & AUDLEY, R. J. Response instructions and the speed of relative judgments. II. Pitch discrimination. *British Journal of Psychology*, 1964, 55, 121-132.

NOTE

1. The prominence of Sheffield weather and the British Royal Family in these stimuli met the requirement that stimuli should denote referents with which the Ss would be familiar but without complete knowledge. Neither the weather nor the age of Royal Family members was predictable or known with complete certainty by any of the Ss.

(Accepted for publication July 27, 1971.)