Cognitive Reference Points in Judgments of Symbolic Magnitude

KEITH J. HOLYOAK AND WESLEY A. MAH

University of Michigan

Research on speeded symbolic magnitude comparisons indicates that decisions are made more quickly when the magnitudes of the stimuli being compared are relatively close to an explicit or implicit reference point. Alternative explanations of this phenomenon are tested by seeking similar effects in nonspeeded rating tasks. In accord with the predictions of discriminability models, rated magnitude differences between stimuli in the vicinity of a reference point are expanded relative to differences between stimuli far from it. The inferred locations of cities along a west–east axis varies systematically depending on which coast, Pacific or Atlantic, is specified as the reference point. Scales derived from the rating data are correlated with the pattern of reaction times obtained in a comparable speeded comparison task. In addition, the distance between the cities nearest the locale of our subjects is subjectively stretched. Reference point effects are also observed when the form of the comparative specifies an implicit reference point at either end of a continuum of subjective size; however, these effects are very small and do not clearly support a discriminability interpretation. Stronger evidence for discriminability effects is obtained when an explicit reference point is established at an arbitrary size value. An implicit scaling model, related to range-frequency theory, is proposed to account for the influence of reference points on relative discriminability of stimulus magnitudes. The implicit scaling model is used to develop an account of how symbolic magnitudes may be learned and of how habitual reference points can produce asymmetries in distance judgments.

Memory, it seems, is not a camera that faithfully records the products of our sensory experience; rather, it is a distorting mirror that reflects the shadows, subtly altered in shape, of what we once perceived. The transformations produced by memory are varied, but often systematic. Thus, we find that recall of stories is distorted in the direction of cultural stereotypes (Bartlett, 1932), that verbal labels influence the recall of pictures (Carmichael, Hogan, & Walter, 1932), and that recognition of visual patterns depends on their similarity to an unseen prototype (Franks &

Bransford, 1971). Memory transformations have also been observed in standard psychophysical judgment tasks such as magnitude estimation (Kerst & Howard, 1978; Moyer, Bradley, Sorensen, Whiting, & Mansfield, 1978). Furthermore, memory performance is determined not only by the information originally encoded, but by the manner in which that information is elicited at the time of its retrieval (Loftus & Palmer, 1974). Information in memory may be transformed by the very processes that act upon it in the course of answering a question or reaching a decision. The present paper focuses on transformations induced by the process of making judgments about remembered geographical distances and object sizes.

SYMBOLIC COMPARATIVE JUDGMENTS

Symbolic Distance Effect

People have fairly detailed information stored in memory about the values of concepts on such continuous dimensions as size, distance, and duration, as evidenced by their ability to make psychophysical judgments about symbolic stimuli. Most research on the processing of subjective magnitudes has used a speeded comparative-judgment paradigm. Subjects are typically presented with a pair of symbolic stimuli, such as two object names, and asked to choose the item that is greater (or lesser) in magnitude along some dimension, such as size. A consistent finding, termed the *symbolic distance effect,* is that reaction time (RT) decreases as the separation between two stimuli on the relevant dimension increases (Moyer & Landauer, 1967). Holyoak and Walker (1976) provided evidence that the distance effect depends on interval properties of subjective magnitude differences, rather than just ordinal properties. Various theorists have suggested that symbolic magnitude is coded in memory by values with interval-scale properties and that these analog magnitude values are compared to make relative judgments (Moyer & Landauer, 1967; Holyoak & Walker, 1976). The comparison process has been characterized as a sequential sampling procedure, such as a random walk process, that reaches a criterion more quickly for larger magnitude differences (Buckley & Gillman, 1974; Holyoak, 1978; Holyoak & Patterson, 1981; Moyer & Dumais, 1978).

Semantic Congruity and Reference Point Effects

A second phenomenon, termed the *semantic congruity effect,* involves the form of the comparative used in the instructions. Subjects are relatively fast to select the greater of two high-magnitude stimuli, while they are faster to select the lesser of two low-magnitude stimuli. For example, subjects can choose the larger member of a pair such as *airplane–truck* more quickly than they can choose the smaller, while the reverse is true

for a pair like *moth–fly*. While numerous specific models of the congruity effect have been advanced, these can be grouped into two broad classes, which we will term *discriminability* and *nondiscriminability* models. Discriminability models, as their name implies, assume that the comparative alters the relative discriminability of symbolic magnitudes. One general view is that the comparative specifies an implicit "reference point" at one end of the magnitude continuum and that differences between items with magnitudes close to the reference point will be more discriminable than differences between items with magnitudes near the opposite pole (Holyoak, 1978; Jamieson & Petrusic, 1975).

The reference point explanation of the congruity effect suggests that comparable variations in relative discriminability might be produced by *explicit* reference points. To test this hypothesis, Holyoak (1978) timed subjects as they decided which of two digits was closer in magnitude to a third (the reference digit). In addition to a distance effect, an effect of nearness to the reference digit was observed. Holding the pair constant, RT increased with distance to the reference digit (e.g., subjects could select the closer member of the pair 3−4 more quickly when the reference digit was 6 rather than 7). Like the congruity effect, this result can be interpreted as evidence that discriminability is increased in the region of a reference point. In general, RT in Holyoak's experiments was a linear function of the ratio of the closer to the further of the two distances being compared. This ratio measure is sensitive to both the distance between the two items and their overall distance from the reference stimulus. Other studies have also yielded evidence suggesting increased discriminability of magnitudes in the vicinity of a reference point for such diverse stimuli as geographical locations (Baum & Jonides, 1979; Evans & Pezdek, 1980), color concepts (te Linde & Paivio, 1979), and semantic categories (Hutchinson & Lockhead, 1977). The assumption that values close to a reference point are relatively discriminable has also been embodied in Kahneman and Tversky's (1979) "prospect theory" of economic choice behavior.

Reference Points in Nonspeeded Judgments

The results of the above RT experiments, while consistent with the hypothesis that discriminability is increased in the region of a reference point, do not rule out alternative explanations. If the discriminability of the difference between a pair of stimuli is indeed affected by the form of the instructions, then it should be possible to find congruity and reference point effects with unspeeded rating tasks. For example, suppose a person is presented with a pair of cities located in the western United States and is asked to rate how much closer one is than the other to the Pacific or the

Atlantic Ocean. Discriminability models predict that in this example the rated distance will be greater when the Pacific, rather than the Atlantic, is specified as the reference point (since the cities are located relatively close to the Pacific). More generally, discriminability models predict that comparative judgments about stimulus pairs in relationship to a reference point can be described by

$$R_{ijk} = J_k\big[f(s_i - s_k) - f(s_j - s_k)\big],   \tag{1}$$

where R_{ijk} is a response related to the subjective difference between stimuli i and j with respect to reference point k; s_i, s_j, and s_k are the subjective scale values of the two stimuli and the reference point, respectively; J_k is a monotonic judgment function associated with reference point k; and f is an increasing, negatively accelerated function. The effect of f is to "expand" differences between stimuli close to the reference point, relative to comparable differences between stimuli far from it. Discriminiability models permit but do not require J_k to differ across reference points. When R_{ijk} is a response on a category scale, we assume that J_k will be approximately linear. If R_{ijk} is a mean RT obtained in a speeded comparison task, J_k will presumably have a different form. However, the function f should remain constant across different types of judgment tasks.

Nondiscriminability models assume that while the form of the instructions may influence speed of responding, it does not alter subjects' subjective impressions of the magnitude differences among the items. Such models predict that nonspeeded judgments will conform to

$$R_{ijk} = J_k(s_i - s_j).   \tag{2}$$

Equation 2 is a special case of Equation 1 that implies the reference point can alter the judgment function, but cannot transform the scale values nonlinearly. (Equation 1 reduces to Equation 2 if f is linear.)

A few previous studies have obtained psychophysical judgments for the same set of stimuli with two different comparatives (Torgerson, 1960; Birnbaum & Mellers, 1978; Banks & Root, 1979). None of these studies reported any evidence of the type of reference point effect predicted by discriminability models.[1]

---

[1] Earlier work in social psychology suggests possible effects of reference points on nonspeeded judgments. Hovland and Sherif (1952) found that judges with extreme opinions made finer distinctions among attitudes close to their own than among neutral or opposing attitudes, as measured by a sorting task. However, this result is apparently fragile (Segall, 1959) and dependent on the type of judgment required (Kelley, Hovland, Schwartz, & Abelson, 1955). Furthermore, any such effect might well reflect real differences across subject populations in the subjective similarities among attitudes.

EXPERIMENT 1

Experiment 1 was modeled after that of Birnbaum and Mellers (1978), who obtained judgments of relative "easterliness" or "westerliness" for pairs of U.S. cities. Seven cities that roughly lie on an imaginary east–west line across the country were selected. These overlapped with, but were not identical to, the set used by Birnbaum and Mellers. The cities were San Francisco (SF), Salt Lake City (SL), Denver (Den), Kansas City (KC), Indianapolis (Ind), Pittsburgh (Pitt), and New York (NY).

Method

Experiment 1A. A design similar to that of Birnbaum and Mellers was used. One group of subjects rated all possible city pairs under two different sets of instructions. In each case subjects were asked to "write a number from 1 to 9 beside each pair indicating how much closer the first [or second] city is to the reference point than the second [or first] city is." In the Pacific version, subjects were told to adopt as a reference point "an imaginary spot sitting a few miles off the PACIFIC OCEAN." The Atlantic version was identical except that the Atlantic Ocean was specified as the reference point. In both versions a rating of 9 indicated that the nearer city was maximally closer than the farther city to the reference point. Judgments were blocked by form of the comparative, with the pairs listed in a different random order for each condition. The order of the two instructions and assignment of random order to instruction condition were counterbalanced across subjects. The two lists were presented to subjects in a single booklet. Subjects were told not to look ahead in the booklet, and prior to the second instruction condition they were advised not to "look back or even think back" to their previous answers. Each list included 42 items (21 pairs of different cities, each in two orders).

A separate group of subjects rated the pairs under Neutral instructions, which simply asked them to rate how far apart each pair of cities seemed to be. In all three versions higher ratings indicated longer intercity distances.

Eighty undergraduates completed the Pacific and Atlantic versions, and an additional 24 completed the Neutral version.

Experiment 1B. To compare the reference point effects obtained in rating and RT paradigms, 14 additional subjects performed a speeded comparative-judgment task with the same city pairs. To make it less obvious that the critical cities formed a linear ordering, we added an additional seven filler pairs to the item set. These fillers were composed of 14 different U.S. cities, including cities on the two coasts. None of these fillers was a critical city, and most of them fell off the east–west line defined by the critical cities.

On each trial subjects judged which of two cities was closer to (or farther from) the Pacific or Atlantic Ocean and pressed one of the two response keys as quickly as possible. Each of the 28 basic pairs (21 critical pairs and seven fillers) was presented with both reference points, both comparatives (*closer* vs *farther*), and in both possible orders, for a total of 224 test trials, which were preceded by 20 practice trials with different city pairs. Each subject received the pairs in a different random order. The stimuli were presented on a CRT screen controlled by a PDP 11/60 computer. On each trial the sequence of events was as follows. First, the reference point (Pacific or Atlantic) was presented for 1 sec. Then the comparative (*closer* or *farther*) was presented below the reference point. A fixation cross simultaneously appeared below the comparative. After a further 500 msec, the fixation cross was replaced by the city pair, centered horizontally on the position previously occupied by the fixation cross. The entire display (reference point, comparative, and cities) then remained in view until the subject pressed a response key or a time limit of 5 sec was exceeded. If the subject

made an error, the word "error" was displayed for 1 sec. The next trial then began automatically.

Results and Discussion

Experiment 1A: Mean category ratings. An initial set of analyses was performed on the data from those subjects who completed the Pacific and Atlantic versions of the rating task. The mean ratings obtained for the 21 pairs in the Pacific and Atlantic conditions are presented in Fig. 1. The horizontal axis represents the ordinal position of the more westerly pair member, and the lines connect pairs at each ordinal distance from 1 (adjacent pairs) to 5. Discriminability models predict that cities in the vicinity of the reference pole will be rated as relatively far apart. The change in instructions should therefore produce a "crossover" interaction at each ordinal distance.
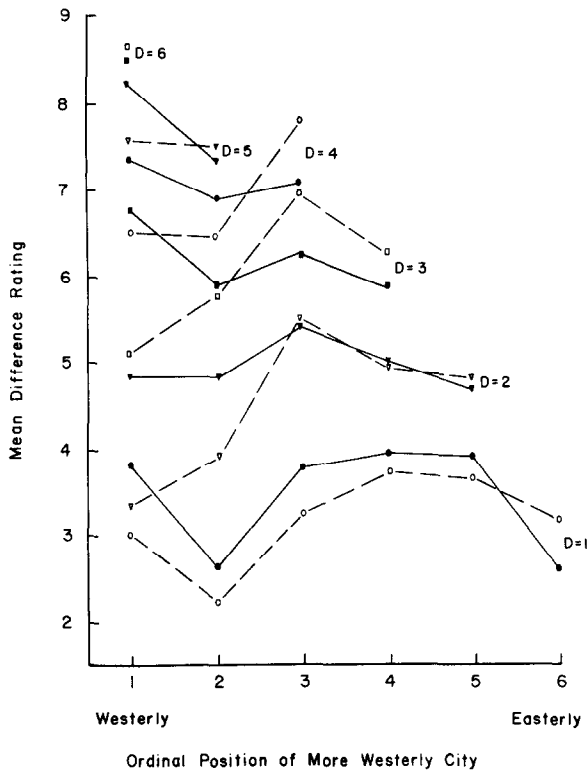

FIG. 1. Mean ratings obtained under Pacific (———) and Atlantic (– – –) instructions as a function of the ordinal position of the more westerly city and of the ordinal distance between the cities (Experiment 1A).

As illustrated in Fig. 1, such interactions were indeed obtained. Analyses of variance were performed on the ratings for the Pacific and Atlantic conditions. Separate analyses were performed for each ordinal distance, and the predicted interaction between the pair locations and the reference pole was assessed by a bilinear trend test. These bilinear trends were highly significant for Distances 1 to 4, $p < .001$. Ratings increased monotonically with ordinal distance, $p < .001$. Neither the order of the cities within pairs nor the order of the two instruction conditions had any consistent effects on the pattern of results.

Equation 2 can be rejected if the rank order of the rated distances differs systematically as a function of the specified reference point, since changes in J_k cannot alter the rank order. We therefore examined the rank order of the ratings given by each subject to the two extreme pairs at each ordinal distance (e.g., San Francisco−Salt Lake and Pittsburgh−New York for Distance 1), under both Pacific and Atlantic instructions. Discriminability models (Equation 1) predict that more subjects will rate the westerly pair higher under Pacific instructions and the easterly pair higher under Atlantic instructions than vice versa. Nondiscriminability models (Equation 2) predict that the two patterns will be equally probable. (Cases in which the subject gives a higher rating to the same pair under both instructions do not differentiate between the two equations.) Binomial tests of these frequency counts supported the prediction of discriminability models at each ordinal distance, $p < .05$. The count at Distance 1, for example, was 17 subjects vs four. Equation 2 can therefore be rejected on the basis of ordinal violations that were statistically reliable across subjects.

Discriminability models derive further support from scales of relative location constructed for each reference point condition (scales A and B in Fig. 2). The mean ratings for the 21 city pairs were fitted by a nonmetric scaling program, CONSCAL (Noma & Johnson, Note 1). The resulting
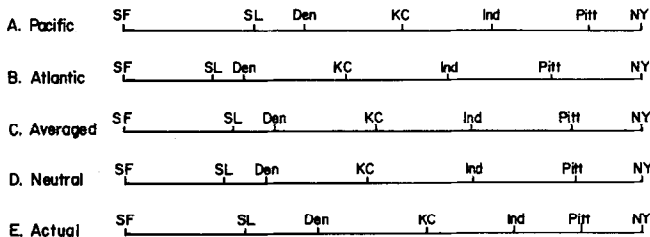

FIG. 2. Scales of relative city locations based on: (A) Pacific instructions; (B) Atlantic instructions; (C) average of scales A and B; (D) Neutral instructions; (E) actual east−west distances. Scales A−D were derived from pair-comparison data (Experiment 1A).

unidimensional solutions had stress values (Kruskal's form 1) less than .02. A separate scale (D in Fig. 2) was derived for the Neutral condition.[2] All the scales in Fig. 2 have been normalized to a constant width. The Pacific and Atlantic scales clearly show the reference point effects discussed above with respect to the mean ratings and frequency counts. The distances among the western cities are relatively expanded on the Pacific scale and compressed on the Atlantic scale; the reverse is true for the eastern cities. Note that the locations of all the interior cities on the Neutral scale, D, are intermediate between their locations on the Pacific and Atlantic scales. It is as if the two reference points had produced symmetrically opposite transformations of the Neutral scale. This impression is heightened when we consider scale C, which was generated by averaging the locations of each city on the Pacific and Atlantic scales. This averaged scale is virtually identical to the Neutral scale.

Scale E was derived by measuring the actual locations of the seven cities, as they project onto an east–west line, using a map of the United States. The distance between Salt Lake City and Denver is underestimated in scales C and D, while that between Indianapolis and Pittsburgh is overestimated. It may be more than coincidental that Ann Arbor, where our subjects were tested, is located between the latter pair of cities. Interestingly, the subjective location scale reported by Birnbaum and Mellers reveals similar overestimations of distances between the cities nearest the site of their study, Champaign-Urbana, Illinois (compare their Fig. 4C and D, 1978, p. 407). These results suggest that residents of America's heartland may not be immune to the type of distance distortion of which New Yorkers are commonly accused.

The exaggerated Indianapolis-to-Pittsburgh distance in scales C and D may in part reflect the fact that all cities from Indianapolis on west—that is, all cities west of Ann Arbor—were displaced to the west on the subjective scales. The Birnbaum and Mellers scale again reveals a comparable effect. Our subjects seem to believe that the Midwest is located entirely in the midwest of the United States, when in fact it extends into the mideast.[3]

Experiment 1B. The data from Experiment 1B allow a direct comparison of reference point effects in rating and RT paradigms. Only the RTs for trials in which the question was "Which city is closer?" will be dis-

---

[2] In the case of the Neutral condition, an initial nonmetric scaling yielded a degenerate solution. Scale D was therefore constructed by a metric scaling procedure. This metric solution yielded a stress value less than .01 when submitted to a confirmatory test using CONSCAL.

[3] The pattern of results obtained in Experiment 1A was also found in an experiment in which the three sets of instructions were administered to independent groups of subjects.

cussed, since the term *closer* was always used in the rating tasks.[4] The overall error rate was 8.1%, and errors were positively correlated with RTs. The mean correct RTs for the 21 critical pairs, as a function of the reference point, the ordinal position of the more westerly city, and the ordinal separation of the two cities are presented in Table 1. The RT pattern generally resembled that obtained in many previous studies of comparative judgments. In particular, RT declined monotonically by 667 msec as ordinal separation increased from one to six steps, and RT was relatively fast for pairs close to the reference point. The latter interaction, corresponding to the congruity effect typically observed in similar RT studies, was significant at Distances 1 and 2, with substantial trends apparent at Distances 3 and 5. Ordinal tests analogous to those performed in Experiment 1A yielded comparable outcomes. At each ordinal distance, the number of subjects with faster RT for the most westerly than the most easterly pair when the Pacific was the reference point, and vice versa when the Atlantic was the reference point, was compared to the number of subjects who exhibited the opposite pattern. As discriminability models predict, more subjects fit the former pattern (six vs zero at Distance 1, and 11 vs zero at Distance 2, $p < .02$ by binomial tests, with nonsignificant trends in the same direction at Distances 3 to 5). Equation 2 can therefore be rejected as a description of the RT data of Experiment 1B, just as in the case of the rating data of Experiment 1A.

Our main concern was to determine the extent to which the different RT patterns obtained for the Atlantic and Pacific reference points can be predicted by the scales derived in the corresponding rating tasks (scales A and B in Fig. 2). We therefore correlated the differences between the two sets of intercity distances with the differences between the two sets of RTs (i.e., Atlantic distance minus Pacific distance vs Atlantic RT minus Pacific RT). If the scale differences predict the rank order of the observed RT differences across reference points, a substantial negative rank-order correlation should be obtained; if not, the correlation should not differ from zero. A significant correlation was in fact obtained, $r_s = -.67, p < .01$. This result indicates that the influence of reference points on choice RT can in some measure be predicted from their influence on the relative discriminability of subjective locations, as determined in an independent rating task.

----

[4] The RT pattern differed substantially when the comparative *farther* was used, and the effect of the reference point was more complex. There were suggestions in the data that subjects used multiple strategies to answer questions with *farther*. For a discussion of alternative possible strategies, see Holyoak (1978, Experiments 4 & 5).

TABLE 1
Mean Correct RT to Choose the City Closer to the Pacific or Atlantic Ocean
(Experiment 1C)

Reference point	Ordinal separation	Ordinal position of more westerly city					
		1	2	3	4	5	6
Pacific	1	1437	2595	2189	2026	2431	2053
Atlantic		1959	2717	2242	2297	1799	1541
Pacific	2	1314	1902	1581	1666	1873	
Atlantic		1750	2078	1583	1674	1195	
Pacific	3	1530	1645	1733	1572		
Atlantic		1559	1701	1540	1317		
Pacific	4	1455	1544	1415			
Atlantic		1474	1503	1425			
Pacific	5	1268	1528				
Atlantic		1500	1448				
Pacific	6	1378					
Atlantic		1503					

EXPERIMENT 2

In Experiment 1 the subjects' task always involved evaluation of the difference between a *pair* of stimuli with respect to a reference point. Would similar results be obtained if the task were to assess the value of a *single* stimulus relative to a reference point? Two of the previous studies that failed to find results comparable to those of the present study involved judgments about a single stimuli (Banks & Root, 1979; Torgerson, 1960). Discriminability models, however, predict that judgments about single stimuli should conform to

$$R_{ik} = J_k[f(s_i - s_k)], \tag{3}$$

where R_{ik} is the response to stimulus i with respect to reference point k, and f is again an increasing, negatively accelerated function. The result will be a "stretching" of relative subjective distances in the region of the reference point, just as we have observed in the case of pair comparisons. Experiment 2 was performed to assess the role of reference points in judgments about the values of single stimuli.

Method

Subjects gave category ratings of the locations of U.S. cities on a 9-point scale. Seven filler cities (including one located on each coast) were added to the seven critical cities previously used in Experiment 1. Different subjects received the list of 14 cities in one of four different random orders. The instructions in Experiment 2 specified the reference point

in one of two different ways. One pair of versions asked subjects to judge how close each city was to either the Pacific or Atlantic coast. This explicit specification of the reference point locations is comparable to that provided in Experiment 1. In the other pair of versions, subjects were asked to judge how far either west or east each city was, leaving the location of the reference point implicit. In all versions a rating of 9 indicated that the city was maximally close to the implied reference point. This means that western cities were to receive the highest ratings under Pacific/West instructions, while eastern cities were to receive the highest ratings under Atlantic/East instructions.

A total of 132 University of Michigan undergraduates completed the rating task, with 33 subjects receiving each of the four versions of the instructions.

Results and Discussion

To examine the overall pattern of results, the ratings each subject gave for the seven critical cities were placed in the order of the actual city locations. The six differences between the ratings given to adjacent cities were then calculated and used as indices of the subjective intercity distances. An analysis of variance was then performed on the rating differences between cities to assess the influence of the reference pole (Pacific/West vs Atlantic/East) and the mode of specifying the reference point (Pacific/Atlantic vs West/East). The predicted influence of the reference point was tested by a bilinear trend test. This trend was highly significant, $t(640) = 5.26, p < .001$, reflecting the relatively wider spacing of cities near the reference pole. This effect was .64 units in magnitude for the two extreme adjacent pairs. The mode of specifying the reference point produced small variations in the overall intercity distances; however, the influence of the reference point on the spacings did not vary significantly between the Pacific/Atlantic and West/East versions, $F(5,640) = 1.01$.

The discriminability prediction was tested by a frequency count, as in the previous experiments. For each subject the rating difference between San Francisco and Salt Lake City was compared to that between Pittsburgh and New York. Under Pacific and West instructions, the distance index was higher for the westerly pair for 71% of the subjects, higher for the easterly pair for 3%, and equal for 26%. Under Atlantic and East instructions the comparable figures were 32, 30, and 38%, respectively. These frequencies differ significantly in the manner predicted by discriminability models, $\chi^2(2) = 26.2, p < .001$.

Scales of relative location were derived by successive interval scaling (Bock & Jones, 1968), combining the Pacific and West data and the Atlantic and East data.[5] The two resulting data sets were analyzed together, after inverting the response scale for the Pacific/West data. Equal dispersions were assumed. Estimated scale values for the cities were normal-

---

[5] The computer program used to perform successive interval scaling was developed by J. E. Keith Smith.

ized to a constant width, separately for each reference point condition. This procedure is equivalent to fitting a constrained form of Equation 3 to the data, which assumes that the judgment functions associated with the two reference points are linearly related to each other. The deviations of the observed from predicted response frequencies did not approach significance, $\chi^2(78) = 77.4, p > .50$. Furthermore, the two resulting location scales were virtually identical to those obtained when the two data sets were scaled separately, removing the constraint on the relationship between the two judgment functions.

Scales F and G in Fig. 3 present the scales of location derived from the Pacific/West and Atlantic/East ratings, respectively. While Scales F and G are not identical to the analogous scales A and B (Fig. 2) derived from pair comparisons, they show the same basic phenomenon—a relative expansion of distances between cities close to the reference point, relative to cities far from it. Reference points thus also appear to affect scales derived by the successive intervals method.

Scale H in Fig. 3 (analogous to scale C in Fig. 2) was obtained by averaging the Pacific/West and Atlantic/East scales. Scale E again presents the actual spacings for comparison. Scale H is somewhat less veridical than scale C (or the Neutral scale D), perhaps because it is based on fewer data per subject; nevertheless, it shows a pattern similar to those apparent in the scales from pair comparisons. The distance from Salt Lake City to Denver is underestimated, and the distance from Indianapolis to Pittsburgh is overestimated. A general tendency to displace cities to the west is also apparent. This westward displacement begins with Pittsburgh, suggesting that the overestimation of the Indianapolis–Pittsburgh distance is at least partially independent of the overall westward shift, since both these cities were displaced to the west. The results of Experiments 1A and 2 thus both provide hints that subjective distances between locations close to the locale of the respondents tend to be expanded.
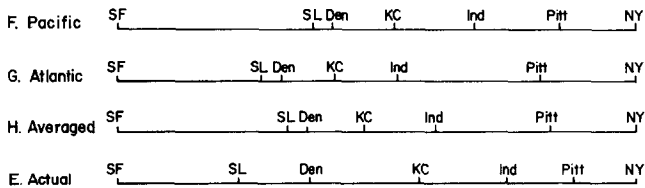

FIG. 3. Scales of relative city locations based on: (F) Pacific/West instructions; (G) Atlantic/East instructions; (H) average of scales F and G; (E) actual east–west distances. Scales F–H were derived by successive interval scaling from ratings of the relative locations of individual cities (Experiment 2).

EXPERIMENT 3

Experiment 3 applied the rating paradigm of Experiment 1A to comparisons of the subjective size of objects. Reference points were manipulated by varying the polar adjective used in the instructions (*smaller* vs *larger*), as in previous RT studies (e.g., Jamieson & Petrusic, 1975).

Method

Subjects rated the size differences between pairs of object names under instructions phrased in terms of either "largeness" or "smallness." Twelve object names were selected from the size norms provided by Paivio (1975). These norms give mean size ratings for each object on a 9-point scale. The 12 test items in order of increasing size were ant, moth, prune, doorknob, hairbrush, flute, lamp, eagle, goat, motorcycle, horse, and garage. The mean ratings of these items ranged from 1.04 to 7.96. All possible pairings of the 12 items were then formed and listed in a random order on a sheet of paper. The smaller object in each pair always appeared on the left. Subjects were then asked to rate the magnitude of the size difference between the items in each pair on a 9-point scale, with higher numbers indicating greater size differences. Two sets of instructions were used. One version included the following: "Each pair of words below names objects that differ in size. In each case the object named on the right is LARGER than the object named on the left. In some cases the right object is only a little larger than the left object, while in other cases the right object is a great deal larger than the left object. In the space beside each pair, please write a number from 1 to 9 indicating *how much larger the right object is than the left object.*" The second set of instructions was identical except that "smaller" was substituted for "larger" throughout, and the positions of the words "right" and "left" were interchanged. In each version higher ratings corresponded to greater size differences. Both versions included instructions to consider only the average or most typical size of each object and to base the ratings on "your intuitive feeling" of the size differences, rather than objective differences in inches and feet.

One hundred thirty-eight Stanford University undergraduates completed the rating task, with half receiving the "larger" and half the "smaller" instructions.

Results and Discussion

The mean category ratings obtained in the two instruction conditions, plotted as a function of the ordinal size of the smaller member of each pair, are presented in Fig. 4. Pairs with ordinal differences of one, two, and three steps in the 12-term series are plotted separately. A very small congruity effect was obtained for these pairs. Subjects rated the size differences between small objects as relatively greater under the "smallness" instructions, while they rated the differences between large objects as relatively greater under "largeness" instructions. The apparent monotonic trends were statistically significant for ordinal distances of 1 $(F(1,1360) = 5.15, p < .025)$ and 2 $(F(1,1224) = 15.5, p < .001)$, but not for 3 $(F < 1)$. At higher distances (not plotted) the effect disappeared entirely. The obtained congruity effects were very small, reaching a maximum of .50 units at Distance 2.

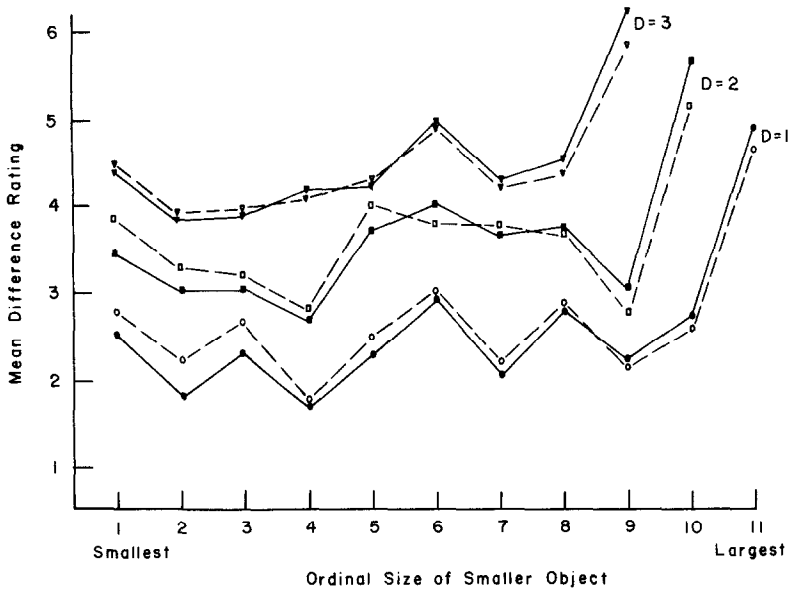Frequency counts analogous to those of previous experiments were

FIG. 4. Mean ratings obtained under "smallness" (− − −) and "largeness" (—) instructions, as a function of the ordinal size of the smaller object and of the ordinal size difference between objects (Experiment 3).

also performed. At each ordinal distance from 1 to 3, the number of subjects in each instruction condition who gave a higher rating to the smallest pair was compared to the number who gave a higher rating to the second largest pair. (Pairs including the largest item were not considered because they consistently yielded the highest ratings regardless of instructions.) While in each case the trends were in the direction predicted by discriminability models, the differences were small and did not approach significance at any ordinal distance. Thus, in Experiment 3, unlike the previous experiments, Equation 2 cannot be rejected as a description of the data.

As in Experiment 1A, the mean ratings for each instruction condition were fit by a nonmetric scaling program (in this case MDSCAL; Kruskal, 1964). For each data set a satisfactory unidimensional solution with stress less than .04 was obtained. The relative sizes of the 12 objects from each solution are given in Fig. 5. The two sets of values have been normalized by equating the scale values of the smallest and largest objects. Although the "smallness" and "largeness" scales show some small differences consistent with discriminability models, the reference point effects obtained in Experiment 3 can charitably be described as miniscule.

FIG. 5. Relative sizes of 12 objects derived from ratings obtained in "smallness" and "largeness" conditions (Experiment 3).

EXPERIMENT 4

Experiment 4 was performed to establish whether *explicit* reference points would have an effect on category ratings involving subjective size. Subjects were presented with word triplets, consisting of a comparison pair and a reference object, and were asked to judge how much closer in size one pair member was than the other to the reference point.

Method

Materials and design. The object names used in Experiment 4 were animal terms taken from the size norms of Holyoak and Mah (1981). These scale values were derived by successive interval scaling of rating data, and range from 0.0 to 10.0. Twenty-four basic test pairs (two sets of 12) were selected. Three triplets were created for each pair by varying the reference object. The three reference animals were either all larger (for 12 of the pairs) or all smaller than both pair members and were selected to have size values approximately 1, 2, or 3 scale units from that of the closer pair member. For example, the pair *bobcat–canary* was associated with the reference animals *alligator, cow,* and *hippo.* The size difference *between* the pair members was also varied. For the 12 basic pairs that were either smaller or larger than their respective reference animals, five pairs were approximately 1 scale unit apart, four were 2 units apart, and three were 3 units apart.

The combination of 24 pairs, each with three reference animals, created a set of 72 test triplets in which the reference animal was either larger or smaller than both pair members. For such triplets a judgment of the extent to which one pair member is closer than the other to the reference point could be reduced to a judgment of the size difference between the pair members. To make it less likely that such a strategy would be used, 18 additional triplets were added, in which one pair member was smaller than the reference animal and the other was larger. The ratings given to these filler items were not analyzed.

The 90 triplets were listed in a booklet, with 15 triplets on each page. Four different random orders were used for different subjects. A pair was presented no more than once, with one reference animal, on any given page. Each triplet was presented on a single horizontal line. The two pair members, separated by a hyphen, appeared at the left margin, and the name of the reference animal was presented near the center of the page. The left pair member was always the closer in size to the reference object.

Procedure and subjects. The written instructions included the following: "The third animal in each item should be regarded as a reference point to which the first two animals will be compared. In particular, the FIRST animal in each comparison pair is closer in size to the reference point animal than the SECOND animal is. Your task is to decide how much closer to the reference point the FIRST animal is." Subjects were told to use a 9-point rating scale, with a rating of 9 indicating the first animal was maximally closer than the second to the reference point. Fifty-three University of Michigan undergraduates served as paid subjects.

Results and Discussion

If subjects compared distances to the reference point, and these distances are more discriminable when the pair is relatively close to the reference point, then the ratings should tend to decrease as distance to the reference point is increased. This result was in fact obtained, with mean ratings of 5.53, 5.04, and 4.75 for reference points 1, 2, and 3 units, respectively, away from the closer pair member. This effect of the reference point was significant treating both subjects and items as random effects (Clark, 1973), $F'(2,103) = 16.5$, $p < .001$, and it did not vary systematically with the size difference between pair members or with the size of the reference point relative to the pair members. The size difference between the pair members also had a strong effect on the mean ratings, which increased from 4.27 to 6.10 as the size difference increased from one to three units.

It is unlikely that the observed reference point effects can be described by Equation 2. Since the comparison pairs were held constant as the reference stimuli were varied, the influence of the reference point cannot be attributed to variations in either the absolute size of the comparison stimuli or in their size differences. Furthermore, distance to the reference stimulus was not confounded with its absolute size, since the reference animals at each distance spanned comparable broad size ranges. The observed influence of reference points on the ratings appears to depend on the *relationship* between the magnitude of the reference point and the magnitudes of the stimuli being compared, as Equation 1 implies.

GENERAL DISCUSSION

Summary

Nonspeeded judgments of the subjective magnitudes of symbolic stimuli can be influenced by the reference point with respect to which the judgments are made. Magnitude differences between stimuli in the vicinity of the reference point are expanded relative to differences between stimuli far from the point. Judgments concerning the locations of cities along a west–east axis depended on which coast, Pacific or Atlantic, was specified as the reference point (Experiments 1 and 2). Furthermore, scales derived from the rating data were significantly correlated with the pattern of RTs obtained in a speeded comparison task in which the reference point was also varied (Experiment 1B). Similar effects on mean ratings were observed when the form of the comparative specified an implicit reference point at either end of a continuum of subjective size (Experiment 3); however, these effects were very small and did not clearly support a discriminability interpretation. Stronger evidence for discriminability effects was obtained when an explicit reference point was established at an arbitrary size value (Experiment 4).

Boundary Conditions on Discriminability Effects

Previous studies investigating nonspeeded judgments have not found evidence that reference points influence discriminability. In contrast, the present experiments, with the exception of Experiment 3, yielded clear positive results. We therefore need to consider the conditions under which discriminability effects apparently are or are not obtained. The Torgerson (1960) and Banks and Root (1979) studies both involved judgments about perceptual stimuli. It is possible that reference points have a greater influence on discriminability in symbolic than in perceptual tasks. It is more surprising that the study by Birnbaum and Mellers (1978) yielded no hint of discriminability effects, since their stimuli and procedure were very similar to those of our Experiment 1A. Birnbaum and Meller's data were highly reliable, yielded a location scale closely resembling our Neutral scale, and conformed to Equation 2. The discrepancy between the two sets of results may depend on subtle variations in the instructions. In the conditions most comparable to our own, Birnbaum and Mellers' subjects evaluated "differences" of either westerliness or easterliness. In contrast, our instructions never explicitly mentioned "differences"; rather, subjects were simply asked to rate "how much closer" to the reference point one city is than the other. Perhaps an emphasis on "differences" may dispose subjects to attend directly to the difference between the subjective locations of the two cities, rather than to their locations relative to the reference point. If so, the subjects' judgment process would follow Equation 2. Equation 1, derived from discriminability models, can be viewed as an extension of Equation 2, which is based on the subtractive theory of psychophysical judgment (Birnbaum & Veit, 1974; Veit, 1978). Equation 1 differs only in that each stimulus-to-reference-point distance is monotonically transformed prior to the subtraction operation. If a reference point is not provided or is ignored, or if the function f is linear, then Equation 1 will reduce Equation 2 as a special case. The data from our Neutral condition suggest that subjects are quite capable of evaluating location differences without comparison to a reference point.

The rating differences obtained in Experiment 3, in which reference points were manipulated by varying the comparative used in the instructions, seem too small to account fully for the congruity effects typically produced in RT tasks. Discriminability effects may be more pronounced for judgments based on linear arrays, in which the enditems provide salient polar reference points (Holyoak & Patterson, 1981). Research in this area has tended to focus on "single-factor" explanations of congruity effects within particular paradigms. Congruity effects, however, may be due to multiple factors, and their relative importance may vary across stimulus continua and experimental procedures.

Implicit Scaling Model of Reference Point Effects

What process model might account for the influence of reference points on discriminability? Most previous comparison models assume that variations in the form of the instructions do not influence the relative discriminability of the stimuli, but rather any of a variety of other processes such as encoding the items, retrieving their memory representations, translating semantic codes, or executing a response (Marschark & Paivio, 1979; Moyer & Dumais, 1978; Banks, Fujii, & Kayra-Stuart, 1976; Banks & Root, 1979). A discriminability model might assume that a reference point causes the subject to selectively attend to the region of the magnitude continuum in its vicinity; due to more efficient processing in the favored region of the scale, items with magnitudes similar to that of the reference point would be compared relatively quickly (Hutchinson & Lockhead, 1977). This type of differential attention model, however, has not been formulated very precisely; furthermore, one variant assumes the subject will quickly "readjust" to the range of the presented stimuli, in which case discriminability effects would not be expected in nonspeeded tasks (Kosslyn, Murphy, Bemesderfer, & Feinstein, 1978).

Yet a further approach would be to assume that for each continuum people code items on two magnitude scales in long-term memory (e.g., both "easterliness" and "westerliness" for cities) and that the items will be spaced differently on each scale. However, such a model does not provide a satisfactory account of the effects of explicit reference points, since it would be necessary to invoke a new scale to account for the facilitated processing of items in the vicinity of every arbitrary reference point. A more promising tack is to develop a model based on the premise that only a single set of scale values is stored in long-term memory, which can be transformed in working memory due to the influence of a reference point. One version of such a model assumes that subjects are implicitly assessing the ratio of the lesser to the greater difference between a stimulus item and a reference point (Jamieson & Petrusic, 1975). If subjective continua are represented psychologically as interval scales, then such judgments about ratios of differences (unlike direct judgments of magnitude ratios) are both conceptually meaningful and empirically possible (Hagerty & Birnbaum, 1978; Veit, 1978).

But while several studies have found that RT is a linear function of the ratio of differences (e.g., Holyoak, 1978), this result is not always obtained (te Linde & Paivio, 1979). The ratio of differences is equivalent to a very specific form of Equation 1, in which f is logarithmic and J_k is exponential. Observed response patterns sometimes strain this form of the equation. In particular, if one of the items in a pair is located at the specified reference point, the magnitude of the farther item should not influence either choice RT or distance ratings, since the ratio of differ-

ences will always be zero. However, distance effects are sometimes obtained even for pairs that include a congruent enditem. The model could assume the reference points are more extreme than any items in the test set (cf. Holyoak, 1978, footnote 2). However, in Experiment 1A one would have expected the subjective positions of the two reference points (the coastlines) to have been virtually identical with the positions of the two extreme cities. The distance of the farther city should therefore have had little or no effect for pairs that included the city located on the specified coast; but in fact, the mean ratings showed a large monotonic increase. It therefore seems preferable to develop a model consistent with the general form of Equation 1, which can account for the data pattern associated with the ratio of differences as a special case.

An *implicit scaling model* can be used to derive Equation 1. There is abundant evidence that category ratings are influenced by contextual effects, as well as by the psychophysical function for the stimuli. Range-frequency theory (Parducci, 1965; Parducci & Perrett, 1971) predicts category judgments as a weighted sum of the psychophysical function and the cumulative density function of the stimulus distribution on the magnitude continuum. Subjects tend to place their category boundaries closer together in those regions of the magnitude range in which the greatest numbers of items are presented.

A model incorporating similar assumptions can be formulated to explain reference point effects. Suppose that the comparison process in a choice RT paradigm involves implicit category scaling. That is, the subject will recode the magnitude value of each item, retrieved from long-term memory, into a value on an internal category scale, which will then be available in working memory. (This implicit scale may be limited to 7 ± 2 values, as suggested by Miller, 1956, in his review of work on absolute judgment.) These category values are assumed to have interval-scale properties, so that the difference between the values for a pair of items can be assessed using the type of sequential sampling process postulated by analog comparison models.

To account for the influence of reference points on relative discriminability, the implicit scaling model asssumes that the width of the categories on the internal scale will increase with distance from a reference point. As a result, implicit category scaling will in effect transform the underlying continuum, stretching the perceived distances between stimuli in the vicinity of a reference point relative to distances between stimuli far from it. This mechanism corresponds to the function f in Equations 1 and 3. The resulting variations in relative discriminability will influence judgments in an overt category rating task, so that reference point effects will be obtained with nonspeeded as well as speeded judgments.

The implicit scaling model seems to provide a reasonable account of all

the present results. In particular, the model can accommodate the observed variations in the magnitude of reference point effects. The precise form of the scale transformation effected by a reference point may depend on a variety of factors. If for whatever reason the implicit scaling process is *not* influenced by a reference point, the model predicts that the response pattern will reduce to Equation 2, in accord with the psychophysical theory of Birnbaum (1978) and Veit (1978). The model suggests that distributional effects could be obtained in symbolic comparison tasks. For example, the same items should be more discriminable when presented in an item set that spans a narrow rather than a broad magnitude range, a prediction for which there is some evidence (Hinrichs, Yurko, & Hu, 1981). Finally, the proposal is readily compatible with the type of analog comparison process that many theorists have advocated.

Magnitude Codes in Long-Term and Working Memory

The implicit scaling model assumes that symbolic magnitudes are coded on a coarse-grained interval scale in working memory, but on a relatively fine-grained interval scale in long-term memory. But if magnitude values in working memory are relatively limited in their resolution, what use is there in having more precise values stored in long-term memory? Also, how can such relatively precise values ever be learned? In general, why is working memory not a "bottleneck" that limits the resolution of magnitude information being stored in or retrieved from long-term memory (cf. Banks, 1977)?

The solutions to these twin puzzles are essentially the same, namely, the introcution of sequential sampling mechanisms. Consider first the process of retrieving symbolic magnitudes, as is required in a relative judgment task. The implicit scaling model, like other analog models, assumes that the comparison process involves an iterative process of retrieving magnitude values for the two items, recoding them into working memory, and assessing the difference between them (Holyoak & Patterson, 1981). While on each cycle of this process the available magnitude codes for the two items will be relatively coarse grained, over repeated iterations the estimated difference between them will become more precise. For example, suppose a person is asked to decide which is larger, an ant or a pea. On any given cycle the two items may receive identical magnitude codes in working memory. If the size values associated with the two concepts in long-term memory are in fact discriminable, however, *pea* will receive a relatively high value on the scale in working memory more often than *ant* will. As a result, a random walk or similar sequential sampling procedure will in general eventually lead to a confident decision that peas are larger than ants.

A similar sequential sampling mechanism can describe how subjective

magnitudes could be learned (an issue that has received little previous discussion). The first time an object (or member of a particular category) is encountered, a coarse magnitude value may be stored in long-term memory for any continuous dimension that is encoded. Subsequent exposures can then be used to revise the initially imprecise estimate. Just as a series of observations can be used to estimate the mean for a population, this sampling process will yield increasingly precise estimates of the mean magnitude value of the stimulus. In the case of a category of stimuli, the resulting representation may more generally take the form of a parametric description of the distribution of magnitude values for the category instances, as is assumed by the category density model of classification learning proposed by Fried and Holyoak (Note 2).

This conception of the learning process suggests how subjective magnitudes might come to deviate systematically from the veridical values. The magnitude of values stored in long-term memory will necessarily depend on the values encoded in working memory each time the stimulus is encountered. It follows that if the codes in working memory are consistently biased in a particular way, the long-term codes will eventually reflect this bias. This suggests a link between two types of distortions observed in Experiments 1A and 2. You may recall that in addition to the instructional effects produced by specifying either the Atlantic or the Pacific coast as a reference point, the data provided suggestive evidence that distances between cities near the locale of our subjects were overestimated (as did the data of Birnbaum & Mellers, 1978). In terms of the implicit scaling model, the former effect reflects transient alterations of the discriminability of location codes in working memory, while the latter effect reflects a stable distortion of the subjective locations stored in long-term memory. When the learning process is considered, it seems that the latter effect could be a product of the former. If people evaluate relative locations with respect to the place they live, distances between nearby locations will seem subjectively larger than comparable distances between remote locations. If the same place is habitually used as a reference point, this distortion will eventually be reflected in the "mental map" stored in long-term memory. Reference point effects may therefore reflect an ecologically important flexibility in people's ability to distribute a limited discriminatory capacity over a given magnitude range. In the geographic domain it is usually desirable to make relatively fine discriminations among distances between nearby locations. In most circumstances the apparent corollary, that local distances are subjectively "stretched" relative to remote ones, is unlikely to have serious adverse consequences.

Another possible explanation of the influence of habitual reference points, also consistent with the implicit scaling approach, is based on the

observation that people will tend to be familiar with more places located in the vicinity of their reference location than elsewhere. If locations serve as retrieval cues for nearby places, the subjective density of known locations will tend to be relatively great in the reference region. Range-frequency theory predicts that response categories will be relatively narrow when placed in a dense region of a continuum; accordingly, an implicit scaling process might stretch the subjective distances between locations in the reference region. The density factor could also account for the consistent underestimation of the distance between Salt Lake City and Denver that we observed in Experiments 1A and 2. In the cognitive maps of our Midwestern subjects, little may lie between these two cities except for a narrow line of Rocky Mountains. (See Thorndyke, 1981, for evidence that intervening locations influence distance judgments.)

Reference Points and Asymmetrical Distance Judgments

The effects of reference points observed in the present study add to other factors that have been shown to distort judgments involving geographical locations (Stevens & Coupe, 1978; Tversky, 1981). The apparent effect of an habitual reference point can be related to the influence of "landmarks" on the acquisition and use of spatial knowledge (Evans, 1980). Sadalla, Burroughs, and Staplin (1980) recently demonstrated that habitual reference points (the most familiar locations on a college campus) produce asymmetrical distance judgments: an ordinary place is judged closer to a reference point than vice versa. This result parallels similar effects on similarity judgments obtained in nongeographical domains, in cases in which one stimulus in a pair is more "prototypical" or "prominent" than the other (Rosch, 1975; Tversky & Gati, 1978).

The Sadalle et al. demonstration implies that judgments of remembered physical distances violate a distance axiom. An important question concerns the locus of the asymmetry effect—does it involve storage or retrieval? If one location is a reference point and another is not, do people have two different interlocation distances stored in long-term memory? Or could there be a single subjective distance between the two locations, which can be transformed in different ways during the judgment process? The framework provided by the implicit scaling model is more compatible with the latter possibility. Consider a task in which subjects make a series of judgments about pairs of remembered locations, in each case evaluating "how close A is to B," so that B serves as the standard to which A is being compared. The implicit scaling model assumes that each distance judgment will involve recoding the location values of A and B, stored in long-term memory, onto an implicit category scale in working memory. On each trial the small set of possible values on this category scale must be distributed across the entire subjective stimulus range. Since a wider

range will yield reduced discriminability, the distance *A–B* will be rated as smaller if the subjective range seems relatively wide.

We assume that the subjective range may vary from trial to trial and can be influenced by the range of locations called to mind by the two test stimuli, and especially by the standard, location *B*. By definition, an habitual reference point is a location with respect to which many other locations have been coded. It follows that a reference point presented as the standard will be a better retrieval cue for other locations than will an ordinary place presented as standard. A reference point standard will therefore be more likely to trigger recall of some remote location that will increase the subjective range. It follows that the subjective range will tend to be greater when a reference point rather than an ordinary place is presented as the standard, so that the *A–B* distance will seem relatively small in the former case. An implicit scaling process could thus produce the asymmetries in distance judgments observed by Sadalla et al.

While research to date has identified a variety of phenomena attributable to the influence of reference points on symbolic comparisons, a great deal remains to be learned about the processes that mediate these effects. The implicit scaling model, while closer to a working hypothesis than to a full-fledged theory, at least provides a tentative framework within which it may be possible to integrate various phenomena associated with reference points. The basic assumption of the model is that relative judgments can reflect not only information stored in long-term memory, but also transformations induced by the judgment process. The distorting mirror of memory need not deceive us, if we can learn how it is fashioned.

REFERENCES

Banks, W. P. Encoding and processing of symbolic information in comparative judgments. In G. H. Bower (Ed.), *The psychology of learning and motivation.* New York: Academic Press, 1977. Vol. 11.

Banks, W. P., Fujii, M., & Kayra-Stuart, F. Semantic congruity effects in comparative judgments of magnitudes of digits. *Journal of Experimental Psychology: Human Perception and Performance,* 1976, **2,** 435–447.

Banks, W. P., & Root, M. Semantic congruity effects in judgments of loudness. *Perception & Psychophysics,* 1979, **26,** 133–142.

Bartlett, F. C. *Remembering.* Cambridge, MA: Cambridge Univ. Press, 1932.

Baum, D. R., & Jonides, J. Cognitive maps: Analysis of comparative judgments of distance. *Memory & Cognition,* 1979, **7,** 462–468.

Birnbaum, M. H. Differences and ratios in psychological measurement. In N. J. Castellan & F. Restle (Eds.), *Cognitive theory.* Hillsdale, NJ: Erlbaum, 1978. Vol. 3.

Birnbaum, M. H., & Mellers, B. A. Measurement and the mental map. *Perception & Psychophysics,* 1978, **23,** 403–408.

Birnbaum, M. H., & Veit, C. T. Scale convergence as a criterion for rescaling: Information integration with difference, ratio, and averaging tasks. *Perception & Psychophysics,* 1974, **15,** 7–15.

Bock, R. D., & Jones, L. V. *The measurement and prediction of judgment and choice*. San Francisco: Holden–Day, 1968.

Buckley, P. B., & Gillman, C. B. Comparisons of digits and dot patterns. *Journal of Experimental Psychology*, 1974, **103**, 1131–1136.

Carmichael, L., Hogan, H. P., & Walter, A. A. An experimental study of the effect of language on the reproduction of visually perceived form. *Journal of Experimental Psychology*, 1932, **15**, 73–86.

Clark, H. H. The language-as-fixed-effect fallacy: A critique of language statistics in psychological research. *Journal of Verbal Learning and Verbal Behavior*, 1973, **12**, 335–359.

Evans, G. W. Environmental cognition. *Psychological Bulletin*, 1980, **88**, 259–287.

Evans, G. W., & Pezdek, K. Cognitive mapping: Knowledge of real-world distance and location information. *Journal of Experimental Psychology: Human Learning and Memory*, 1980, **6**, 13–24.

Franks, J. J., & Bransford, J. D. Abstraction of visual patterns. *Journal of Experimental Psychology*. 1971, **90**, 65–74.

Hagerty, M., & Birnbaum, M. H. Nonmetric tests of ratio vs. subtractive theories of stimulus comparison. *Perception & Psychophysics*, 1978, **24**, 121–129.

Hinrichs, J. V., Yurko, D. S., & Hu, J.-M. Two-digit number comparison: Use of place information. *Journal of Experimental Psychology: Human Perception and Performance*, 1981, **7**, 890–901.

Holyoak, K. J. Comparative judgments with numerical reference points. *Cognitive Psychology*, 1978, **10**, 203–243.

Holyoak, K. J., & Mah, W. A. Semantic congruity in symbolic comparisons: Evidence against an expectancy hypothesis. *Memory & Cognition*, 1981, **9**, 197–204.

Holyoak, K. J., & Patterson, K. K. A positional discriminability model of linear-order judgments. *Journal of Experimental Psychology: Human Perception and Performance*, 1981, **7**, 1283–1302.

Holyoak, K. J., & Walker, J. H. Subjective magnitude information in semantic orderings. *Journal of Verbal Learning and Verbal Behavior*, 1976, **15**, 287–299.

Hovland, C. I., & Sherif, M. Judgmental phenomena and scales of attitude measurement: Item displacement in Thurstone scales. *Journal of Abnormal and Social Psychology*, 1952, **47**, 822–832.

Hutchinson, J. W., & Lockhead, G. R. Similarity as distance: A structural principle for semantic memory. *Journal of Experimental Psychology: Human Learning and Memory*, 1977, **3**, 660–678.

Jamieson, D. G., & Petrusic, W. M. Relational judgments with remembered stimuli. *Perception & Psychophysics*, 1975, **18**, 373–378.

Kahneman, D., & Tversky, A. Prospect theory: An analysis of decision under risk. *Econometrica*, 1979, **47**, 263–291.

Kelley, H. H., Hovland, C. I., Schwartz, M., & Abelson, R. P. The influence of judges' attitudes in three methods of attitude scaling. *Journal of Social Psychology*, 1955, **42**, 147–158.

Kerst, S. M., & Howard, J. H., Jr. Memory psychophysics for visual area and length. *Memory & Cognition*, 1978, **6**, 327–335.

Kosslyn, S. M., Murphy, G. L., Bemesderfer, M. E., & Feinstein, K. J. Category and continuum in mental comparisons. *Journal of Experimental Psychology: General*, 1977, **106**, 341–375.

Kruskal, J. B. Multidimensional scaling by optimizing goodness of fit to a nonmetric hypothesis. *Psychometrika*, 1964, **29**, 1–27.

Loftus, E. F., & Palmer, J. C. Reconstruction of automobile destruction: An example of the interaction between language and memory. *Journal of Verbal Learning and Verbal Behavior,* 1974, 13, 585–589.

Marschark, M., & Paivio, A. Semantic congruity and lexical marking in symbolic comparisons: An expectancy hypothesis. *Memory & Cognition,* 1979, 7, 175–184.

Miller, G. A. The magical number seven, plus or minus two: Some limits on our capacity for processing information. *Psychological Review,* 1956, 63, 81–97.

Moyer, R. S., Bradley, D. R., Sorenson, M. H., Whiting, J. C., & Mansfield, D. P. Psychophysical functions for perceived and remembered size. *Science,* 1978, 200, 330–332.

Moyer, R. S., & Dumais, S. T. Mental comparison. In G. H. Bower (Ed.), *The psychology of learning and motivation.* New York: Academic Press, 1978. Vol. 12.

Moyer, R. S., & Landauer, T. K. Time required for judgements of numerical inequality. *Nature (London),* 1967, 215, 1519–1520.

Paivio, A. Perceptual comparisons through the mind's eye. *Memory & Cognition,* 1975, 3, 635–647.

Parducci, A. Category judgment: A range-frequency model. *Psychological Review,* 1965, 72, 407–418.

Parducci, A., & Perrett, L. F. Category rating scales: Effects of spacing and frequency of stimulus values. *Journal of Experimental Psychology Monograph,* 1971, 89, 427–452.

Rosch, E. Cognitive reference points. *Cognitive Psychology,* 1975, 7, 532–547.

Sadalla, E. K., Burroughs, W. J., & Staplin, L. J. Reference points in spatial cognition. *Journal of Experimental Psychology: Human Learning and Memory,* 1980, 6, 516–528.

Segall, M. H. The effect of attitude and experience on judgments of controversial statements. *Journal of Abnormal and Social Psychology,* 1959, 58, 61–68.

Stevens, A., & Coupe, P. Distortions in judged spatial relations. *Cognitive Psychology,* 1978, 10, 422–437.

Te Linde, J., & Paivio, A. Symbolic comparison of color similarity. *Memory & Cognition,* 1979, 7, 141–148.

Thorndyke, P. W. Distance estimation from cognitive maps. *Cognitive Psychology,* 1981, 13, 526–550.

Torgerson, W. S. Quantitative judgment scales. In H. Gulliksen & S. Messick (Eds.), *Psychological scaling: Theory and applications.* New York: Wiley, 1960.

Tversky, A., & Gati, I. Studies of similarity. In E. Rosch & B. B. Lloyd (Eds.), *Cognition and categorization.* Hillsdale, NJ: Erlbaum, 1978.

Tversky, B. Distortions in memory for maps. *Cognitive Psychology,* 1981, 13, 407–433.

Veit, C. T. Ratio and subtractive processes in psychophysical judgment. *Journal of Experimental Psychology: General,* 1978, 107, 81–107.

REFERENCE NOTES

1. Noma, E., & Johnson, J. *Confirmatory multidimensional scaling.* Human Performance Center Technical Report No. 60, University of Michigan, August 1977.

2. Fried, L. S., & Holyoak, K. J. *Induction of category distributions: A framework for classification learning.* In preparation, 1981.