JOURNAL OF VERBAL LEARNING AND VERBAL BEHAVIOR **15**, 287–299 (1976)

Subjective Magnitude Information in Semantic Orderings

KEITH J. HOLYOAK AND JANET H. WALKER

Stanford University

Subjects compared the magnitudes of pairs of concepts from the semantic orderings of time, quality, and temperature, choosing either the concept that was *longer/better/warmer* or the concept that was *shorter/worse/colder*. Decision time depended on two factors. First, decision time decreased as the subjective difference between the magnitudes of the two concepts being compared increased. Second, the decision was made more quickly when the form of the comparative in the question was congruent with the scale position of the terms being compared. For example, it was easier to choose the *longer* term of the pair *decade–century*, and the *shorter* term of the pair *second–minute*. The results demonstrate that the semantic representations of ordered terms contain subjective magnitude information.

Recent studies of mental comparisons have revealed an intriguing phenomenon termed the "symbolic distance effect" by Moyer and Bayer (in press). The basic phenomenon is that the greater the psychological difference between members of a pair of symbols, the faster people can compare their magnitudes. For example, Moyer and Landauer (1967) presented subjects with pairs of digits, and found that decisions about which digit was larger were faster if the difference between the digits was large (e.g., *2* and *8*) rather than small (e.g., *7* and *8*). In addition to other digit studies (Buckley & Gillman, 1974; Parkman, 1971; Sekuler, Rubin, & Armstrong, 1971), distance effects have been reported for alphabetic comparisons (Parkman, 1971; Lovelace & Snodgrass, 1971), comparisons of the normative size of objects (Moyer, 1973; Paivio, 1975), and comparisons of items learned in an arbitrary linear ordering along some dimension (Potts, 1972, 1974; Moyer & Bayer, in press).

While the existence of the distance effect has been clearly established, the explanation for it is less certain. In fact, it is not clear that a single explanation will be adequate for all forms of the phenomenon. For comparisons of items from arbitrary orderings, it has been suggested that a spatial representation of the complete ordering underlies the distance effect (Potts, 1972, 1974; Trabasso & Riley, 1975). For comparisons of digits and of animal sizes, the emphasis has been on information about the absolute magnitude of individual items (Moyer, 1973; Moyer & Landauer, 1967; Paivio, 1975). For example, Moyer (1973) proposed that people compare the sizes of animals from memory by making "internal psychophysical judgments" on the basis of analogue representations that preserve animal size.

The present study investigated the possibility that such absolute magnitude information may also be used in comparisons involving natural language concepts that are ordered along continuous semantic dimensions. The orderings studied were the terms commonly used to

The order of authorship was determined alphabetically. We thank Gordon Bower, Herbert Clark, Arnold Glass, Roger Shepard, Edward Smith, Fred Woocher and Glenn Kleiman for their helpful comments, and Wayne Miller for his help in testing subjects. This research was supported by NIMH Grants MN 13950-06 to Gordon H. Bower and MH20021 to Herbert H. Clark and by a Canadian National Research Council Postdoctoral Fellowship to Janet Walker. After September 1976, send reprint requests to Dr. Keith J. Holyoak, Human Performance Center, University of Michigan, 330 Packard Road, Ann Arbor, Michigan 48104.

express time (e.g., *second*, *day*), quality (e.g., *good*, *poor*), and temperature (e.g., *hot*, *cool*). These concepts, unlike other items for which a symbolic distance effect has been reported, appear to be ordered on the basis of linguistic knowledge. The orderings of arbitrary items and of the letters of the alphabet have no semantic basis. The digits have a necessary ordering, but their meanings depend on the rules of arithmetic rather than of language. In the case of size comparisons, Paivio's (1975) finding that judgments can be made more quickly when the objects are presented as pictures rather than as words suggests that size information is more directly related to the perceptual than to the linguistic system. While some word meanings do form semantic contrasts on the size dimension (e.g., *mountain-hill*), such pairs can in fact be compared more quickly than would be predicted from their subjective size differences (Holyoak, 1976).

The terms of time, quality, and temperature each form an ordered scale, although these scales have somewhat different semantic properties. The quality and temperature terms form polar oppositions, while the time terms form a hierarchic scale (Leech, 1974). A polar scale has a positive and a negative end (e.g., *good-bad*), and a "middle ground" which is theoretically neutral or normative (e.g., *average*). The temperature and quality terms refer to (possibly overlapping) regions of the scale, rather than to clearly defined points. In contrast, a hierarchic scale such as time has no neutral midpoint, and the terms are categorically exclusive. It is possible that the results from a magnitude comparison task might differ for these two types of semantic scales.

Models of the Comparison Process

The present study used the magnitude comparison task to investigate the semantic representation of ordered concepts. The questions of interest concerned the form of the magnitude information included in the meaning of words, and how such information is used in making comparative judgments. We will outline three possible models of the comparison process for semantic orderings, and the predictions of each for the magnitude comparison task. These models will be referred to as the analogue model, the grouping model, and the relational model.

The fundamental assumption of the analogue model is that the semantic representation of each concept includes absolute magnitude information (i.e., a measure independent of those for other terms on the scale).[1] A person can judge the relative magnitude of two concepts by comparing the absolute magnitude information stored with each. Many theorists who have assumed that analogue information is used in magnitude comparisons (e.g., Moyer, 1973; Moyer & Landauer, 1967) have not proposed an explicit process model, but have simply suggested an analogy between symbolic and perceptual comparisons. However, Buckley and Gillman (1974) and Luce and Green (1972) have proposed analogue comparison models that can account for distance effects. The basic assumption underlying these proposals is that retrieval of magnitude information is a continuous process. While in the analogue model information retrieved at one point in the process does not differ qualitatively from information available at a different point, it is assumed that the precision of the information increases with time. The initial information will be sufficient to distinguish between concepts very different in magnitude (e.g., *day-decade*), but more precise information (which will take longer to retrieve) will be necessary to distinguish between concepts that are relatively similar in magnitude (e.g., *day-week*). The comparison

[1] The term "analogue" will be used throughout to contrast with "discrete" or "categorical". In this sense it will be used interchangeably with "continuous". By "analogue value" we mean a point on a continuous dimension, not a numerical value. The term "absolute" will be used to contrast with "relative". We do not intend it to imply anything about the veridicality of the information.

process will therefore produce a symbolic distance effect.

The grouping model is a categorical model for mental comparisons. It assumes that terms are categorized, as well as ordered relative to each other, and that the category information is more available. For example, our quality terms might be coded in memory in three groups: *perfect*, *excellent*, and *good* as one group, *average* and *fair* as another, and *poor* and *awful* as the third. The grouping model predicts that between-group comparisons (e.g., *perfect–average*) will be relatively fast, since these decisions can be made on the basis of easily retrieved categorical information. However, within-group comparisons (e.g., *perfect–good*) will be relatively slow, since such decisions will depend on accessing the less available information about the ordering of terms within a cluster.

If the group boundaries are the same for all subjects on all trials, the grouping model predicts a discrete step function for the distance effect: between-group comparisons will be uniformly fast, while within-group comparisons will be uniformly slow. However, a weaker version of the model might assume that the group boundaries fluctuate from subject to subject or from trial to trial. The resulting distance function would then be continuous. The more similar the magnitudes of two concepts, the more likely they will be categorized together, and hence the slower the mean reaction time to compare them. It can be argued that the assumption of fluctuating groupings is somewhat questionable for semantic orderings. Since magnitude relations between ordered and semantic concepts are not arbitrary, but rather form part of the meanings of the words, it would be surprising if different speakers categorized these concepts in radically different ways.

If the flexible grouping assumption is nevertheless adopted, the predictions of the grouping and of the analogue model become very similar. However, the two models suggest that the distance effect will be controlled by different factors. The grouping model predicts that decision time will increase with the probability that two concepts are categorized together, while the analogue model predicts that decision time will increase with the subjective similarity between the magnitudes of the two concepts. While these two measures are certain to be highly correlated, it is possible that one will be a better predictor of decision time than the other. To permit this comparison of the models, independent measures of groupings and subjective distances were collected.

Both of the above models predict a symbolic distance effect, and both assume that comparisons based on semantic orderings involve processes similar to those used in comparisons based on nonlinguistic orderings. In contrast, the relational model does not predict a distance effect for semantic orderings. In its general form, this model assumes that people store only information about the relative magnitude of one concept to another, rather than the absolute magnitudes of individual concepts. Accordingly, the model does not predict any necessary relationship between comparison time and distance. A more specific version of the relational model might assume that people must decompose the meaning of a concept into a relational definition in order to perform a magnitude comparison. This possibility can be illustrated most clearly for the time scale, for which most intervals are defined in terms of the next smaller interval; e.g., an hour is 60 min, a week is 7 days. If these definitions are the basis for the comparison process, these adjacent pairs should be maximally easy, since the relative magnitude of the two terms will be retrieved immediately. Nonadjacent pairs, on the other hand, will require some additional computation and therefore produce longer decision latencies.

Do Semantic Orderings Show a Congruence Effect?

The representation of semantic orderings in memory may be further clarified by comparing different forms of the magnitude question. Clark (1969) has proposed a principle of "semantic congruence," according to which a comparative judgment can be made more easily if the form of the question matches the form of the information that must be evaluated. In dealing with arbitrary ordered relationships, for example, decisions about "who is taller" are faster than decisions about "who is shorter" when the relationships are learned in the form "Joe is taller than Henry". However, when the original information is presented as "Henry is shorter than Joe", decisions about "who is shorter" are easier.

Banks, Clark, and Lucy (1975) found this congruence effect in perceptual comparisons of the relative height of circle-and-line figures that were described to subjects either as "balloons" or as "yo-yos". Subjects had to choose either the figure that was "higher" or the figure that was "lower." For the balloons, people were faster to choose which one was higher; for the yo-yos, they were faster to choose which was lower. Assuming that balloons are normally coded in terms of "highness" and yo-yos in terms of "lowness," the semantic congruence principle predicts this result. The Banks et al. finding demonstrates that an apparently simple perceptual comparison can be affected by linguistic variables.

It is possible that the comparisons between semantically-ordered concepts will also be affected by the congruence of information and question. For example, suppose that terms for long time intervals (*decade*, *century*) can be compared more easily if the question is *Which is longer?*, but that terms for short intervals (*second*, *minute*) can be compared more easily for the question *Which is shorter?* An interaction between the comparative used in the question and the scale position of the concepts being compared may have important implications for the nature of the magnitude information contained in representations of ordered concepts. The congruence effect is not predicted by any of the models so far outlined for the comparison process, since additional assumptions about the congruence of question and stimulus have not been specified. The integration of congruence and comparison models will be postponed until after discussion of the results.

METHOD

Pairs of words drawn from the natural-language scales for time, quality, and temperature were presented to subjects who were timed as they chose the term that was greater or lesser in magnitude. There were 11 time intervals: *millenium*, *century*, *decade*, *year*, *month*, *week*, *day*, *hour*, *minute*, *second*, *millisecond*; seven quality terms: *perfect*, *excellent*, *good*, *average*, *fair*, *poor*, *awful*; and six temperature terms: *torrid*, *hot*, *warm*, *cool*, *cold*, *frigid*. The terms were chosen for their familiarity and obvious ordering, excluding terms like *fortnight* and *lukewarm*. The end terms on each scale were possibly less familiar, but were included to minimize the effect of the "end anchor" strategy found in studies with artificial materials (e.g., Potts, 1974), in which no distance effect is obtained for pairs involving end terms.

Subjects saw all possible pairs of items in both orders (e.g., both *second–minute* and *minute–second*). The pairs were typed side-by-side in uppercase letters on white cards, with a double space separating the two words. The trials for each scale were presented in a separate block, with 110 trials for the time scale, 42 for the quality scale, and 30 for the temperature scale. Each subject was asked to decide either which term was *longer/better/warmer* (the unmarked question) or which term was *shorter/worse/colder* (the marked question) (see Clark, 1969, for a discussion of linguistic markedness). Half of the subjects chose the unmarked member of the pair for all scales, and half chose the marked member.

Pairs were presented in a tachistoscope with the subject initiating each trial by pressing a "ready" button. After a 1-second delay the word pair appeared and the subject pressed the response button (left or right) which corresponded to the position of the word which was greater (or lesser) in magnitude. The display disappeared with the response and was replaced by a lighted field with a fixation cross. The position of the fixation cross corresponded to the space between the words. Subjects were instructed to respond as quickly as possible while maintaining high accuracy. There were no reading order instructions.

At the beginning of the session, subjects were shown all three scales to check their agreement with the orderings. The terms were typed in order in a column, with the most extreme term (marked or unmarked, depending on the question form) at the top. All subjects agreed with the orderings. Twenty practice trials consisting of magnitude comparisons for pairs of digit names (e.g., *three–seven*) preceded the test trials. The markedness of the question for the practice trials (*bigger* or *smaller*) was the same as for the test trials. The order of the three test blocks was counterbalanced across subjects and the order of trials within each block was randomized individually for each subject. Twenty-four Stanford University undergraduates participated either for pay or for course credit.

Two other independent groups of subjects completed questionnaires in a grouping task and in a rating task designed to study the properties of the scales. In the grouping task, the terms on the scales were typed in order in three columns, with the most positive term at the top. Subjects were instructed to divide the words into groups according to their similarity of meaning, by drawing in the group boundaries. They were told to use from two to four groups for each scale, depending on how many groups seemed appropriate. Seventy-seven subjects completed the grouping questionnaire. In the rating task, all 91 pairs of words (one order for each pair) were typed in a four-page booklet. The order of the pairs was random for each scale, and the order of the three scales was counterbalanced across subjects. Each word appeared equally often on the left- or right-hand side of a pair. Subjects were instructed to rate each pair on a seven-point scale indicating how far apart the concepts were in subjective magnitude, with a rating of 7 indicating maximum distance. Twenty-one subjects provided distance ratings.

RESULTS

Overview of the Analyses

The results for the three scales will be reported separately. For each scale, two sets of analyses of variance were performed on the decision time data. The first analysis examined decision time as a function of the ordinal difference between the terms in a pair (distance or step size). For each scale an unweighted means analysis of variance for distance was performed both with and without pairs which included the end terms in the ordering. In no case was the overall pattern of results affected by inclusion of the end terms. Unlike results obtained with artificial materials, end term pairs in the present study were not always compared most quickly, and there was an effect of distance for pairs which included an end term. The second unweighted means analysis of variance examined the effect of question markedness on decision time for pairs of terms adjacent in the orderings. In both sets of analyses the order of terms in a pair (e.g., *hour–day* vs. *day–hour*) was included as a factor, but since order did not affect the results in any systematic way, the results will be reported collapsed across this variable.

In addition, stepwise multiple regression analyses were used to compare grouping measures, rated distance, and ordinal distance as predictors of the overall decision time data (collapsing over the two forms of the question). In each analysis we determined whether the partial correlation of each variable with decision time was significant after the variance attributable to the correlated variables was accounted for. If this partial correlation was significant for one of the variables, but not for any others, it was concluded that the former variable accounted for a significant amount of variance that could not be explained by any of the remaining factors.

The results from the grouping task were scored by calculating for each pair the proportion of times subjects placed the terms in the same group. For the rating task, the mean rating of the difference in magnitude was calculated for each pair of items. In order to ensure that the reliabilities of the two sets of norms were comparable, split-half correlations were calculated for each measure for each of the three scales. These reliability estimates were very high for both measures, with correlations ranging from .96 to .99 for the grouping measure and from .95 to .98 for the distance measure.

In order to obtain a visual representation of the subjective distances between the concepts, the rated distances were subjected to nonmetric multidimensional scaling of the sort developed by Shepard (1962) and Kruskal (1964). We obtained best-fitting one-dimensional spatial representations of each scale using Kruskal's M-D-SCAL 5M (with "stress formula 1" and the "primary" approach to ties). To avoid entrapment in local minima, we specified an evenly-spaced initial configuration with the items ordered from least to greatest in magnitude. For all three scales the one-dimensional solution had acceptably low stress. However, the nonmetric solution for the temperature scale was "degenerate" (Shepard, 1962), with the terms dividing into two groups (*torrid/hot/warm* and *cool/cold/frigid*), such that distances were larger for all between-group comparisons than for any within-group comparison. The ordinal information in the data was therefore insufficient to determine a fully metric representation of the temperature scale. We therefore obtained a metric solution for temperature by finding the best fit to a linear function. Figure 1 displays the scaled distances and corresponding stress values for the best nonmetric solutions for the time and quality scales, and the best metric solution for the temperature scale.

Multiple regression analyses were used to compare the effectiveness of rated and scaled distance as predictors of decision time (collapsing over the two forms of the comparative) for each scale. The two distance measures were quite similar in their predictive power: rated distance was somewhat better for the time scale, but scaled and rated distance were equally good predictors of decision time for the other two scales. The scaled distances displayed in Figure 1 therefore seem to capture the predictive properties of the ratings.

The Time Scale

The overall distance effect for the time scale is shown in Figure 2, both including and excluding the end terms *millenium* and *millisecond*. In both cases, decision time decreased with step size, $F(7, 154) = 21.6$, $p < .001$ without the end terms and $F(9, 198) = 9.02$, $p < .001$ including the end terms. Decision time decreased by 180 msec as step size increased from one to four and then effectively leveled off. The slight increase in decision time at longer step sizes when end terms are included is probably attributable to the greater proportion of trials including one of the relatively less familiar end terms. There

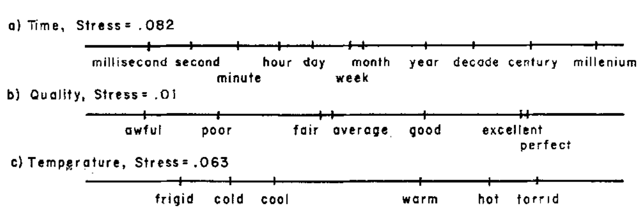

FIG. 1. Nonmetric one-dimensional scaling solutions for the time and quality terms, and the metric one-dimensional scaling solution for the temperature terms.

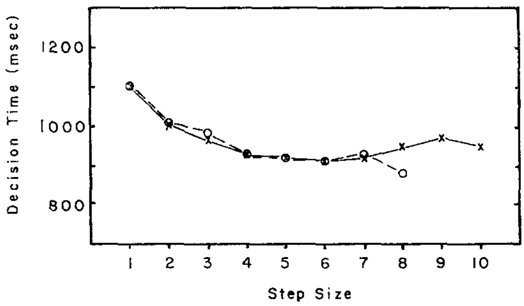

FIG. 2. Decision time as a function of step size for the time terms, including the end terms (×—×), and excluding the end terms (○- -○).

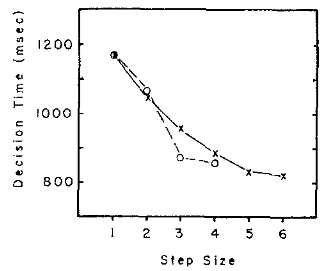

FIG. 4. Decision time as a function of step size for the quality terms, including the end terms (×—×), and excluding the end terms (○- -○).

was no overall effect of question markedness, nor did the step size effect differ with the form of the question.

The responses for adjacent pairs only are shown in Figure 3 as a function of the form of the question. The semantic congruity effect is readily apparent. Pairs referring to longer time intervals were compared faster when the question was *Which is longer?*, while decisions about pairs referring to shorter time intervals were faster when the question was *Which is shorter?*, $F(9, 198) = 5.19$, $p < .001$. The size of the interaction is essentially monotonic with the position of the pairs on the scale. The overall differences among the 10 adjacent pairs fell just short of significance, $F(9, 198) = 1.91$, $p = .06$.

Rated distance was a better predictor of decision time ($r^2 = .402$) than either grouping ($r^2 = 0.364$) or simple ordinal distance ($r^2 = 0.346$). In neither case was the R^2 significantly increased by the addition of any variable beyond rated distance, while the effect of rated distance remained significant even after the variance attributable to either of the other variables was accounted for.

The overall error rate for the time scale was 2.2%. Almost all of the errors were made with adjacent pairs (5.0% for the adjacent pairs vs. 1.6% for all other pairs).

The Quality Scale

The overall distance effect for the quality scale is shown in Figure 4, both with and without the end terms *perfect* and *awful*. In both cases, decision time decreased monotonically with increasing step size, $F(3, 66) = 40.2$, $p < .001$ without the end terms, and $F(5, 110) = 44.1$, $p < .001$ including the end terms. Decision time decreased 348 msec from step size 1 to step size 6. Again, the form of the question (*better* or *worse*) did not affect decision time in the distance analysis.

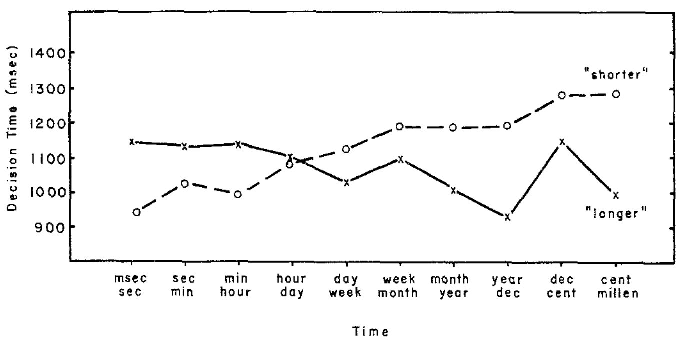

FIG. 3. Decision time as a function of scale position and question markedness for adjacent pairs of time terms (dec = decade, cent = century, millen = millenium).

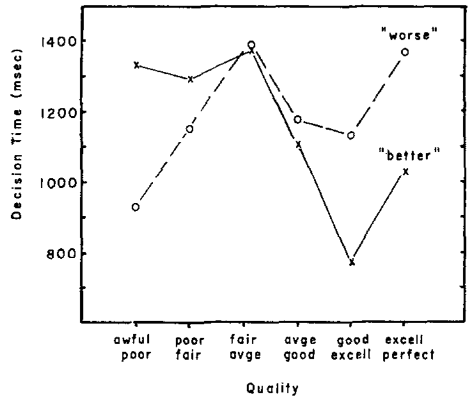

FIG. 5. Decision time as a function of scale position and question markedness for adjacent pairs of quality terms (avge = average, excell = excellent).

The decision times for adjacent pairs are shown in Figure 5 as a function of the form of the question. The congruity effect was highly reliable for the quality scale, $F(5, 110) = 9.55$, $p < .001$. Terms near the positive end of the scale were compared faster for the unmarked comparative *better*, while terms toward the negative end were faster for the marked comparative *worse*. The size of the congruity effect was a monotonic function of scale position, with almost no difference between the two forms of the question for the "neutral" pair *fair–average*.

Overall decision times for the six adjacent pairs differed significantly, $F(5, 110) = 8.85$, $p < .001$. The main source of the effect appears to be the very slow comparison for the pair *fair–average*. People had considerable difficulty determining that *average* is a more positive term than *fair*. This result is predicted by the analogue model, which assumes that decision time varies inversely with subjective distance. The scaling solution for quality (Figure 1b) shows *fair* and *average* to be very close in subjective magnitude. One would also predict on the basis of subjective distance that the pair *excellent–perfect* would be relatively difficult as well, and this prediction receives some support, as Figure 5 indicates.

The effectiveness of subjective distance as a predictor of decision time was also supported by the multiple regression results. As with the time scale, rated distance was a better predictor of decision time ($r^2 = 0.694$) than either ordinal distance ($r^2 = 0.623$) or grouping ($r^2 = 0.625$). Addition of either ordinal distance or grouping to the rated distance variable did not significantly improve the R^2, while rated distance had a significant effect on decision time independent of the other two variables.

The overall error rate was low (4.9%) with virtually all of the errors occurring on adjacent pairs (14.2%) and very few (1.2%) for all other distances. Among the adjacent pairs, most of the errors were made on the pair *average–fair*.

The Temperature Scale

Decision time as a function of step size is shown in Figure 6 both including and excluding end terms. Decision time decreased significantly as distance increased, $F(2, 44) = 50.4$, $p < .001$ without the end terms, and $F(4, 88) = 8.72$, $p < .001$ including the end terms. The source of this overall distance effect will be discussed in detail below. The form of the question (*warmer* or *colder*) did not affect decision time in the distance analysis.

Mean decision times for adjacent pairs are shown separately in Figure 7 for the two forms of the question. As with the other scales, the congruity effect was highly significant, $F(4, 88) = 14.3$, $p < .001$, and the size of the interaction was monotonic with ordinal position of the pair. The pair *torrid–hot* was

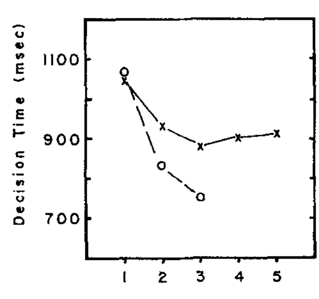

FIG. 6. Decision time as a function of step size for the temperature terms, including the end terms (×—×), and excluding the end terms (○- -○).

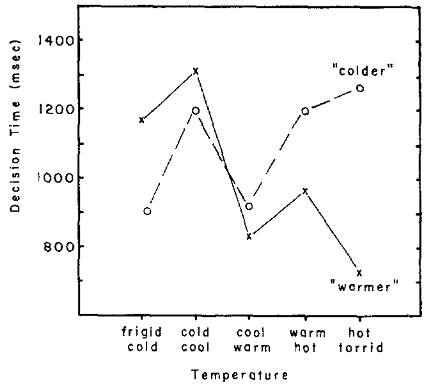

FIG. 7. Decision time as a function of scale position and question markedness for adjacent pairs of temperature terms.

compared 535 msec faster when the comparative was *warmer*, while the pair *cold–frigid* was compared 271 msec faster when the comparative was *colder*. The two forms of the question produced nearly equal decision times for the middle pair *cool–warm*.

Collapsing across the two forms of the question, the five adjacent pairs differed significantly in decision times, $F(4, 88) = 10.9$, $p < .001$. Unlike the results from the previous scales, comparisons with the temperature scale appeared to follow a categorical effect. Intuitively, the six terms can be divided into two groups: the three "hot" terms (*torrid*, *hot*, *warm*) and the three "cold" terms (*cool*, *cold*, *frigid*). Considering all possible comparisons, decision times were slow when the terms were from the same group and uniformly fast when the terms were from different groups, regardless of step size. The effect of step size in Figure 6 is entirely due to the relatively slow RTs for the pairs of distance one and two for which both items are from the same group.

These results thus initially appear to support the categorical grouping model. However, less than half of the subjects in the grouping task separated the terms into the two groups proposed above. Most of the other subjects produced three groups: a "hot" group (*torrid*, *hot*), a "medium" group (*warm*, *cool*), and a "cold" group (*cold*, *frigid*). The distance model fares better: the scaled distances (Figure 1c) show the widest spacing between *cool* and *warm*, correctly predicting that *warm–cool* should be the fastest adjacent pair. Over all the pairs, the rated distance of the terms was a much better predictor ($r^2 = 0.427$) of the decision times than either the grouping measure ($r^2 = 0.273$) or ordinal distance ($r^2 = 0.184$). The addition of grouping to rated distance did not significantly increase R^2. Ordinal distance had a significant partial correlation with RT after rated distance was accounted for, but in the opposite direction (i.e., larger ordinal distances were associated with longer RTs). This residual effect was apparently due to the extra difficulty of the less familiar end terms, which appear relatively often in pairs with large ordinal distances.

Again the overall error rate was low (3.1%), with 7.5% errors for the adjacent pairs and only 0.9% for pairs at all other step sizes. However, no errors were made on the pair *warm–cool*.

DISCUSSION

The present results firmly establish the existence of both the symbolic distance effect and the semantic congruity effect for natural semantic scales. For pairs from all three scales, decision time for comparative magnitude judgments decreased with increasing ordinal distance of the terms in the pair. In addition, more detailed analyses revealed that subjects' ratings of the subjective distance between terms predicted decision time even better than did ordinal distance. For pairs of terms adjacent in the ordering, the difficulty of the pair being compared depended on the congruence of the form of the question (unmarked or marked) and the position of the terms on the scale (unmarked end or marked end).

Evaluation of the Three Comparison Models

These results permit an evaluation of the possible models of the comparison process that were discussed earlier: the relational model, the grouping model, and the analogue distance model. The relational model, which predicted that decision times for comparisons of adjacent pairs would sometimes be fastest, was not supported by the data. Comparisons between adjacent terms were the most difficult, as measured by both decision time and errors, even when the adjacent concepts were relationally defined (e.g., *week* and *day*). It is clear, of course, that the relational definitions of concepts are often explicit (such as the fact that a week is composed of 7 days) and that people can readily retrieve them. However, such relational information apparently is not the basis for the magnitude comparison process. It therefore appears that the semantic representations of concepts such as the time terms include not only discrete relational information but also readily accessible information about absolute magnitude.

The present results also allow a comparison of the categorical grouping model and the analogue model as explanations of the distance effect. The strong version of the grouping hypothesis predicted that the distance effect would be discrete—comparisons which crossed a fixed category boundary should be uniformly fast and comparisons within a category should be uniformly slow. However, we found no evidence that people divide the semantic scales into firm categories that can be used to predict the distance effect. The distance effect appeared to be discrete only with the temperature scale and, in that case, the grouping suggested by the decision time data did not correspond to a unique grouping actually provided by normative subjects.

The weaker version of the grouping model avoids the incorrect prediction that the distance effect will be discrete, by assuming that the category boundaries will vary from subject to subject or from trial to trial. However, this model still suggests that the best predictor of comparison time will be the probability with which people place each pair of items in the same group. In contrast, the analogue model suggests that ratings of the subjective distance between each pair of items will predict comparison time more effectively. For all three scales, the rated distance for a pair was a better predictor of decision time than was the probability of the pair being grouped together, while split-half correlations showed that the two measures were equally reliable. In no case did adding the grouping variable to rated distance significantly improve the decision time prediction. The analogue model, which assumes that ordered concepts contain absolute magnitude information as part of their meanings, thus appears to provide the most successful account of the distance effect in the present study.

The present results, however, certainly do not rule out all categorical models of the distance effect. One could presumably formulate a variant of the grouping model to account for the obtained distance effect, by assuming that semantic scales are divided into variable nested and overlapping groups, including "groups" with a single member. However, such a model would seem to have lost much of the empirical force of the original strong version of the grouping hypothesis.

Semantic Congruence and the Comparison Process

The present results extend the principle of semantic congruence, previously reported for artificial orderings and perceptual comparisons (Banks et al., 1975; Clark, 1969; Trabasso & Riley, 1975) to comparisons of semantically ordered concepts. There appear to be several possible approaches to incorporating the congruence effect into a comparison model, all of which are broadly compatible with the available data. Three models can be differentiated with respect to their stands on three basic issues—the content of the retrieved information used to reach a decision, the relationship of the retrieved material to the question, and the type of process used to compare the retrieved information about each term. On another basic issue, the content of the permanent semantic representation, all three models agree in assuming that continuous magnitude information is included in the permanent representation of ordered terms. Definitive evaluations of these models are not possible given the present data.

Banks, Fujii, and Kayra-Stuart (in press) have proposed a model of digit comparisons that could readily be extended to semantic scales. The model postulates two independent stages. In the first stage, a discrete linguistic code (either LARGE or SMALL) is retrieved for each of the digits on the basis of continuous magnitude information in the digits' representations. The code retrieved for any term is derived probabilistically, such that the relative probability of the two codings varies with the subjective magnitude of the term. That is, smaller digits are more likely to be coded SMALL, while larger digits are increasingly likely to be coded LARGE. If the codes for the two terms mismatch, the correct term can be chosen immediately. The probability of a mismatch increases with increasing distance between the two terms, yielding a higher proportion of fast stage one responses for widely separated pairs of digits. This accounts for the distance effects. If the two codes match initially, a second stage is required in which one of the terms must first be recoded. For example, if the initial codes are LARGE and LARGE, one of the codes must be converted to LARGE_+ (i.e., "larger"). The codes LARGE and LARGE_+ are sufficient to answer the question "Which is larger?" However, if the question is "Which is smaller?", the semantic codes must then be converted to SMALL_+ and SMALL so as to be congruent with the question. The extra time required for recoding and conversion processes is the source of the congruity effect.

Although the permanent memory representation of the magnitude of the digits is assumed to be continuous, the Banks et al. model assumes that the codes actually used in the comparison process are discrete (in fact, binary) categorical labels. While the decision time results obtained in the present study are compatible with this model, the grouping results make us hesitant to accept the model as a description of the data. The model bases its predictions on the probabilities with which categorical labels are applied to terms on the scale. The probability with which two terms were placed in the same category in the grouping task would therefore be expected to predict decision time more accurately than would the ratings of the distance between the terms. This prediction was not confirmed by the regression analyses reported above. In addition, Jamieson and Petrusic (1975) report data concerning a congruence effect with mental size comparisons that is also inconsistent with the Banks et al. model.

A second possible model assumes that analogue magnitude information is not only stored, but is retrieved for use in the comparison process. This model assumes that the information is retrieved in a form congruent with the question. Congruent retrieval is made possible by the fact that each term is coded in two ways, one being a measure of extent on the unmarked dimension (e.g., "longness") and the other a measure of extent on the marked dimension (e.g., "shortness"). The key assumption of this second model is that the relative availability of these two measures differs with the position of the term on the scale. That is, for terms referring to long time intervals, "longness" values are more accessible, while for terms referring to short time intervals, "shortness" values are more easily retrieved. The form of the question determines which set of values is retrieved and compared to reach a decision. (This is in contrast to the Banks et al. model, in which the values retrieved are determined by the magnitude of the terms, not by the form of the question.) Decision time for the comparative *shorter* will therefore increase as the "shortness" values become less accessible for terms towards the longer end of the scale, while decision time for *longer* will increase as the "longness" values become less accessible for terms towards the shorter end of the scale. The distance effect arises in the comparison of the retrieved values according to some general comparison mechanism, such as the statistical sampling procedure of Luce and Green (1972).

Both of the above models invoke separate processing stages to account for the congruity and distance effects. A third possible model (related to those of Marks, 1972, and Jamieson and Petrusic, 1975) attempts to explain both effects in terms of discriminability. A major assumption of this model is that scales in memory, unlike some physical scales, are conceptually bounded at both ends. For example, the physical temperature scale has an absolute zero but no upper limit. However, the assumption of the model being considered is that the conceptual temperature scale has two end points, which are approximated by the concepts *frigid* and *torrid*. By this assumption the extremes of the scale are at a finite (and reasonably small) distance from any point on the scale. In this model a decision about relative magnitude is based on the difference between the two terms relative to one of the end points. The end point chosen for the comparison is determined by the question. For example, the instruction to choose the colder of two terms will cause the subject to compare the distances of the two terms from the cold end point, while the instruction to choose the warmer term will cause him to compare distances from the hot end point. It is a well established psychophysical principle that differences at low stimulus magnitudes are more discriminable than equal differences at high stimulus magnitudes. By analogy, it should therefore be the case that if two terms are close to the extreme specified by the question, the magnitude difference between them will be more discriminable than if they are both far from the appropriate end. "Cold" terms, for example, will be more discriminable than "warm" terms when compared to the "cold" end point (as will occur when the question is "Which is colder?") while "hot" terms will be more discriminable than "cold" terms when compared to the "hot" end point (with the question "Which is warmer?"). While the relation between decision time and discriminability can be formulated in different ways (Jamieson & Petrusic, 1975; Marks, 1972), models of this type share the basic assumption that both distance and congruity effects are due to the discriminability of the magnitude difference between terms.

Implications for Semantic Theory

Unlike the ordered concepts examined in the present study, terms from most semantic categories are not explicitly ordered (e.g., types of animals). Most work in semantic memory has been concerned with these unordered "taxonomic" categories (Leech, 1974). The differences between unordered concepts have most often been expressed by discrete meaning contrasts, as "male" and "female," for example, contrast on gender. Such semantic markers make it possible to represent the fact that in most cases different concepts from the same semantic category are incompatible in meaning. For example, a human cannot be both a bachelor and a spinster, and an animal cannot be both a dog and a cat (Glass & Holyoak, 1975; Holyoak & Glass, 1975; Katz, 1972; Leech, 1974). However, this concern with discrete meaning components has resulted in a relative neglect of continuous properties such as extent or duration in discussions of semantic descriptions (Walker, 1975).

Further, since semantic theories have most commonly dealt with discrete rather than analogue properties, it has been suggested that analogue properties are fundamentally incompatible with the formalism of semantic theories (e.g., Paivio, 1975). It does not seem necessary to accept this extreme view. Some semantic theorists have used other forms of meaning components to deal with problem cases. For example, Katz (1972) has introduced essentially analogue semantic markers in order to account for the definitions of verb tenses. An important feature of the language is that some semantic categories are ordered and others are not. A complete semantic theory must be able to explain the properties of both types of knowledge. The inclusion of analogue magnitude information in semantic representations may make it possible to extend semantic theories to describe nondiscrete properties and order information.

REFERENCES

BANKS, W. P., CLARK, H. H., & LUCY, P. D. The locus of the semantic congruity effect in comparative judgments. *Journal of Experimental Psychology: Human Perception and Performance*, 1975, **104**, 35–47.

BANKS, W. P., FUJII, M., & KAYRA-STUART, F. Semantic congruity effects in comparative judgments of digit pairs. *Journal of Experimental Psychology: Human Perception and Performance*, in press.

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

CLARK, H. H. Linguistic processes in deductive reasoning. *Psychological Review*, 1969, **76**, 387–404.

GLASS, A. L., & HOLYOAK, K. J. Alternative conceptions of semantic memory. *Cognition*, 1975, **30** 313–339.

HOLYOAK, K. J. Symbolic processes in mental comparisons. Unpublished Ph.D. dissertation, Stanford University, 1976.

HOLYOAK, K. J., & GLASS, A. L. The role of contradictions and counterexamples in the rejection of false sentences. *Journal of Verbal Learning and Verbal Behavior*, 1975, **14**, 215–239.

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

KATZ, J. J. *Semantic theory*. New York: Harper & Row, 1972.

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

LEECH, G. *Semantics*. Harmondsworth: Penguin Books, 1974.

LOVELACE, E. A., & SNODGRASS, R. D. Decision times for alphabetic order of letter pairs. *Journal of Experimental Psychology*, 1971, **88**, 258–264.

LUCE, R. D., & GREEN, D. M. A neural timing theory for response times and the psychophysics of intensity. *Psychological Review*, 1972, **79**, 14–57.

MARKS, D. F. Relative judgment: A phenomenon and a theory. *Perception & Psychophysics*, 1972, **11**, 156–160.

MOYER, R. S. Comparing objects in memory: Evidence suggesting an internal psychophysics. *Perception & Psychophysics*, 1973, **13**, 180–184.

MOYER, R. S., & BAYER, R. H. Mental comparison and the symbolic distance effect. *Cognitive Psychology*, in press.

MOYER, R. S., & LANDAUER, T. K. Time required for judgments of numerical inequality. *Nature (London)*, 1967, **215**, 1519–1520.

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

PARKMAN, J. M. Temporal aspects of digit and letter inequality judgments. *Journal of Experimental Psychology*, 1971, **91**, 191–205.

POTTS, G. R. Information processing strategies used in the encoding of linear orderings. *Journal of Verbal Learning and Verbal Behavior*, 1972, **11**, 727–740.

POTTS, G. R. Storing and retrieving information about ordered relationships. *Journal of Experimental Psychology*, 1974, **103**, 431–439.

SEKULER, R., RUBIN, E., & ARMSTRONG, R. Processing numerical information: A choice time analysis. *Journal of Experimental Psychology*, 1971, **89**, 75–80.

SHEPARD, R. N. The analysis of proximities: Multidimensional scaling with an unknown distance function, I & II. *Psychometrika*, 1962, **27**, 125–140, 219–246.

TRABASSO, T., & RILEY, C. A. On the construction and use of representations involving linear order. In R. L. Solso (Ed.), *Information processing and cognition: The Loyola Symposium*. Hillsdale, New Jersey: Lawrence Erlbaum Associates, 1975.

WALKER, J. H. Real-world variability, reasonableness judgments, and memory representations for concepts. *Journal of Verbal Learning and Verbal Behavior*, 1975, **14**, 241–252.

(Received September 26, 1975)