

Comparative Judgments with Numerical Reference Points

KEITH J. HOLYOAK

University of Michigan

A model of subjective magnitude comparisons is explored, which assumes that subjects compare symbolic stimulus magnitudes with respect to a reference point. The reference point may be established implicitly by the question (e.g., "Which is larger?" vs "Which is smaller?") or be presented explicitly (e.g., "Choose the stimulus closer to X."). The model was tested in five experiments in which subjects judged which of two comparison digits was closer to (or further from) a reference digit. Regression analyses in three experiments revealed that reaction time depended on the ratio of the distances from the comparison items to the reference point. The other two experiments provided evidence that subjects can strategically vary the processes by which they compare stimuli to a reference point. The results indicated that subjects can perform various types of "analog arithmetic" using either the linear number scale or a nonlinear scale of subjective digit magnitude.

A major property of the human memory system is the ability to store and manipulate magnitude information. Many physical dimensions (e.g., size, distance, and duration) are perceptually continuous, and we can code the dimensional values of stimuli into memory. In addition, many conceptual dimensions are also subjectively continuous. These include complex attributes such as pleasantness, and such abstract but fundamental concepts as number. Natural languages incorporate a variety of devices for expressing information about continua, including adverbial quantifiers (e.g., "moderately," "shortly before"), ordered semantic categories (e.g., terms for intervals of time), and aspects of the tense system for verbs. A full understanding of cognitive processing must of necessity include an understanding of the representation of magnitude information in memory.

Despite its importance, the issue of magnitude representation has sometimes been overlooked in investigations of semantic memory.

This study has been aided by discussions between the author and Bill Banks, Gordon Bower, Clyde Coombs, John Jonides, Keith Patterson, Keith Smith, and Janet Walker. I am particularly grateful for the comments of Herbert Clark, as well as for the editorial comments of Zenon Pylyshyn and several anonymous reviewers. I thank Barbara Bushey for her help in testing subjects and John Patterson for programming assistance. Peter Ward of the University of Michigan Statistics Department provided extensive help with data analyses. This research was supported by NSF Grant No. BNS77-01211 and a Rackham Grant to the author; pilot work was supported by NIMH Grant Nos. MH13950-08 to Gordon Bower and 20021 to Herbert Clark. Reprint requests should be sent to the author at University of Michigan, Human Performance Center, 330 Packard Road, Ann Arbor, Michigan 48104.

However, a number of recent studies have begun to investigate the processes by which people use information in memory to evaluate relative magnitude. A typical procedure is to present a subject with a pair of symbols and to measure the subject's reaction time (RT) to choose the term that is greater or lesser in magnitude (e.g., to choose the larger or smaller of two digits). A major phenomenon that has emerged from such pair comparison studies is the "symbolic distance effect"—the greater the psychological difference between members of a pair of symbols stored in memory, 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., 3 and 9) rather than small (e.g., 3 and 4). In addition to other digit studies (Banks, Fujii, & Kayra-Stuart, 1976; Buckley & Gillman, 1974; Parkman, 1971; Sekuler, Rubin, & Armstrong, 1971), distance effects have been obtained with several other continua, including remembered object sizes (Holyoak, 1977; Jamieson & Petrusic, 1975; Moyer, 1973; Paivio, 1975) and semantically ordered terms such as adjectives of quality (e.g., *good*, *fair*) (Holyoak & Walker, 1976).

Moyer and Landauer (1967) were the first to point out that the symbolic distance effect has a parallel in perceptual comparison tasks. For example, when a subject is asked to choose the longer of two lines, decision time decreases as the magnitude of the length difference is increased (Woodworth & Schlosberg, 1954, p. 33). Further, in both symbolic and perceptual comparisons the distance effect follows a psychophysical function: A given difference between two stimuli evokes a quicker reaction the smaller their absolute values. For example, decision time for digit pairs is faster for low pairs (e.g., 1–2) than for high pairs (e.g., 7–8), suggesting that the subjective scale is a logarithmic or similar compression function of numerical magnitude (Banks et al., 1976). Using a variety of different dimensions, Griggs and Shea (1977), Holyoak and Walker (1976), and Moyer and Bayer (1976) have provided evidence that RT depends on at least interval properties of subjective magnitude differences, rather than just ordinal properties.

A second major phenomenon that has emerged from the magnitude comparison task involves the form of the comparative 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 information to be evaluated. A kind of congruity effect is found in the magnitude comparison task. For example, subjects may be asked to choose the larger or else the smaller of two digits. Banks et al. (1976) have shown that subjects can choose the larger of two relatively "large" digits (e.g., 8 and 9) more quickly than they can choose the smaller; conversely, subjects can choose the smaller of two relatively "small"

digits (e.g., 1 and 2) more quickly than they can choose the larger. In general, a comparison can be made more quickly if the items are drawn from near the end of the scale consistent with the form of the question.

Congruity effects have been found with a variety of symbolic comparisons. In addition to digit comparisons these include RT to make color preference judgments (i.e., decisions about which of two colors is most or least preferred) (Shipley, Coffin, & Hadsell, 1945; Shipley, Norris, & Roberts, 1946), mental size comparisons (Jamieson & Petrusic, 1975), and comparisons of terms from semantic orderings (Holyoak & Walker, 1976). Similar congruity effects have been found with perceptual judgments. The latter include judgments of relative pitch (Wallis & Audley, 1964) and of brightness (Audley & Wallis, 1964); e.g., people can choose the higher of two "high" tones more quickly than they can choose the lower, while they can choose the lower of two "low" tones more quickly than they can choose the higher. In addition, similar effects have been found with more complex comparisons such as judgments of the relative age of pairs of children vs adults (Ellis, 1972) and of the relative skin color (lighter or darker) of pairs of whites or blacks (Friend, 1973). In some experiments the congruity effect takes the form of a "crossover" interaction, so that the relative difficulty of the two questions is actually reversed at opposite ends of the scale. Other times a weaker "funnel effect" is found, so that one question is easier overall, but its advantage is reduced at the end of the scale consistent with the other question.

Models of the Comparison Process

There is no dearth of proposed models of relative judgments (for critical reviews of early proposals, see Banks, Clark, & Lucy, 1975; Banks et al., 1976; Moyer & Bayer, 1976). Many cannot account for all aspects of the empirical phenomena outlined above. For example, Parkman's (1971) "counting model," which assumes that on each trial the subject implicitly counts up to the smaller digit, cannot explain the residual effect of magnitude difference nor the congruity effect (Banks et al., 1976). Other proposals are not easy to generalize beyond specific task domains. Moyer and Bayer (1976), for example, explain the congruity effect by assuming that the item set forms a linear "mental array" and that on each trial the subject searches inward to the test items from the end of the ordering congruent with the question (e.g., the search would begin at the "large" end when the comparative is *larger*). But it is not clear how such a search process could operate when the test items are drawn from an "infinite set" (e.g., a size comparison task in which items are not repeated, so that the test items do not form an explicit linear ordering). Yet congruity effects are also found in experiments that use an infinite set procedure (Banks & Flora, 1977).

A particularly intriguing notion has been the idea that symbolic comparisons involve an "internal psychophysics" (Moyer & Landauer, 1967). The central assumption is that symbolic magnitude is coded in memory in a continuous or analog form, and that analog magnitude values are compared directly in the relative judgment task. The parallels between symbolic and perceptual comparisons are taken as evidence that memory and perception to some extent involve common mental processes. But while such a conclusion would be of great importance to our conception of cognition, analog comparison models have proved difficult to formulate in a satisfactory way. One possibility is that symbolic stimuli are represented as mental images, which can be perceived internally in a manner analogous to the perception of physical objects. However, it seems counterintuitive to suppose that our conception of abstract dimensions such as number depends on explicit imagery. Other theorists have maintained that while judgments of relative magnitude involve an analog comparison process, the representational format is abstract rather than imaginal. This position has been formulated most precisely, as a random walk comparison process, by Buckley and Gillman (1974). Their particular model, however, is unable to account for congruity effects (Banks et al., 1976).

The congruity effect in general poses serious problems for interpretations of magnitude comparisons. For example, it would be useful to be able to interpret comparison RT as a measure of the psychological distance between stimuli; i.e., a pair of symbols that are compared relatively slowly would be considered psychologically close (Shepard, Kilpatrick, & Cunningham, 1975). But if the pattern of RTs across pairs changes with the form of the question, such a direct interpretation of RT as a distance measure becomes questionable (Jamieson & Petrusic, 1975). The congruity effect is especially problematic for analog comparison models, because at face value it is clearly a linguistic effect, suggesting that mental magnitude information may in fact always be processed in a linguistic form.

The semantic coding model. Indeed, the congruity effect has been used as the central evidence for a linguistic coding model of the comparison process. This model has been developed by Clark (1969) and Banks et al. (1975) and has been formulated most explicitly by Banks et al. (1976) for digit comparisons. Although the permanent memory representation of the magnitude of the digits is assumed to be continuous, the semantic coding model assumes that the codes actually used in the comparison process are discrete (in fact, binary) categorical tags. The model postulates three serial stages, of which the last two are probabilistic. In the first stage a discrete linguistic code (either LARGE or SMALL) is generated for each of the digits. The code generated for any item is derived probabilistically, such that the relative probability of the two codes varies with the subjective magnitude of the term. That is, the 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 items mismatch, the correct term can be chosen immediately. The probability of a mismatch increases with the distance between the two items, yielding a higher proportion of fast stage one responses for widely separated pairs of digits. This accounts for the distance effect. If the two codes match initially, a second stage is required in which one of the terms is 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 were "Which is smaller?" a third stage would be required to convert the codes to SMALL+ and SMALL so as to be congruent with the question. The extra time required for this conversion process is the source of the congruity effect.

The semantic coding model provides the most explicit account of the relative judgment process currently available. It provides a clear procedure for mental comparisons and has been formulated in a manner permitting quantitative evaluation. Banks et al. (1976) demonstrated that the semantic coding model could predict RT for digit comparisons more accurately than any of the previously proposed quantitative comparison models. The model can readily be extended to nonnumerical dimensions as well.

Reference point models. It is possible, however, to incorporate the congruity phenomenon into an analog comparison model. Such a model assumes that a person can judge the relative magnitude of two symbols by comparing the analog magnitude values stored with each. The comparison process is viewed as continuous; i.e., information retrieved at one point in the process does not differ qualitatively from information available at a different point. However, it is assumed that the precision of the information increases with time. Initial information will be sufficient to distinguish between concepts very different in magnitude (e.g., the digits 3 and 9), but more precise information (which takes longer to generate) will be necessary to distinguish between concepts that are relatively similar in magnitude (e.g., 3 and 4).

An additional assumption must be added to account for the congruity phenomenon. This assumption is related to the unfolding theory developed for preferential choice behavior (Coombs, 1950) and involves the notion of an "ideal point" or "reference point" specified by the question. Reference point models assume that scales in memory are conceptually bounded at both ends, and that a decision about relative magnitude is based on the difference between the two items *relative to one of the endpoints*. The endpoint chosen for the comparison is determined by the question. For example, the instruction to choose the smaller of two digits will cause the subject to compare the distances of the two digits to the lower bound (presumably set somewhere below 1), while the instruction to choose the larger digit will cause him to compare distances from the upper bound

(presumably set above 9 for the digit set 1–9). If differences at low stimulus magnitudes are more discriminable than equal differences at high stimulus magnitudes, then the magnitude difference between two terms will be more discriminable if they are close to the extreme specified by the question than if they are both far from the appropriate end. “Large” digits will therefore be more discriminable when compared to the upper reference point (as will occur when the question is “Which is larger?”), while “small” digits will be more discriminable when compared to the lower reference point (with the question “Which is smaller?”).

The idea that discriminability can be influenced by the implicit reference point of the comparison is not new. Work in social psychology has demonstrated that a person is better able to discriminate between attitudes if they are relatively similar to his own (Hovland & Sheriff, 1952; Kelley, Hovland, Schwartz, & Abelson, 1955; Segall, 1959). In the context of reaction time measurement, reference point models have been formulated in different ways. Marks (1972) assumed that distance from the reference point influenced RT via the “discriminal dispersion” of the subjective stimulus magnitudes (Thurstone, 1927). Jamieson and Petrusic (1975) present an alternative formulation, a “discrepancy ratio” model (or “distance ratio,” as it will be referred to below), that was originally considered by Greenberg (1963) in a preferential choice situation. It is assumed that subjects compare two stimuli by comparing the distance of S_1 from the reference point, $d[S_1, RP]$, with the distance of S_2 from the reference point, $d[S_2, RP]$. It is assumed that RT decreases monotonically as the ratio $d[S_1, RP]/d[S_2, RP]$ departs from unity. However, no version of a reference point model has been tested quantitatively.

Aims of the Present Paper

A major purpose of the present paper is to investigate the role of reference points in a symbolic comparison task that is somewhat more complex than the usual pair comparison procedure. The task to be considered is directly suggested by the central claim of reference point models, namely, that internal magnitudes are compared with respect to an implicit reference point specified by the question. If so, the model should predict the pattern of results in studies in which an *explicit* reference point is given. The experiments reported here involve *triplets* of digits, where the subject’s task is to decide which of two digits is closer in magnitude to the third (the reference point). This procedure bears some resemblance to both the standard relative judgment task with pairs of stimuli and to studies of mental arithmetic (Groen & Parkman, 1972; Parkman & Groen, 1971; Restle, 1970). We shall see that the triplet task in fact appears to share some mental processes with both of these other tasks involving number manipulations. For this reason triplet comparisons can provide a rich new source of evidence for evaluating comparison models.

Digits were chosen as stimuli for several reasons. Because number representation has received a fair amount of study, it is possible to find converging evidence for assumptions that must be made about the subjective representation of numerical magnitude. Number stimuli are particularly amenable to quantitative analyses, and the semantic coding model has been formulated most precisely for digit comparisons. Also, since people know the rules of arithmetic, they can make accurate judgments about pairs of digits that straddle the reference point (e.g., 4 and 8 for the reference point 5). Finally, the cognitive representation of number concepts is of considerable intrinsic interest. In particular, detailed studies of numerical comparisons may provide some insight into the nature of the processes people use to perform "analog arithmetic" (Restle, 1970).

The plan of this paper is as follows. A reference point model is outlined, along with various predictions for the triplet comparison tasks. Five experiments are then reported. These test specific predictions of the reference point model and also provide evidence relating to alternative strategies for making numerical comparisons.

Outline of a Reference Point Model

The reference point model to be explored in the present study can be described at several levels. These include (1) a functional description of the predicted overall RT pattern, (2) a series of psychological processing stages, and (3) an underlying mechanism for the comparison stage. As we shall see, it is by no means the case that all aspects of performance in the triplet task are clearly determined by the data to be reported here. In particular, Experiments 3 and 4 will suggest that subjects can vary their processing strategies under certain conditions. Nevertheless, a preliminary sketch of the general model being considered will make it easier to understand the implications of the data to be reported below.

Functional description: The distance ratio. A reference point model makes two basic predictions. (1) RT will increase monotonically with the distance of the stimuli being compared from the functional reference point. (2) RT will decrease monotonically with the difference between the stimuli with respect to the reference point. In the present study we will see that this RT pattern can be described in a simple way: RT for a given triplet (S_1 , S_2 , RP), where S_1 and S_2 are stimuli to be compared and RP is the reference point, is a linear function of the distance ratio, expressed as follows: $RT = \alpha[(d[S_C, RP]) / (d[S_F, RP])] + K$; i.e., RT is a linear function of the ratio of the closer to the further distance, plus a constant. This formulation of the distance ratio, which resembles a Weber fraction, has several desirable properties. The ratio ranges between asymptotes of 0 and 1 (it is undefined only in the degenerate case where the two stimuli and the RP have the same value). Accordingly, the constant K can be interpreted as the

total time required to encode the digits, reach a decision in a minimal amount of time, and make a response. Theoretically, RT will equal K when the pair of stimuli are maximally discriminable (ratio = 0). If the pair of stimuli are equidistant from the RT, and hence entirely indiscriminable, the ratio will be 1. The parameter α is therefore interpretable both as a proportionality constant and as the maximum time the subject will spend on the decision process before initiating a response (by guessing if necessary). The ratio is sensitive to both the distance from RP and the difference between the distances. For example, if $d[S_C, RP] = 2$ and $d[S_F, RP] = 3$, the ratio will be .67. Increasing the distances to 3 and 4 (which holds the difference constant) will increase the ratio to .75, while increasing the difference, by choosing distances 2 and 4, will decrease the ratio to .50. Since the general claim is that difficulty of a comparison will increase as the distance ratio approaches unity, the reference point model predicts a positive correlation between RT and error rate.

An important question concerns the appropriate scale of measurement for defining the distances between stimuli and RP. The obvious candidate is the linear number scale. However, this assumption leads to difficulties. In the case of simple pair comparisons, this would predict that the congruity effect should be symmetrical for *smaller* vs *larger*. But it is not. The data of Banks et al. (1976) show an RT advantage for low digit pairs (e.g., 1–2) vs high digit pairs (e.g., 8–9) even when the comparative is *larger*. The advantage of the low pairs is simply magnified when the comparative is *smaller*. However, the distance ratio can describe such an asymmetry in the congruity effect if the ratio is computed using subjective magnitude differences, where subjective magnitude is a monotonic decelerating function of numerical magnitude. That is, the model must assume that higher digits are psychologically closer together than are low digits. This is by no means an ad hoc assumption. In fact, studies using a variety of techniques other than RT measurement have provided evidence that subjective numerical magnitude increases in a less than linear fashion (Banks & Hill, 1974; Rule, 1969; Rule & Curtis, 1973; Shepard et al., 1975), although there has been some dispute as to whether the subjective magnitude function is logarithmic or a power function. The model presented here will assume that subjective digit magnitudes can be approximated by a logarithmic transformation of the linear scale.¹

The present model therefore assumes that in the case of pair comparisons, subjects are comparing the subjective distances from the

¹Power functions were also tested in the context of the distance ratio. In addition, a number of alternative formulations of the reference point notion, including possible elaborations of Marks' (1972) model, were also tested. Since none of the alternatives appeared to be as consistently successful in describing the obtained results as was the distance ratio with a logarithmic subjective magnitude function, only the latter will be presented here.

items to an implicit RP.² But what about the triplet paradigm? Consider the case where the subject must decide whether 2 or 3 is closer to 5. What distances will the subject compare to make this decision? One possibility is an implicit mental subtraction process, yielding distance values of 3 and 2, respectively. Alternatively, the subject may be able to directly estimate the “subjective distance” from each digit to the RP. That is, he may be able to tap the same subjective magnitude information apparently used in the pair comparison procedure. In fact, the distance ratio can be used to contrast these two possibilities. If subjects compare distances derived by subtraction, the ratio calculated using linear distances should be the better predictor of RT. But if subjects compare distances derived from subjective magnitudes, the ratio calculated using log values should be the better predictor.

Furthermore, it is possible to generate predictions about the “optimal” scale of measurement. It seems reasonable to assume that manipulation of subjective magnitude values is a more elementary, and hence faster, process than subtraction. For this reason, subjects tend to compare subjective magnitude information *whenever such information is sufficient to generate a correct response*. Subjective magnitude information will *not* produce accurate responses in mental arithmetic tasks; e.g., the correct answer to $2 + 3 = ?$ depends on linear relations among the digits. This is consistent with Parkman and Groen’s (1971) evidence that subjects in fact use a linear counting process to do mental arithmetic. What should be expected for triplet comparisons? First, note that triplets may be either *unilateral* (e.g., 2 and 3 for RP = 5) or *bilateral* (e.g., 3 and 6 for RP = 5). For unilateral pairs, it is always the case that the stimulus with the lesser subjective distance from the RP is the closer (e.g., on a log scale 3 is still closer to 5 than 2 is). But the same is not true for bilateral pairs. For example, on a log scale 8 is actually closer to 5 than 3 is. That is, for bilateral pairs it is sometimes the case that subjective magnitude differences cannot predict the correct answer to the question “Which is numerically closer?”

This analysis leads to the following hypothesis: Subjects will evaluate

²To test the reasonableness of this assumption, the distance ratio with a logarithmic magnitude function was applied to a set of data for *larger* and *smaller* judgments with digit pairs, published by Banks et al. (1976). The combined data for Experiments 1 and 2 of Banks et al. were fitted by the equation $RT = \alpha((d[S_C, RP]/d[S_F, RP])) + K$, where S_C and S_F were respectively the natural logarithms of the digits that were closer to and further from the implicit RP. The RP was set below $\ln 1$ (0) for the *smaller* question and above $\ln 9$ (2.20) for the *larger* question. A nonlinear parameter estimation procedure was used to find the best fitting RPs for the two comparatives (on a natural log scale these were -1.17 for *smaller* and 2.88 for *larger*). This four-parameter equation (the two RPs, α , and K) accounted for 68% of the variance in the RTs obtained for a total of 54 cases. This overall fit was close to that provided by the four-parameter semantic coding model, and as good or better than that provided by two other alternatives (Moyer & Landauer, 1973; Parkman, 1971) that were tested against the same data by Banks et al. (1976).

subjective distances for unilateral triplets but will perform a linear “mental subtraction” process for bilateral pairs. RT for unilateral pairs will therefore depend on the distance ratio computed on a log scale, while RT for bilateral pairs will depend on the distance ratio computed on a linear scale. Since the linear process is assumed to be relatively slow, it also follows that for triplets with equal linear differences with respect to the RP (e.g., the pairs 3–4 vs 4–7 for $RP = 5$), RT will be slower for bilateral than for unilateral pairs. It is interesting to note that such a difference would have a parallel in preference judgments. Subjects produce more intransitivities in their preference orderings of stimuli if the stimuli lie on opposite sides of the “ideal” or maximally preferred stimulus (when stimuli are ordered along some unidimensional scale) than when the stimuli lie on the same side (Coombs, 1958). There are probably various factors that generally contribute to the difficulty of bilateral cases, but presumably it is only with numbers that this difficulty could be reflected in the scale of distance measurement.

It should be noted that an advantage for the log over the linear version of the distance ratio for unilateral pairs will not always be detectable, since for some subsets of digit triplets the two measures will be almost perfectly correlated. But in some specific cases the two measures will make substantially different predictions. For example, the distance ratios for the pairs 1–2 vs 8–9 for $RP = 5$ are identical on a linear scale (.75); but on a log scale, the ratio will be higher for the larger pair (.57 vs .80).

Processing stages. The distance ratio is simply a convenient way to describe the predicted RT pattern; however, the model as described above clearly involves a series of psychological processing stages. Besides the usual encoding and response stages, the present model suggests three stages in the decision process: (1) identifying whether a triplet is unilateral or bilateral, (2) assessing the distance from each comparison digit to the RP, and (3) comparing the two derived distances.

What variables might affect these three stages? The first stage is necessary if subjects actually use different distance scales for unilateral vs bilateral triplets. Presumably, it should be easier to identify on which side of the RP a digit falls if it is relatively far from the RP. However, such an effect will tend to be obscured by the opposite effect of distance from RP predicted for later stages (see below). The clearest prediction is that overall RT will decrease if an RP is used with only unilateral or only bilateral pairs, since then this initial stage should be eliminated.

The second stage, distance computation, is hypothesized to differ for unilateral vs bilateral triplets. If the distances are derived by a subtraction process for the bilateral cases, it is reasonable to suppose that this process would be similar to the mental addition processes studied by Parkman and Groen (1971). Mental addition appears to be based on a rapid counting process, or at any rate on some kind of process that requires time directly

proportional to the numerical magnitude of the smaller of the two numbers being added (Groen & Parkman, 1972). In the triplet task, the linear differences could be generated by a count from the RP to each of the digits being compared (or vice versa). Assuming the two distances are computed serially, the time for the generation of the two distances would be proportional to the sum of the two RP-to-digit numerical distances. Notice that in contrast to the Parkman and Groen model for mental addition, a counting process is only one component of the hypothesized decision strategy for bilateral triplets. The Parkman and Groen model does not include a comparison stage of the sort described below, nor does it refer to a *ratio* of linear quantities.

If *subjective* distances are generated for unilateral triplets, a linear subtraction process will presumably not be used. Accordingly, the sum of the two numerical distances should not be related to distance generation time for unilateral triplets. If generation time is nonetheless proportional to the magnitude of the distances generated, the sum of the distances on a logarithmic scale may predict generation time. Alternatively, it is possible that for unilateral cases generation time is actually independent of the distance from the RP to the comparison digits.

The third stage of the decision process involves a comparison of the two derived RP-to-digit distances in order to select the smaller distance, which corresponds to the digit closer to the RP. It is assumed that this comparison stage is based on a mechanism sensitive to the magnitude difference between the two derived distances. The subject will perform multiple internal observations of the relevant magnitudes in order to estimate their true values with sufficient precision to achieve an acceptable average degree of accuracy in his responses. One possibility is a random walk process (Buckley & Gillman, 1974). Such a process could be formulated in terms of either the difference or the ratio of the two distances. In the latter case, RT would decrease in a symmetrical fashion as the ratio of the two distances departs in either direction from unity. One way the random walk process could operate is as follows. An accumulator would be set to 0 ($\log 1$) at the beginning of each trial. After each successive internal observation of the ratio d_1/d_2 (where d_1 and d_2 might be the distances to the RP from the left and right comparison digits, respectively), a value proportional to the estimated (positive or negative) quantity $\log d_1 - \log d_2$ would be added to the accumulator. The accumulator would be incremented by additional observations until it reaches either a positive criterion, c , or a negative criterion, $-c$. The subject will then respond with the left or the right digit. If the ratio is relatively close to 1, the process will on average take longer to reach the correct criterion; in addition, it will have a higher probability of reaching the other criterion first, triggering an error.

While the present experiments will not test the nature of the comparison process in detail, it is useful to remember that such a random walk process

can account for a variety of important factors that are likely to influence the decision process. If the internal variability of the distances is relatively high (for example, because the magnitude of the reference point is especially variable), the probability of reaching the wrong criterion will also be relatively high. Accordingly, the subject may shift his decision boundaries further from 0 to maintain a given accuracy level. Such a criteria shift would produce variations in the rate of change in RT with the distance ratio across different reference points. Instructions that stress speed over accuracy will cause the subject to set his criteria closer to 0, producing faster RTs but higher error rates. If response probabilities in the experiment are unequal, the subject may set criteria that are not symmetrical around 0. Presumably, the subject has available a rational process for making rapid comparisons under conditions of uncertainty.

Evaluating assumptions of the model. It should be clear that while the processing stages outlined above are conceptually distinct, it is difficult and perhaps impossible to empirically tease apart and test all of the assumptions involved. For example, the pair-to-RP distances (at least for bilateral pairs) may affect both the distance generation stage and the comparison stage. Recall that the distance ratio measure is sensitive to both the sum of the distances *and* the difference between them. As a result, it is difficult to determine whether the distance-to-RP has its effect at the generation stage, the comparison stage, or both. Furthermore, it is not necessarily the case that these two stages are strictly sequential. It is possible that each internal observation in a random walk process requires both generation and comparison of the two distances, so that both stages are carried out repeatedly and perhaps partially in parallel on a given trial. Similarly, it is difficult to determine precisely how processing differs for unilateral vs bilateral triplets. One possibility is that the entire difference lies in the output from the generation stage, and that the comparison stage for the two types of triplets is identical. Alternatively, the comparison stage may also differ (e.g., subjects may assess the difference between the two distances for bilateral cases, the ratio for unilateral cases). Again, these possibilities are difficult to distinguish because the distance ratio may reflect factors operating in both the generation and comparison stages.

These considerations determined the way in which the data to be reported below were analyzed. Rather than attempting to fit the data with regression equations that include all the factors that could potentially influence each processing stage (cf. Sternberg, 1977), several types of analyses were carried out. First, the major ordinal predictions of a reference point model (an increase in RT with distance from the comparison digits to the RP, a decrease in RT with the difference between the two distances) were tested by analyses of variance. Second, the data were fitted by relatively simple regression equations relating RT to the distance ratio. These analyses test the possible different scales of

measurement used for unilateral vs bilateral cases in the generation and comparison stages. Finally, additional regression analyses were used to provide at least some evidence concerning the details of the processes by which subjects generate and compare the distances between digits.

EXPERIMENT 1

The reference point model predicts that asking subjects to choose the digit closer to 1 vs 9 will produce "congruity" effects similar to those obtained with the questions "Which is smaller?" vs "Which is larger?" (Banks et al., 1976). Further, a new "congruity" effect should be found for the reference point 5—pairs close to the center of the series should be relatively fast, while extreme pairs should be relatively slow (i.e., the plot of RT vs serial position should be V shaped). A possible extension of the semantic coding model that might predict such a pattern will be discussed below. However, a coding model that assumed all comparisons are made using the binary codes SMALL and LARGE would fail to produce any answer for judgments of closeness to 5 (since knowing that a digit is SMALL or LARGE provides no information about its distance from 5). This problem would also arise for any model that explains congruity effects by invoking prestored binary magnitude codes (e.g., the "multiple subscales" model mentioned by Holyoak and Walker, 1976).

Method

Subjects were timed as they decided which of two digits (referred to as "the pair") was numerically closer to a third (the reference point). All possible triplets were formed using reference points 1, 5, and 9 with the following exceptions: (1) "Pattern matches" were not included; i.e., the same digit never appeared as both the reference point and as a member of the pair. To balance the stimulus sets for RPs 1 and 9, neither 1 nor 9 was used in the pairs for either RP. (2) Pairs equidistant from 5 (e.g., 4–6) were not used with 5 as RP. All pairs were presented in both possible orders (e.g., 1–2 and 2–1), producing a total item set of 132 triplets (42 for RPs 1 and 9, 48 for RP = 5).

The stimuli were presented on TV screens connected to an IBM 1800 computer. Up to three subjects were tested simultaneously in semiisolated booths. The sequence of events on each trial was as follows. The word READY appeared in the center of the screen for 1000 msec and was then replaced by a plus sign. This fixation cross stayed on for 500 msec and then was replaced by the digit triplet. The left and right digits in the pair appeared one character space to the left and right, respectively, of the position previously occupied by the fixation cross. The reference point digit appeared one character space above the position of the fixation cross. The triplet remained on until the subject pressed one of two response buttons, up to a maximum of 4 sec. Subjects pressed the left button if the left digit was closer to the RP, and they pressed the right button if the right digit was closer to the RP. Reaction time and correctness were recorded automatically on each trial. If 4 sec elapsed without a response, an error RT of 4 sec was recorded. If the subject made an error, the triplet offset was followed by display of the word ERROR for 1000 msec. Upon completion of these events, the screen remained blank until all subjects being tested had completed the current trial. The next trial then began automatically.

The session began with presentation of 20 practice triplets, using RPs other than 1, 5, and 9.

The test session involved five cycles through the test items. Presentation order was randomized individually for each subject on each cycle. The program paused after the first and third cycles to allow rest breaks. The cumulative error percentage was displayed on the screen at these breaks to provide feedback to the subjects. Subjects were told to respond as quickly as possible but to avoid errors. The entire session lasted about 50 min.

Eight University of Michigan undergraduates served as paid subjects.

Results and Discussion

Preliminaries. Several general aspects of the data need to be described initially, since these were common to all the experiments to be reported here and determined the form in which the results will be presented. First, there was a monotonic decrease in both RT and error rate across the five trial blocks (from a mean RT of 1018 msec on block 1 to 963 msec on block 5). However, the basic pattern of results across the 66 triplets showed no apparent systematic change with practice. The correlation between mean RTs for the items on blocks 1–2 vs blocks 3–5 was $r = .90$. Accordingly, all detailed analyses were performed on mean RTs obtained after collapsing over blocks. Second, order of the pairs (closer digit on the left or right) had no apparent consistent effect on the results. In Experiment 1 mean RT was 970 msec for left responses vs 992 msec for right responses, and the correlation between item means for left and right responses was $r = .87$. All reported analyses were consequently performed after collapsing across the two presentation orders. Third, subjects almost never exceeded the 4-sec deadline (this occurred on just a single trial). In fact, only .4% of the responses exceeded 2500 msec. This value was taken as a cutoff point for wild scores, and RTs above 2500 msec were truncated to 2500 msec. Finally, overall error rate was fairly low (5.2%), and error RTs appeared to be too infrequent to merit separate analyses. As a practical matter, including error RTs in computing mean RTs did not change the overall pattern (the correlation between mean item RTs calculated with and without error RTs was $r = .997$). As a result, all analyses will be reported on mean RTs that include errors. This basic pattern held in all the experiments, and essentially the same procedures will be followed in reporting the later results.

Qualitative analyses. Figure 1 presents the mean RTs for the adjacent pairs used with the three reference points. Considering just the results for RPs 1 and 9, two clear commonalities with *smaller* and *larger* judgments (Banks et al., 1976) are apparent. First, the adjacent pairs differ significantly in RT, $F(5,35) = 4.95$, $p < .01$, and most of this variance is due to an increase in RT for higher digit pairs. The RT functions for RP = 1 and 9 show a dip at the pair 4–5; interestingly, this pair is also somewhat faster than a log magnitude scale would predict in the data of Banks et al. (1976) and also in data analyzed by Shepard et al. (1975). Second, a clear “congruity” effect was obtained. The advantage of RP = 1 over RP = 9

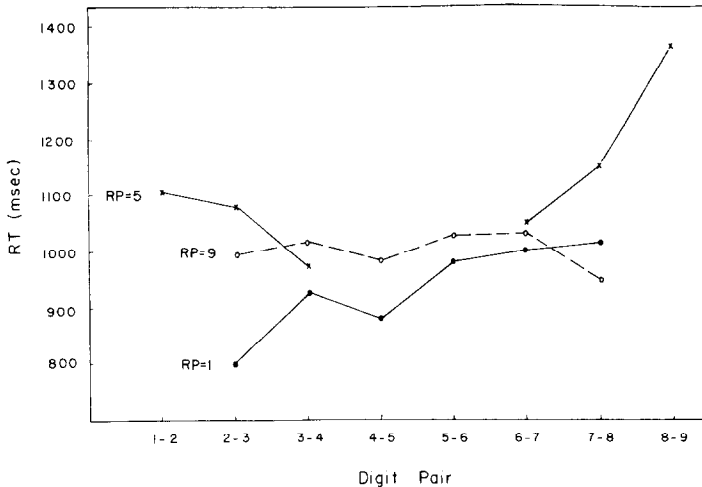


FIG. 1. RT to select the digit closer to 1 (● — ●), 9 (○ - - - ○), or 5 (x — x) for adjacent pairs.

was reduced for the higher pairs, $F(5,35) = 3.35, p < .025$. While the form of this interaction differs somewhat from that obtained by Banks et al., this may in part be explained by assuming that the implicit RPs for *smaller* vs *larger* judgments with digits are not set exactly at 1 and 9 (see footnote 2). However, the present results do show one significant departure from the results obtained with comparative questions. While *larger* produces faster overall RT than *smaller*, RP = 9 produced slower RTs than RP = 1. This discrepancy will be considered below.

The results for RP = 5 show a very different pattern. As predicted, the serial position function is V shaped—RT is relatively fast for the central pairs 3-4 and 6-7 and very much slower for the extreme pairs 1-2 and 8-9, $F(5,35) = 11.3, p < .001$. Furthermore, the function is asymmetric—the pairs above 5 are slower than the pairs below 5. This is what would be expected if RT depended on a log-scale distance ratio. A similar asymmetric V shaped function was obtained for the four pairs with magnitude difference of two (mean RTs of 1018, 911, 1070, and 1188 msec for pairs 1-3, 2-4, 6-8, and 7-9, respectively, $F(3,21) = 35.1, p < .001$). As Fig. 1 shows, pairs with RP = 5 were considerably slower than pairs with RP = 1 or 9.

The "congruity" prediction of the reference point model can also be tested for the bilateral pairs with RP = 5. For example, the pairs 4-7, 3-8, and 2-9 all have the same numerical magnitude difference with respect to 5 (one), but the mean distance of the pairs from 5 increases from 4-7 to 2-9. Consequently, the reference point model predicts that RT should be faster for a central pair like 4-7 than for an extreme pair like 2-9. Figure 2 presents mean RT for bilateral pairs as a function of the distance of the pair

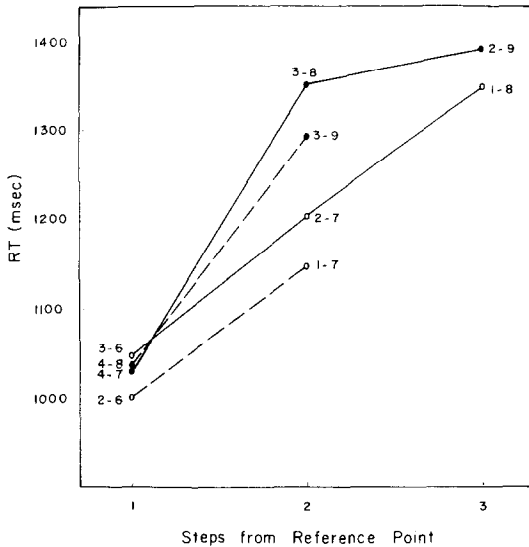


FIG. 2. RT to select the digit closer to 5 as a function of the distance of the pair from 5 (solid line = magnitude difference of one, broken line = magnitude difference of two; filled circle = lower digit closer to 5, open circle = higher digit closer to 5).

from 5. The data are broken down according to whether the lower or higher digit is closer to 5 and whether the magnitude difference with respect to 5 is one or two. The "congruity" prediction of the model is clearly supported by these data. For the pairs with a magnitude difference of one, RT increases 319 msec from the most central to the most extreme pairs, $F(2, 14) = 18.6, p < .01$. The pairs with a difference of two show a 195-msec increase, $F(1,7) = 29.1, p < .01$. These results clearly indicate that the simple difference between the two stimulus-to-RP distances is not sufficient to predict comparison latencies.

Table 1 presents mean RTs as a function of magnitude difference for the various types of triplets. In general, RT declined monotonically as the

TABLE 1
MEAN RT AS A FUNCTION OF MAGNITUDE DIFFERENCE

	Magnitude difference					
	1	2	3	4	5	6
RP = 1	934	876	848	824	798	805
RP = 9	1004	976	902	837	835	769
RP = 5, unilateral	1119	1046	1079	—	—	—
RP = 5, bilateral	1231	1122	1059	—	—	—

difference increases, $F(5,35) = 22.1$, $p < .001$ for $RP = 1$ and 9, and $F(2, 14) = 19.0$, $p < .001$ for $RP = 5$. In addition, the advantage of $RP = 1$ over $RP = 9$ declined with increasing magnitude difference, $F(5, 35) = 3.42$, $p < .025$. An interaction with magnitude difference was also obtained for the unilateral vs bilateral pairs with $RP = 5$. While bilateral pairs were processed more slowly when the magnitude difference was one, the RT difference between the two types of pairs declined with increasing magnitude difference, $F(1,14) = 6.31$, $p < .025$ (for a test of the monotonic trend). As we shall see below, these interactions can be captured in a quantitative model.

The effect of magnitude difference for the bilateral pairs is of particular interest in that it allows a separation of the effect of magnitude difference from the effect of the "split," or simple difference between the two digits in the pair. These two measures are always equivalent for unilateral pairs; but for some bilateral pairs the magnitude difference with respect to 5 *decreases* as the split is increased. RT was clearly determined by magnitude difference in these cases. Thus RT increased for the series 1-6, 1-7, 1-8 (1034, 1151, and 1366 msec, respectively), in which an increase in split produced a decrease in magnitude difference.

An interesting test for a role of subjective magnitude information in decisions about bilateral pairs involves pairs with equal numerical magnitude differences, such as 3-8 vs 2-7, in which either the lower or the higher digit is closer to 5. Note that on a log scale 7 is closer to 5 than 3 is, while 2 is further from 5 than 8 is. If subjects are sensitive to subjective magnitude differences for bilateral pairs, RT should in general be faster for pairs in which the higher digit is closer, given that the numerical difference is equal. The obtained RT differences were in this direction for five of the six possible comparisons; results for the five matched pairs with magnitude differences of one and two are presented in Fig. 2. Collapsing over all cases, pairs in which the higher digit was closer to 5 were evaluated 63 msec more quickly than pairs in which the lower digit was closer, $F(1,7) = 4.47$, $p < .10$. While this difference is rather marginal, it does suggest that subjects may sometimes use subjective magnitude differences to evaluate bilateral pairs. As we will see shortly, such an evaluation strategy is at most a secondary effect for bilateral pairs.

An error analysis produced results quite consistent with the obtained RT differences. Mean error rate was 3.2% for $RP = 1$, 2.0% for $RP = 9$, 4.8% for unilateral pairs with $RP = 5$, and a substantial 14.6% for the bilateral pairs. Across all 66 items, the correlation between RT and error rate was $r = .69$, $p < .01$.

Quantitative fit of the distance ratio. Recall that a question of particular interest concerned the optimal scale of measurement for computing the distance ratio. Table 2 presents the correlations between RT and the distance ratio (closer divided by further) computed on both a log and a

TABLE 2
 INTERCORRELATIONS AMONG RT AND DISTANCE RATIO COMPUTED
 ON A LOG OR LINEAR SCALE

	RT and ratio _{log}	RT and ratio _{linear}	Ratio _{log} and ratio _{linear}
RP = 1	.85	.88	.99
RP = 9	.89	.85	.98
RP = 5, unilateral	.74	.50	.87
RP = 5, bilateral	.67	.84	.55

linear scale, as well as the correlations between the two ratio values. These correlations are computed separately for the four types of triplets. The two ratios are almost perfectly correlated for RP = 1 and 9, and both were excellent predictors of RT, accounting for approximately 75% of the variance in each case. For the unilateral pairs with RP = 5, however, the intercorrelation of the two ratios decreased, and a substantial advantage for the ratio_{log} emerged (a difference of 30% in variance accounted for). The ratio_{log}, of course, is able to account for the asymmetry in the V-shaped serial position curves obtained for RP = 5 (Fig. 1). For the bilateral pairs the two ratios were still less correlated with each other, and the ratio_{linear} accounted for 26% more variance than did the ratio_{log}.

The pattern of predictive power apparent in Table 2 is further supported by the results of regression analyses. When both ratios were used simultaneously to predict RT for all the unilateral pairs, the ratio_{log} was significant beyond the .0001 level, while the ratio_{linear} did not account for significant additional variance ($p = .09$). However, the ratio_{log} was not significant when the ratio_{linear} was entered into the regression equation first. Conversely, simultaneous prediction of RTs for the bilateral pairs produced an effect of the ratio_{linear} beyond the .0001 level, while the ratio_{log} was not significant ($p = .16$). In this case, the ratio_{linear} remained significant even when the ratio_{log} was entered into the regression equation first.

The results thus argue that unilateral pairs are evaluated on the basis of subjective distances from the reference point, while bilateral pairs are evaluated on the basis of linear numerical distances. But in both cases the overall decision process requires time proportional to the distance ratio.

However, while the distance ratio predicted the RTs within RPs very well, the fit deteriorated when the data for all RPs were predicted at once. The problem, as we have already observed, is that pairs with RP = 9 were slower than pairs with RP = 1, while unilateral pairs with RP = 5 were slower yet. The simple regression equation using the distance ratio cannot describe these effects of the specific reference point. There is some evidence that the difficulties with 9 and 5 have slightly different sources.

Holyoak (Note 1) reports an experiment essentially identical to the present one except that a 2-sec delay separated presentation of the RP and the pair. Under those conditions, decisions about which digit was closer to 9 were in fact made about 20 msec *faster* than decisions about which digit was closer to 1. This suggests that 9 may require more time for initial encoding as a reference point. Intuitively, 1 seems to be clearly marked as the lowest digit (perhaps more so than zero—see Shepard et al., 1975), while 9 is not so clearly the uppermost digit. While this kind of effect is not directly reflected in the distance ratio, it is consistent with the kind of stochastic comparison process on which judgments are assumed to be based. In terms of the random walk process described earlier, if the variability of the internal subjective distances is increased, it becomes necessary to adopt a more conservative response criterion in order to maintain a constant level of accuracy. Such a criterion shift would be reflected in an increase in α , the proportionality constant relating RT to change in the distance ratio.

In contrast to pairs with RP = 9, Holyoak (Note 1) found that unilateral pairs with RP = 5 were relatively slow even with a 2-sec delay. The natural explanation is that 5 was used as an RP for bilateral as well as unilateral pairs. The bilateral pairs are particularly difficult, and the evidence suggests that subjects identify bilateral pairs and use a different procedure for evaluating the pair-to-RP distances. This extra decision stage would be expected to slow judgments about both unilateral and bilateral pairs for RP = 5.

These considerations suggested an expanded regression model that could be applied to the entire RT pattern obtained in Experiment 1. Since the slope of the effect of magnitude difference on RT changed from RP = 1 to RP = 9, and from unilateral to bilateral cases for RP = 5 (as reported above), different proportionality parameters relating RT to the ratio were allowed for the four item types. In addition, an extra parameter of constant processing time, C , was allowed for all cases where RP = 5, to account for the extra processing stage required to identify unilateral vs bilateral pairs. The full regression model then becomes:

$$RT = \begin{cases} \alpha_1(\text{ratio}_{\log}) + K & \text{when RP} = 1 \\ \alpha_2(\text{ratio}_{\log}) + K & \text{when RP} = 9 \\ \alpha_3(\text{ratio}_{\log}) + K + C & \text{when RP} = 5 \text{ (unilateral)} \\ \alpha_4(\text{ratio}_{\text{linear}}) + K + C & \text{when RP} = 5 \text{ (bilateral)}. \end{cases}$$

The correlation between obtained and predicted RT for the 66 items was $R = .92, R^2 = .85$. All six parameters were significant beyond the .02 level. The values of the proportionality parameters were respectively 188, 416, 486, and 620 msec for RP = 1, RP = 9, RP = 5 (unilateral), and RP = 5 (bilateral). The estimate of overall constant processing time, K , was 750 msec, and the estimate for the additional constant, C , for RP = 5 was 109 msec. Inspection of the scatter plot for predicted vs obtained RT revealed

no systematic departures from linearity. The distance ratio thus provides a rather good fit to the observed RTs, given the assumptions that subjects can alter their response criteria for different reference points, and that an extra processing stage is added for RPs that have both unilateral and bilateral cases.

EXPERIMENT 2

We noted earlier that a model that assumes magnitude comparisons are based solely on the binary codes of LARGE and SMALL cannot produce a decision for judgments of closeness to the central reference point 5. However, such a model could be revised in a simple way to account for the results of Experiment 1. The model could introduce a new magnitude category, MIDDLE, that is the most probable code for central digits. Selecting the digit closer to 5 can then be translated into an instruction to choose the digit coded MIDDLE+. With this addition, the revised model can accommodate the V-shaped "congruity" effect found for $RP = 5$.

It might be argued that "middle" may well be a natural magnitude category, just like "large" and "small." Clearly the next step, then, must be to determine whether a new form of the congruity effect can be produced for *any* arbitrary reference point. If so, a model based on prestored magnitude codes would be forced to the unparsimonious extreme of postulating a separate "category" for each item. Experiment 2 was designed to provide evidence on this point.

Method

Subjects were again presented with triplets of digits and asked to decide as quickly as possible which of two digits was numerically closer to a third. The stimuli included both unilateral and bilateral pairs. The unilateral items in Experiment 2 were constructed using all possible pairs with magnitude differences of one, two, and three, e.g., the pairs 1–2, 1–3, and 1–4. A subset of all possible reference points from 2 to 8 was included, so that each comparison pair was associated with two or three RPs above and/or below it (e.g., the pair 5–6 was used with the lower RPs of 2, 3, and 4, and the higher RPs of 7 and 8). The resulting item set comprised 66 unilateral items: 28 for magnitude difference = 1, 22 for magnitude difference = 2, and 16 for magnitude difference = 3. In addition, 21 bilateral items were also included in the design, with three triplets representing each of the RPs from 2 to 8. The bilateral triplets were selected to represent a wide range of magnitude differences (1–4). Since all pairs were presented in both possible orders, the total number of triplets was 174.

The apparatus and presentation procedure were identical to that used in Experiment 1. Six University of Michigan undergraduates served as paid subjects. Each subject went through the item set 14 times in five 1-hr experimental sessions (two blocks were completed the first day and three blocks each succeeding day).

Results and Discussion

The first two blocks of trials for each subject were counted as practice, and analyses were performed on mean RTs obtained by collapsing over the

other 12 blocks and over presentation order of the pairs (item means for blocks 1-7 vs 8-14 correlated $r = .96$, and means for the two presentation orders correlated $r = .92$). RTs over 2500 msec (1.3%) were truncated to that value.

Mean RTs for all the unilateral items are presented in Table 3. For all three magnitude differences, mean RT invariably increased as the reference point moved from one to two steps away from the pair, with an average increase of 85 msec, $t(5) = 2.80, p < .05$. On average, moving the RP from two to three steps away had no additional effect on RT (a mean increase of 1 msec). There was some variability in this difference across pairs, however. In particular, in some cases RT declined from the middle to the furthest RP, although the closest RP always produced the most rapid RT. These deviations from the pattern predicted by the reference point

TABLE 3

RT TO CHOOSE THE DIGIT CLOSER TO THE REFERENCE POINT FOR UNILATERAL PAIRS^a

Comparison pair	Steps from Reference Point					
	Lower RP			Higher RP		
	-3	-2	-1	1	2	3
Difference = 1						
1-2				777(3)	877(6)	818(8)
2-3				841(4)	854(6)	878(8)
3-4				796(5)	870(6)	921(8)
4-5		1012(2)	924(3)	791(6)	885(7)	906(8)
5-6	994(2)	1067(3)	909(4)	873(7)	1016(8)	
6-7	1033(2)	973(4)	876(5)			
7-8	1022(2)	1065(4)	865(6)			
8-9	1004(2)	1076(4)	983(7)			
Difference = 2						
1-3				742(4)	816(6)	840(8)
2-4				819(5)	836(6)	868(8)
3-5				788(6)	859(7)	885(8)
4-6		945(2)	929(3)	819(7)	889(8)	
5-7	994(2)	924(3)	837(4)			
6-8	968(2)	912(4)	827(5)			
7-9	1024(2)	1087(4)	887(6)			
Difference = 3						
1-4				690(5)	770(6)	762(8)
2-5				762(6)	812(7)	840(8)
3-6				818(7)	848(8)	
4-7		906(2)	869(3)			
5-8	928(2)	944(3)	881(4)			
6-9	930(2)	979(4)	889(5)			

^a Note: The reference point for each pair is in parentheses.

model follow a rough pattern—a decline in RT is more likely when the pair is most remote from the RP. The largest discrepancies occurred when the RP was 2 (the most distant RP on a log scale) and the pair was relatively high (digits of 5 and above). The only other deviation of any magnitude was obtained when the pair 1–2 was associated with the maximally distant RP = 8. As we shall see, these deviations do not represent large quantitative departures from the predictions of the distance ratio; nevertheless, we will explore this result further in Experiments 3 and 4.

The data in Table 3 show very clear effects of the subjective magnitude of the digits. Reading down the columns, RT tends to increase with magnitude of the comparison pair. The correlation between RT and the sum of the two digits in each pair was $r = .70$ for pairs with magnitude difference = 1, $r = .69$ for magnitude difference = 2, and $r = .91$ for magnitude difference = 3. In addition, a very interesting effect is apparent in the italicized RTs in Table 3. These are pairs which can be matched in numerical distance from either higher or lower RPs (e.g., the pair 4–5 is one step below RP = 6, and one step above RP = 3). Note that on a log scale, these pairs are actually further from the lower reference point in all these matched cases. As the ratio_{\log} would therefore predict, RT is in each case longer when the pair is matched with a lower rather than a higher RP, $t(5) = 2.87, p < .05$. This result argues that the rapid RT found when the RP is adjacent to one of the comparison digits is not due simply to subjects noting that fact, since such triplets do not produce uniformly fast RTs. Even when the RP is adjacent, effects of the subjective distance are found.

Table 4 presents mean RT as a function of magnitude difference for both unilateral and bilateral cases. A monotonic decrease in RT with increasing difference was obtained in both cases. Considering numerical differences of one to three, the decline in RT was highly significant, $F(2,10) = 149, p < .001$. The unilateral pairs were faster overall, $F(1,5) = 6.62, p < .05$; but as in Experiment 1, the advantage of the unilateral pairs decreased as the magnitude difference increased, $F(2,10) = 20.3, p < .01$. This again argues that different processes operate in judgments about unilateral vs bilateral pairs.

The error rate in Experiment 3 was again low, and the only substantial difference was between unilateral and bilateral cases (1.1 vs 6.1%).

Quantitative fit of the distance ratio. As in Experiment 1, regression analyses were used to compare distance ratios computed on a log vs linear scale. The results were very clearcut. For the unilateral pairs, RT correlated $r = .82$ with the ratio_{\log} but only $r = .51$ with the $\text{ratio}_{\text{linear}}$, a difference of 42% of the variance. When the variance attributable to the ratio_{\log} was partialled out, the residual partial correlation of the $\text{ratio}_{\text{linear}}$ and RT was actually negative, reflecting the tendency of some pairs to decline in RT with the furthest RP. The ratio_{\log} remained significant even when the $\text{ratio}_{\text{linear}}$ was entered into the regression equation first. This clear

TABLE 4
MEAN RT AS A FUNCTION OF MAGNITUDE DIFFERENCE

	Difference			
	1	2	3	4
Unilateral pairs	925	886	852	—
Bilateral pairs	1172	1084	879	848

advantage of the ratio_{\log} for unilateral cases confirms the various effects of subjective magnitude on RT reported above.

The analysis for bilateral pairs was equally decisive. Mean RT correlated $r = .89$ with the $\text{ratio}_{\text{linear}}$, and only $r = .46$ with the ratio_{\log} , a difference of 58% in variance accounted for. Simultaneous regression of the two measures produced no significant effect of the ratio_{\log} ($p = .67$). In contrast, the $\text{ratio}_{\text{linear}}$ was significant even after the ratio_{\log} was first entered into the regression equation.

Since RP-specific effects were not apparent in Experiment 2, a simple three parameter version of the ratio model was used to predict RTs: $\text{RT} = \alpha_1(\text{ratio}_{\log})(z) + \alpha_2(\text{ratio}_{\text{linear}})(1 - z) + K$, where $z = 1$ for unilateral pairs, 0 for bilateral. The correlation between predicted and obtained RT for all 87 items was $R = .91$, $R^2 = .82$. Parameter estimates were $\alpha_1 = 403$, $\alpha_2 = 804$, and $K = 723$. These estimates are reasonably similar to the estimates obtained for the comparable parameters in Experiment 1.

The results of Experiments 1 and 2 provide very strong evidence that effects of distance from a reference point can be obtained for any digit used as reference point.³ A model that assumes numerical comparisons are based on prestored categorical codes would require a separate category for each digit in order to account for these results. This unparsimonious conclusion can only be avoided by postulating some form of computational process that derives categorical codes. This possibility will be considered below, but first we will return to another issue raised by Experiment 3.

EXPERIMENT 3

While the quantitative fit of the distance ratio was at least as good in Experiment 2 as in Experiment 1, we noted one qualitative discrepancy. For the unilateral items, RT tended to decrease when the RP was maximally distant. While this trend could perhaps be discounted as noise, there is reason to consider it more seriously. So far we have assumed that subjects always choose the closer digit by estimating the distance of each

³A further experiment, using adjacent unilateral pairs and all the bilateral pairs that can be formed for RPs of 4 and 6, produced essentially the same results as Experiment 2 ($R = .86$ for the three-parameter ratio model).

digit to the RP and then comparing the two distances. The reference point model as initially outlined is intended to describe this "direct comparison" strategy. However, an alternative decision strategy is possible for trials on which a unilateral triplet is presented (once the subject has identified the item as unilateral). The subject might (1) decide whether the pair is higher or lower than the RP, and then (2) choose the smaller digit if the pair is above the RP, or the larger digit if the pair is below the RP. Time to complete the first stage of the decision should depend on distance from the RP—the further the two comparison digits are from the RP, the easier it will be to decide whether the digits are higher or lower than the RP. The second stage, selecting the larger or smaller digit, would then be independent of the RP. The overall result of using this "indirect comparison" strategy would therefore be a decrease in RT as the distance between the pair and RP increases. Since this alternative strategy is maximally easy when the direct comparison procedure is maximally difficult (i.e., when the pair and the RP are far apart) subjects in Experiment 2 may have tended to use the indirect strategy for just such items, producing the observed decline in RT for the most distance RPs.

At first glance it might seem that this indirect process would be too time consuming overall to constitute a viable alternative strategy. However, a comparison of the information-processing demands of the two possible strategies suggests this need not be so. The first stage of the direct comparison procedure requires that the subject assess the subjective distance of each stimulus to the RP, while the first stage of the alternative procedure requires only that the subject determine whether the stimuli are larger or smaller than the RP. In statistical sampling terms, less information is required to determine the *direction* of a difference than the *size* of the difference. Furthermore, the direction of the difference between the RP and each digit must be assessed in any case, in order to determine whether the item is unilateral or bilateral. Note, however, that the reference point model suggests that choosing the larger or the smaller digit in the second stage of the process will still involve a comparison of distances to an *implicit* reference point (the reference point associated with either *smaller* or *larger* judgments). It may well be the case, however, that the internal representations of these implicit reference points are less variable than the representation of any arbitrary digit, which in a statistical model would make it relatively easy to assess distances to the *larger* or *smaller* RPs.

There is some evidence that the indirect comparison strategy is in fact sometimes preferred in perceptual judgment tasks. Greenberg (1963) originally suggested both the decision strategies discussed above in the context of an experiment in which subjects chose the darker, lighter, or more neutral of two shades of gray; he interpreted his results as support for the indirect strategy. A related result is reported by Krantz (1967), who found that subjects produced relatively high error rates in judging which of

two colors was more similar to a standard when one of the comparison stimuli was actually identical to the standard.

Why did subjects in Experiments 1 and 2 appear to generally use the direct comparison procedure? One reason may be that the alternative strategy is only viable for unilateral pairs. For bilateral pairs it breaks down completely (since the fact that one digit is lower than the RP and one is higher yields no information about which is closer). Since unilateral and bilateral cases were randomly intermixed in both the experiments, subjects may have tended to simply use the direct comparison procedure for all cases. Experiment 2, which produced some evidence that the alternative strategy was used occasionally, contained a rather low proportion of bilateral cases (only 24% of the items were bilateral). This suggests an experimental test of the analysis just outlined. If subjects are presented with only unilateral pairs, and informed about the alternative decision strategy, RT may then actually decrease with distance from the RP.

Method

To form the unilateral pairs, each pair of adjacent digits was paired with each possible reference point digit (except that digits 1 and 9 were not included, nor any triplets in which a comparison item was identical to the reference point). No bilateral cases were included. All pairs appeared in both orders, yielding 60 triplets in all. The apparatus and procedure were essentially the same as in Experiments 1 and 2. Six University of Michigan undergraduates served as paid subjects. Each subject went through the item set 10 times in a 1-hr session, with rest breaks after the third and seventh blocks. The order of each block of trials was individually randomized for each subject. Subjects were informed that the digit pair would always be either lower or higher than the RP. They were told that a good procedure for deciding which digit was closer was to choose the smaller digit if the pair was above the RP, and the larger digit if the pair was below the RP. The session began with presentation of 20 unilateral practice items with nonadjacent pairs.

Results and Discussion

Table 5 presents mean RT as a function of distance from the RP, collapsing over items, blocks, and subjects. An earlier experiment (cited in

TABLE 5
MEAN RT FOR UNILATERAL PAIRS AS A FUNCTION OF DISTANCE
FROM THE REFERENCE POINT

	Steps from reference point				
	1	2	3	4	5
Experiment 3 (no bilateral cases)	845	861	842	834	810
Earlier study (see text) that also included bilateral cases	673	752	789	784	777

footnote 3) had tested exactly the same unilateral pairs, but with bilateral cases intermixed. For purposes of comparison, the data from this other study (for the unilateral cases only) are also presented in Table 5. (Since the apparatus and subject population differed between the earlier study and Experiment 3, the overall RT differences between the two experiments are difficult to interpret).

The two experiments produced markedly different results. Whereas in the earlier experiment RT increased with distance from the RP (as in Experiments 1 and 2), in Experiment 3 it decreased (with a slight residual advantage when the RP was just one step from the pair). The linear trend in Experiment 3, while relatively small, was highly significant, $F(1,20) = 11.4, p < .01$. The residual effect of pair-to-RP distance was nonsignificant, $F(3,20) = 1.89, p > .20$.

While it therefore appears that subjects in Experiment 3 tended to use the indirect comparison strategy rather than the direct comparison procedure, some individual differences in performance were evident when the consistency of the subjects in their use of comparison strategies was examined. As a measure of consistency, the mean RTs for each of the five steps from the RP, reported in the top line of Table 5, were calculated separately for each subject and then intercorrelated. For five of the six subjects all of the intercorrelations were positive (ranging from $r = .28$ to $r = .94$, with a mean of $r = .63$). All of these subjects showed some trend toward faster RTs for pairs far from the RP, indicating use of the indirect comparison strategy. The means for the remaining subject, however, were negatively correlated with those of all the other subjects (a mean correlation of $r = -.49$). This subject showed a clear *increase* in RT with distance from the RP, indicating use of the direct comparison strategy. These individual differences suggest that while experimental conditions may favor one decision process over another, subjects may differ in their preferred comparison procedure.

Additional evidence suggests that the indirect comparison strategy predominantly used in Experiment 3 also operates on subjective magnitude information. Table 6 presents mean RTs for all triplets broken down by the magnitude of the pair, from 2-3 to 7-8. A very strong linear increase in RT with magnitude was obtained, $F(1,25) = 177, p < .001$ (a 146-msec

TABLE 6
MEAN RT AS A FUNCTION OF MAGNITUDE OF THE PAIR

Pair					
2-3	3-4	4-5	5-6	6-7	7-8
760	811	819	885	889	906

increase in RT from 2–3 to 7–8). Just as in the previous experiments, the results of Experiment 3 indicate that the higher digit pairs are subjectively closer in magnitude. The overall error rate in Experiment 3 was 3.7%, and errors and RT were positively correlated, $r = .38, p < .05$.

EXPERIMENT 4

In all of the preceding experiments, subjects judged which of two digits was *closer* to a reference digit. A natural question is how the decision strategies used for *closer* judgments would be altered if subjects were asked to decide which digit is *further* from a reference digit. This issue is important both in developing an understanding of possible comparison strategies and in evaluating the role of linguistic codes in the judgment process.

The use of the negative preposition *from* in the question “Which digit is further from the RP?” suggests that this question may be more complex than the corresponding *closer* question. At least three comparison strategies might be used to answer *further* questions, and each predicts a different RT pattern. (1) Subjects could choose the digit closer to the RP by the direct comparison procedure and then make the opposite response. This would produce an increase in RT with distance from the RP for both questions, with the *further* question adding a constant increment of processing time. (2) For unilateral pairs, subjects could choose the digit closer to the RP by the indirect comparison strategy and then make the opposite response. This would produce a decrease in RT with distance from the RP for both questions, but the *further* question would again add a constant increment of processing time. (3) For unilateral pairs, subjects could answer *further* questions by using a variant of the indirect strategy. This would involve (a) deciding whether the pair is higher or lower than the RP and then (b) choosing the smaller digit if the pair is below the RP, or the larger digit if the pair is above the RP. This strategy has the advantage of producing the correct response for *further* without first requiring a decision about which digit is the closer. This strategy would also produce a decrease in RT with distance from the RP for the *further* question, but since it is not linked to the *closer* judgments, the functions for the two questions need not be parallel. Note that this third strategy has the same advantages for unilateral pairs as the similar indirect strategy does for the *closer* question; and, moreover, it eliminates the need to compute the closer digit and then switch the response. There is therefore reason to expect that for unilateral pairs subjects will prefer the third strategy for answering *further* questions.

A categorical coding model can also be developed for *further* judgments; moreover, such a model might also account for the results of Experiments 1 and 2 without the unparsimonious feature of requiring categorical codes for each digit. Like the reference point model, a categorical coding model

could postulate a computational process that assesses the distance between each comparison digit and the reference point. However, the output of each computation could be a binary categorical code, CLOSE vs FAR. The probability that the computed code is FAR will increase with distance of the digit from the RP. The model then predicts an effect of the comparative in the triplet paradigm that is entirely analogous to that found in the pair comparison task. Since the question in all the experiments reported above was “Which digit is *closer*?”, this version of the coding model predicts the obtained result—RT was relatively fast when the pair was close to the RP. But the model also predicts that this pattern critically depends on the question. If the question were “Which is *further*?”, the model predicts that pairs far from the RP would be evaluated relatively quickly. It is possible that close pairs could have a residual advantage, particularly if the time required for the initial evaluation of stimulus-to-RP distance increases with the magnitude of the distance. But the advantage of the close pairs should at least be attenuated if the *further* question is used.

This formulation of a categorical coding model is very similar to an analog comparison model. Not only does it assume that continuous magnitude information is stored in memory (as does the Banks et al., 1976, model), but it also postulates a kind of “analog arithmetic” process that operates on this information. Indeed, the only conceptual distinction between this categorical model and the reference point model concerns whether the final step of the comparison process involves categorical codes.

Also, note that the interaction between question and distance from RP, while predicted by the categorical model, could also result from the third processing strategy for *further* described above. That is, if subjects use an indirect comparison strategy more often for *further* than for *closer*, a similar interaction would result. Experiment 4 was designed to provide evidence on how subjects answer *closer* and *further* questions for unilateral pairs.

Method

The same adjacent unilateral pairs used in Experiment 3 were also used in Experiment 4. Again, no bilateral cases were included. Both questions were used with each triplet, producing a total of 120 test items. The apparatus was the same as in the previous experiments, and the procedure was similar. Each trial began with presentation of the word CLOSER or FURTHER centered on the TV screen. After 1500 msec this was replaced by the triplet of digits, one centered above the other two, and the subject was required to press the button corresponding to the lower digit that was closer to or else further from the upper digit. In other respects, the procedure was the same as in the previous experiments.

Six University of Michigan undergraduates served as paid subjects. Each completed eight blocks of trials over two 45-min sessions. During each session subjects were given rest breaks after each trial. All items were individually randomized for each subject and each block, with *closer* and *further* questions intermixed.

Results and Discussion

The first trial was considered practice, and analyses were based on the final seven blocks. RTs over 2500 msec (1.2%) were truncated to that value. Table 7 presents the mean RTs and error rates for the two questions as a function of distance from the RP.

Overall, *further* judgments required 74 msec more processing time than did *closer* questions, $F(1,5) = 12.02, p < .025$. However, the two questions differed significantly in the effect of distance from the RP, $F(4,20) = 4.70, p < .01$. For *further*, RT declined monotonically (a total of 132 msec) as the reference digit moved from one to five steps away from the pair. For *closer*, on the other hand, distance from RP had relatively little effect on RT (a maximum difference of 57 msec), and the function was nonmonotonic, with the longest RT obtained when the RP was three steps from the pair. For all 30 basic triplets, the correlation between RTs for *closer* and *further* was only .53.

These results clearly disconfirm the first two possible strategies for making *further* judgments outlined earlier. Subjects certainly did not choose the closer digit and then switch their response. The other two possibilities are more viable. First, subjects may have consistently used the indirect comparison strategy for *further*, producing a decline in RT with distance from RP, and a mixture of direct and indirect strategies for *closer*, producing a less consistent overall effect. Or alternatively, subjects may have made their judgments in accord with the categorical coding model, using the binary codes CLOSE and FAR.

While the evidence is not conclusive, additional analyses lend some support to the multiple strategies hypothesis. First, the nonmonotonic function for *closer* is problematic for the categorical model. While admittedly post hoc, this nonmonotonicity might be explained by assuming subjects used a mixture of processing strategies. Also, unlike previous experiments, the error rate data do not entirely conform to the RT pattern.

TABLE 7
 MEAN RT AND ERROR RATE FOR UNILATERAL PAIRS AS A FUNCTION OF DISTANCE FROM THE REFERENCE POINT (EXPERIMENT 4)

		Steps from RP				
	Question	1	2	3	4	5
RT	Closer	1194	1238	1252	1205	1200
	Further	1358	1336	1313	1265	1186
Percentage errors	Closer	2.74	1.64	1.79	.89	.00
	Further	2.86	3.87	3.17	2.08	1.19

Even though the error rates were very low (1.4 and 2.6% for *closer* and *further*, respectively), they show a reliable pattern: Errors decline with distance from RP, $F(1,20) = 11.68$, $p < .01$ (by a linear trend test); and furthermore, this trend does not differ for the two questions, $F < 1$. While none of the hypotheses considered predicts this discrepancy between RTs and errors for *closer*, this finding perhaps suggests that subjects did not use a single strategy in making *closer* judgments.

More direct evidence for strategic variations was obtained when individual differences were examined. As in Experiment 3, mean RTs as a function of distance from RP were calculated for each subject and intercorrelated (separately for the two questions). For *further*, the pattern was very consistent across subjects. All 15 intercorrelations among the data for the six subjects were positive, ranging from $r = .27$ to $r = .93$, with a mean of $r = .72$. All subjects thus showed some degree of decrease in RT with increased distance from the RP, as the group data indicates. But for *closer*, the subjects showed extreme variations in their RT patterns. The intercorrelations among subjects ranged from $-.52$ through $.01$ to $.85$, with a mean of just $.14$. Examining each subject's data for *closer* individually, RT generally increased with distance from RP for one subject, showed almost no effect for another, and was substantially nonmonotonic for the other four. This apparent inconsistency even at the individual level may have been due to changes in strategies across trials, but the amount of data available per trial did not appear to warrant more detailed analysis.

EXPERIMENT 5

The results of Experiment 4 indicate that for unilateral items, RT shows a more pronounced decline with distance from the reference point for *further* than for *closer* questions.⁴ While the data from individual subjects suggested that this interaction resulted from a mixture of processing strategies for *closer*, the group data were reasonably consistent with a computational categorical coding model. The multiple strategies that can be used to process unilateral pairs, especially when both questions are used, make it difficult to provide a strong test of alternative models.

Fortunately, the situation is much clearer for bilateral pairs. Recall that the indirect comparison strategies for both *closer* and *further*, which

⁴In a further experiment unilateral pairs were intermixed with bilateral pairs, and subjects again answered both *closer* and *further* questions. An interaction between question and distance from RP was also obtained for unilateral pairs in this experiment. However, the form of the interaction changed: RT *increased* with distance from RP for *closer*, while relatively small differences were obtained for *further*. This result suggests that the inclusion of bilateral cases diminishes use of indirect strategies for both questions, but less so for *further* (for which the alternative direct comparison procedure is more difficult). It appears difficult to account for the influence of bilateral cases on the processing of unilateral cases without allowing for such strategic variations.

appeared to predominate in Experiments 3 and 4, are only applicable to unilateral cases. For bilateral pairs, where the digits straddle the reference point, it is not possible to reduce triplet comparisons to *smaller* and *larger* judgments. Accordingly, the only processing strategy available to subjects should be the direct comparison procedure. For bilateral cases, then, subjects should be left with only one procedure for accurately choosing the further digit: judging which digit is closer by the direct comparison process and then switching their response. Accordingly, the reference point model should be able to account for the RT pattern for both questions simply by assuming an additive latency increment for *further* questions. In contrast, if a categorical coding model is to explain the effects of distance from a reference point, the model must predict a congruity effect for *closer* vs *further*: RT should increase with distance from the reference digit for *closer* and decline with distance from the reference digit for *further* (or at least show a less pronounced increase). Experiment 5 was designed to test these competing hypotheses.

Method

All possible bilateral pairs were generated for RPs = 3, 4, 6, and 7, producing a total of 40 basic triplets. Each pair occurred in both possible orders, and each item was used with both the *closer* and *further* question, yielding a total of 160 test items. The apparatus was the same as in the previous experiments, and the procedure was identical to that of Experiment 4. Ten University of Michigan undergraduates served as paid subjects. Each completed 12 blocks of trials over four 45-min sessions. During each session subjects were given rest breaks after the first and second trial blocks. All items were individually randomized for each subject and each block, with *closer* and *further* questions intermixed.

Results and Discussion

The first two trial blocks were considered practice, and analyses were based on the final 10 blocks. RTs over 3000 msec (.5%) were truncated to that value. The effects of question and distance from the RP were examined in analyses of variance performed separately on the RTs for each magnitude difference (1–3). In order to have balanced designs, the analyses were also performed separately for the extreme RPs, 3 and 7, and for the central RPs, 4 and 6. The mean RTs and error rates are presented in Table 8. These analyses can be briefly summarized. The effect of distance from RP was invariably significant ($p < .001$). Collapsing over the two questions, RT increased in each case as distance from the RP increased. In each case RT was slower (by an overall mean difference of 186 msec) when the question was “Which is further?” as opposed to “Which is closer?” ($p < .001$ in all analyses).

In all six analyses, the mean square error was actually considerably lower for the interaction effect than for the main effect of distance from RP, indicating that the tests for nonadditivity of the effects of question and

TABLE 8
 MEAN RT AND ERROR RATE AS A FUNCTION OF DISTANCE FROM THE PAIR
 TO THE REFERENCE POINT (BILATERAL CASES)

	Question	Magnitude difference	Steps from RP				
			RP = 3 and 7		RP = 4 and 6		
			1	2	1	2	3
RT	Closer	1	1046	1314	1011	1397	1466
		2	930	1252	939	1218	1352
		3	939	1172	940	1119	—
	Further	1	1303	1505	1244	1608	1541
		2	1134	1413	1162	1353	1405
		3	1161	1331	1140	1254	—
Percentage error	Closer	1	5.10	3.96	5.52	9.69	6.04
		2	2.08	2.50	2.29	3.75	3.54
		3	1.04	1.25	1.46	2.50	—
	Further	1	2.81	5.21	4.17	6.56	8.75
		2	2.92	4.38	1.98	3.54	5.42
		3	2.92	3.54	1.46	2.71	—

distance from RP were relatively powerful. Nevertheless, the interaction did not approach significance in any of the four analyses in which the RP was either one or two steps from the pair. The interaction was significant for magnitude differences of one and two, RP = 4 and 6 ($p < .025$). In both cases the RT advantage of *closer* decreased as distance from RP increased. However, examination of the error data in Table 8 reveals that these RT differences were accompanied by speed-accuracy tradeoffs. In both cases, *further* actually produced a lower error rate than *closer* when the RP was one or two steps from the pair and a higher error rate than *closer* when the RP was three steps away. Also, the error rate for *closer* in each case declines from step two to step three. It appears that all the mean RTs in Table 8 that caused departures from the parallel functions predicted by the reference point model were accompanied by compensating changes in error rates. In addition, the correlation between mean *closer* vs *further* RT for all 40 items was very high, $r = .91$ (as contrasted with the much lower correlation of .53 obtained for unilateral cases in Experiment 4). The overall correlation between error rates for the two questions was also substantial, $r = .60$. These results therefore provide considerable support for the predictions of the reference point model and little evidence that a categorical coding model can account for the effects of distance from a reference point.

TABLE 9
MEAN RT AND ERROR RATE AS A FUNCTION OF MAGNITUDE DIFFERENCE

		Magnitude difference			
		1	2	3	4
RT	Closer	1211	1105	1042	988
	Further	1419	1272	1221	1157
Percentage error	Closer	6.33	2.74	1.56	1.67
	Further	5.13	3.37	2.66	2.08

Table 9 presents mean RTs and error rates as a function of magnitude difference (1–4) for the two questions. The apparent decrease in RT as the difference is increased was highly significant, $F(3,27) = 236, p < .001$, as was the advantage of *closer* over *further*, $F(1,9) = 95.1, p < .001$. Although the RT pattern was very similar for both questions, the interaction was significant, $F(3,27) = 4.72, p < .01$. This effect was small in magnitude and almost entirely due to the relatively large increment in RT for *further* when the magnitude difference was one (41 msec larger than the next largest difference). Moreover, examination of the error data reveals that the error rate was actually lower for *further* than *closer* when the magnitude difference was one, so this small degree of nonparallelism in the RT data can again be ascribed to a speed–accuracy tradeoff.

As in Experiments 1 and 2, RT for the bilateral pairs of Experiment 5 (averaged over the two questions) correlated much more highly with the $ratio_{linear}$ than with the $ratio_{log}$ ($r = .87$ vs $.46$, a difference of 55% in variance accounted for). The residual partial correlation of the $ratio_{log}$ again was negligible ($r = .08$) and nonsignificant. Regression analyses were also used to compare the predictive power of the $ratio_{linear}$ with that of a variety of measures that might be related to an extension of Parkman's (1971) counting model. The effect of the $ratio_{linear}$ remained highly significant after the variance due to either the closer distance, the further distance, or the sum of the two distances was partialled out. It therefore appears unlikely that a "pure" counting model could account for comparisons involving bilateral pairs. However, each of the other three distance measures correlated significantly with the residual variance not explained by the $ratio_{linear}$, with the sum achieving the highest partial correlation ($r = .59$). This residual effect of the sum provides additional evidence that for bilateral pairs subjects generate digit-to-RP distances by a subtraction or counting procedure, which depends on the numerical difference.⁵

⁵For unilateral pairs (Experiments 1 and 2 and footnote 3) the residual effect of the numerical sum on RT was inconsistent and generally nonsignificant. Similarly, the residual effect of the sum computed on a log scale was small and sometimes negative. These results

The distance ratio was used to predict RT for all 80 cases (two forms of the question for each triplet) by assuming an additive increment in RT for *further*. The three-parameter model was $RT = \alpha(\text{ratio}_{\text{linear}}) + K + F(z)$, where $z = 1$ for *further*, 0 for *closer*. The correlation between predicted and observed RT was $R = .88$, $R^2 = .78$, with estimated parameters of $\alpha = 994$, $K = 667$, and $F = 186$. As in the previous experiments in which subjects used the direct strategy to make comparative judgments, the distance ratio again provides a very good quantitative description of the obtained RT pattern.

GENERAL DISCUSSION

Summary and Implications

The present results demonstrate that there is a direct relationship between RT in digit comparison tasks and the ratio of the distances of the stimuli to an explicit or implicit reference point. The analysis presented here represents a synthesis of a number of ideas that have emerged in the area of symbolic magnitude comparisons. Most notably, the present proposal incorporates the notions that such comparisons involve an analog process (Moyer & Landauer, 1967) that can be simulated by a random walk (Buckley & Gillman, 1974), that the question has an effect on the comparison process (Banks et al., 1976), that this effect results from a change of reference point (Marks, 1972), and that distances from a reference point may be evaluated as a subjective ratio (Jamieson & Petrusic, 1975). By bringing these ideas together it seems possible to provide a more general treatment of the process of comparing symbolic magnitudes than has been previously offered. In particular, the reference point model can be applied to both judgments of the larger or smaller of two digits and to judgments about which member of a digit pair is closer to or further from a reference digit.

The results of Experiments 3 and 4, however, place important qualifications on the claim that any single model can provide a unified treatment of symbolic comparisons. These results indicate that subjects can be quite flexible in selecting optimal processing strategies to fit particular experimental conditions; any model will have to accommodate such strategic variations. These results are directly relevant to the problem

suggest that for unilateral pairs the effect of distance from the RP to the comparison digits is due to the comparison stage, rather than the generation stage (see the discussion in the Introduction). In contrast, it appears that for bilateral pairs distance from RP affects the generation stage, and perhaps the comparison stage as well. Consistent with the latter possibility is the fact that the proportionality constant relating RT to the distance ratio was always larger for bilateral than for unilateral pairs. Since the ratio measure is influenced by the sum of the distances, a larger α value would be expected if distance from RP were affecting *both* the generation and the comparison stages.

of differentiating strategy-dependent and strategy-invariant memory processes (Anderson & Bower, 1973) in the context of relative judgment. It appears that subjects can perform the same task using alternative sets of mental operations on a single underlying representation of digit magnitudes. The reference point model claims that comparison time has a strategy-invariant relationship to the *functional* reference point; however, subjects can strategically vary the functional reference point, as well as the nature of the decision made about the relation between a stimulus and a reference point (e.g., evaluating the direction vs the size of the difference). Such strategic variations need not render models untestable, since it is possible to specify variables (e.g., instructions, presence or absence of bilateral items, form of the question) that should influence subjects' decision strategies. It is necessary, however, for models of relative judgment to separate strategy-dependent and strategy-invariant components as clearly as possible.

The present study is particularly pertinent to the problem of understanding the processes involved in "analog arithmetic." From this point of view, the most intriguing result that emerged from the application of the reference point model to the digit triplet paradigm was the apparent relationship between the processing of unilateral and bilateral pairs. In both cases RT depended on a distance ratio, but the optimal scale of measurement differed in the two cases. For unilateral pairs subjects apparently access subjective digit magnitudes, which can be approximated by a log transform of the linear scale. But for bilateral pairs such subjective "feelings of magnitude" are poor predictors of the correct response; as a result, subjects appear to use a subtraction or counting procedure, which depends on linear distance. In the context of the reference point model it thus proved possible to separate processes that operate on subjective magnitude values from mental arithmetic processes of the kind explored by Parkman and Groen (1971). More generally, this analysis provides a rather novel test of a stage model. The different measurement scales indicate differences in the processes that assess digit-to-RP distances for the two kinds of triplets (the distance generation stage), while the similar dependencies between RT and the ratio indicate that the actual comparison stage may be fundamentally the same in both cases. Whether this kind of analysis will prove useful in other situations remains to be seen.

The Role of Reference Points

An important conceptual issue concerns the psychological rationale for the use of reference points in comparative judgment tasks. In the triplet paradigm the rationale is clear: The correct answer depends on the relationship of the two comparison digits to an explicit reference point. But why should subjects consider implicit reference points at the ends of the

continuum when making pair comparisons, such as *smaller* and *larger* judgments? In this situation, with reference points at the extremes, extremes, $||S_1 - RP| - |S_2 - RP||$ is algebraically equivalent to the simple difference, $|S_1 - S_2|$. So why should subjects apparently complicate their decision process by comparing stimuli to reference points, rather than directly to each other? For if they directly compared the two stimuli, they could simply select the greater stimulus as the larger and the lesser stimulus as the smaller, producing parallel functions for *larger* and *smaller* judgments and eliminating the congruity effect.

One possibility is that prominent reference stimuli are simply difficult or impossible to ignore, as in certain perceptual tasks (e.g., a circle appears relatively small when surrounded by larger circles). It is also possible that implicit reference points actually facilitate the judgment process for questions involving bipolar adjectives. Bipolar adjectives effectively divide a continuum into two subscales on opposite sides of a neutral midpoint or "secondary reference point" (Clark, Carpenter, & Just, 1973). Note that the comparative morpheme *er* has the meaning of "more" (i.e., "greater"). But *er* can be added to either adjective in a bipolar pair (e.g., *richer* vs *poorer*). The meaning of a term like *poorer* is not precisely equivalent to "less rich"; rather, it means "more poor," i.e., more extreme on the subscale bounded by the neutral point and the negative pole of the continuum (Clark, 1969; Clark et al., 1973). In contrast, the *er* in *richer* implies an increase in magnitude in the direction of the positive pole. Accordingly, the abstract concepts "greater" and "lesser" that presumably underlie the comparative judgment process depend on a specification of directionality for their interpretation. That is, polar adjectives specify a particular end of the scale with respect to which "greater" and "lesser" may be defined. Subjects may therefore make comparisons with respect to the poles of the continuum in order to clarify to themselves how the meaning of the comparative maps onto the abstract concepts of "greater" and "lesser."

This analysis leaves open the possibility that under some circumstances subjects will *not* need to consider implicit polar reference points when making pair comparisons. For example, it is possible that with sufficient practice subjects could make *larger* and *smaller* judgments by directly comparing the two stimulus magnitudes and choosing the greater or else the lesser stimulus (i.e., the stimulus that is more or less large). If so, congruity effects obtained in pair comparison tasks should tend to diminish with extended practice. In contrast, the kind of reference point effects reported in the present paper, which are obtained with interior reference points, should not be eliminated by any amount of practice.

The above rationale for the role of reference points in pair comparisons, which hinges on a semantic analysis of comparatives, points out the fact that analog comparison models and linguistic analyses need not be

incompatible. In addition, some other aspects of the processing of comparatives directly implicate linguistic factors. In Experiment 4 the negative question "Which is further from the RP?" was overall slower than the positive "closer" question; this difference was not directly accounted for by the differential processing strategies that were discussed. More generally, Clark et al. (1973) have reported several experiments relating comparative judgment latencies to differences in the complexity of the procedural rules for applying the comparatives. These linguistic effects appear to fall outside the scope of analog comparison models.

Other Types of Congruity Effects

The reference point model provides a possible explanation for the effects of the background on the form of the congruity effect in perceptual judgments (Audley & Wallis, 1964). In judgments of the *brighter vs darker* of two lights, Audley and Wallis obtained a true "crossover" interaction when the background luminance was intermediate between that of the darker and lighter test stimuli, but a "funnel" interaction (an overall advantage for *brighter* judgments) when the background was darker than all of the test stimuli. The reference point model can explain these effects by assuming that the implicit reference points for *darker* and *lighter* (i.e., the subjects' conception of the perceptual "ideal points" for the two questions) were set near the extremes of brightnesses used in the experiment, *including* the background. In the neutral background condition the reference points would be more or less symmetrical around the test stimuli, producing a crossover interaction. But in the dark background condition the *darker* reference point would be more extreme, so that all test stimuli would be relatively far from it. The *darker* question would therefore be relatively slow overall, producing a funnel effect. (For an alternative explanation of the Audley and Wallis results, see Banks et al., 1975.)

The reference point model is intended to account for cases in which the relative ease of alternative relative judgments (e.g., "Which is larger?" vs "Which is smaller?") depends on the magnitude of the stimuli on the relevant dimension. However, some other kinds of congruity effects reported in the literature likely fall outside the scope of the present model. For example, Clark (1969) showed that subjects given the information "John is better than Bill and Bill is better than Pete" are faster to answer the question "Who is best?" than the question "Who is worst?", while the reverse is true if the initial information is expressed in terms of "worse." In this completely linguistic question-answering task, it seems most likely that a version of the semantic coding model applies.

Congruity effects are also obtained in absolute judgment tasks. For example, Clark and Brownell (1975) found that subjects could determine

that an arrow was pointing up more quickly if it was high in the visual field, while they could determine that an arrow was pointing down more quickly if it was low. Here the interaction is between codes on two related perceptual dimensions. In a somewhat similar paradigm, Banks et al. (1975) found that if subjects were led to interpret ambiguous drawings as "balloons," they could select the higher more quickly than the lower; but if they interpreted the drawings as "yo-yos," the reverse was true. In this case the interaction appears to be between the question and the "standard" directional relation between the type of stimuli (as interpreted) and the position of an observer. Whether a change of reference point is in some way involved in producing this effect is unclear. However, this congruity effect is clearly different from those dealt with in the present paper. In particular, the congruity effect obtained by Banks et al. (1975) did not depend on the magnitude of the stimuli; e.g., it was easier to judge the higher of two "balloons" regardless of their absolute or relative height. It appears most likely that there are a number of different sources for phenomena that have been termed "congruity effects."

In particular, it is useful to distinguish two general kinds of explanations for congruity effects. To take a concrete example, why will people take more time to choose the smaller than the larger of the pair *moose–elephant*? One explanation is that the absolute magnitude code for each of the terms, "large," will interfere with the processing of the term *smaller*. The semantic coding model is an explicit formalization of this idea. Another explanation is that the two objects actually "seem less different" in size when the comparative *smaller* is used. The reference point model embodies this notion. It is quite possible that both of these basic explanations are correct, at least in some circumstances. To find evidence that comparatives can actually influence subjects' "perception" of magnitude differences, it may be useful to investigate judgments that are not made under speed instructions (e.g., direct ratings of the size of magnitude differences).

Further Issues

A question of obvious importance is to what extent the reference point model will prove applicable to nonnumerical dimensions, particularly dimensions on which stimuli vary continuously, rather than in discrete steps (as do the digits). A recent thesis by Baum (Note 2) provides some encouraging evidence for the generality of the model. Baum extended the triplet paradigm used here to comparisons of real-world distances (with well-known campus locations as stimuli). Subjects judged which of two locations was closer to a reference location. As in the present experiments with digits, latencies to compare actual distances from memory were linearly related to the distance ratio.

There are a variety of other interesting problems to which the notion of cognitive reference points may prove relevant. One question concerns the circumstances under which congruity effects will be found with perceptual

judgments. As noted earlier, congruity effects have been obtained for judgments of relative pitch (Wallis & Audley, 1964) and of brightness (Audley & Wallis, 1964). But recently Moyer and Bayer (1976) and Banks (1977) have reported experiments on size comparisons in which congruity effects seem to be found only when the comparison involves memorized stimuli. In these experiments the variation in the size of the perceptual stimulus items was rather small, whereas Audley and Wallis used more extreme differences along the dimensions they studied. A speculative possibility is that perceptual congruity effects will in general be found only when a wide magnitude range is used. If subjects use their background knowledge of the dimension to establish very extreme implicit reference points, then a narrow range of "intermediate" stimuli will appear more or less equidistant from both the reference points. Memorizing the stimuli, however, may draw attention to the range of the actual stimulus set, so that the end items may essentially become the reference points, resulting in a much more pronounced congruity effect. A similar process may produce the "end anchor" effects found in studies of judgments involving order in a linear series taught by the experimenter (Potts, 1974; Trabasso & Riley, 1975; Woocher, Glass, & Holyoak, 1978). Alternatively, on some perceptual dimensions subjects may find it relatively easy to compare stimulus magnitude differences directly, without processing polar reference points at all.

A closely related issue involves the general question of factors that may influence implicit reference points. It is possible that reference points will be influenced to some degree by the range of magnitude values that the subject expects to be represented in the experimental item set. For example, the upper reference point in a numerical comparison task would presumably increase if the item set contained two-digit numbers. The effects of such shifts in the postulated reference point may provide interesting tests of the model. These possibilities are related to work on "adaptation level" phenomena (Helson, 1964), which has demonstrated effects of the magnitude range of the items on the subjective "central tendency" of the presented stimuli. It is possible that analogous effects can be found for the subjective upper and lower bounds of the stimuli. Whatever the eventual outcome, further explorations of the reference point notion will hopefully aid our understanding of how magnitude information is processed in memory and in perception.

REFERENCES

- Anderson, J. R., & Bower, G. *Human Associative Memory*. Washington, D.C.: V. H. Winston, 1973.
- Audley, R. J., & Wallis, C. P. Response instructions and the speed of relative judgments. I. Some experiments on brightness discrimination. *British Journal of Psychology*, 1964, 55, 59-73.
- Banks, W. P. Encoding and processing of symbolic information in comparative judgments. In

- G. H. Bower (ed.), *Psychology of learning and motivation*, Vol. 11. New York: Academic Press, 1977.
- 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., & Flora, J. Semantic and perceptual processes in symbolic comparisons. *Journal of Experimental Psychology: Human Perception and Performance*, 1977, **3**, 278–290.
- 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., & Hill, D. K. The apparent magnitude of number scaled by random production. *Journal of Experimental Psychology (Monograph)*, 1974, **102**, 353–376.
- 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.
- Clark, H. H., & Brownell, H. H. Judging up and down. *Journal of Experimental Psychology: Human Perception and Performance*, 1975, **1**, 339–352.
- Clark, H. H., Carpenter, P. A., & Just, M. A. On the meeting of semantics and perception. In W. G. Chase (Ed.), *Visual Information Processing*. New York: Academic Press, 1973.
- Coombs, C. H. Psychological scaling without a unit of measurement. *Psychological Review*, 1950, **57**, 145–158.
- Coombs, C. H. On the use of inconsistency of preferences in psychological measurement. *Journal of Experimental Psychology*, 1958, **55**, 1–7.
- Ellis, S. H. Interaction of encoding and retrieval in relative age judgments: An extension of the “crossover” effect. *Journal of Experimental Psychology*, 1972, **94**, 291–294.
- Friend, K. E. Perceptual encoding in comparative judgments of race. *Memory and Cognition*, 1973, **1**, 80–84.
- Greenberg, M. G. J-scale models for preferential choice behavior. *Psychometrika*, 1963, **28**, 265–271.
- Griggs, R. A., & Shea, S. L. Integrating verbal quantitative information in linear orderings. *Memory and Cognition*, 1977, **5**, 287–291.
- Groen, G. J., & Parkman, J. M. A chronometric analysis of simple addition. *Psychological Review*, 1972, **79**, 329–343.
- Helson, H. *Adaptation-level theory: An experimental and systematic approach to behavior*. New York: Harper, 1964.
- Holyoak, K. J. The form of analog size information in memory. *Cognitive Psychology*, 1977, **9**, 31–51.
- 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.
- Jamieson, D. G., & Petrusic, W. M. Relational judgments with remembered stimuli. *Perception and Psychophysics*, 1975, **18**, 373–378.
- Kelley, H. H., Hovland, C. I., Schwartz, M., & Abelson, R. P. The influence of judge’s attitudes in three methods of attitude scaling. *Journal of Social Psychology*, 1955, **42**, 147–158.
- Krantz, D. H. Small-step and large-step color differences for monochromatic stimuli of constant brightness. *Journal of the Optical Society of America*, 1967, **57**, 1304–1316.
- Marks, D. F. Relative judgment: A phenomenon and a theory. *Perception and Psychophysics*, 1972, **11**, 156–160.
- Moyer, R. S. Comparing objects in memory: Evidence suggesting an internal psychophysics. *Perception and Psychophysics*, 1973, **13**, 180–184.

- Moyer, R. S., & Bayer, R. H. Mental comparison and the symbolic distance effect. *Cognitive Psychology*, 1976, **8**, 228–246.
- 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 and Cognition*, 1975, **3**, 635–647.
- Parkman, J. M. Temporal aspects of digit and letter inequality judgments. *Journal of Experimental Psychology*, 1971, **91**, 191–205.
- Parkman, J. M., & Groen, G. J. Temporal aspects of simple addition and comparisons. *Journal of Experimental Psychology*, 1971, **89**, 335–342.
- Potts, G. R. Storing and retrieving information about ordered relationships. *Journal of Experimental Psychology*, 1974, **103**, 431–439.
- Restle, F. Speed of adding and comparing numbers. *Journal of Experimental Psychology*, 1970, **83**, 274–278.
- Rule, S. J. Equal discriminability scale of number. *Journal of Experimental Psychology*, 1969, **79**, 35–39.
- Rule, S. J., & Curtis, D. W. Conjoint scaling of subjective number and weight. *Journal of Experimental Psychology*, 1973, **97**, 305–309.
- Segall, M. H. The effect of attitude and experience on judgments of controversial statements. *Journal of Abnormal and Social Psychology*, 1959, **58**, 61–68.
- Sekuler, R., Rubin, E., & Armstrong, R. Processing numerical information: A choice time analysis. *Journal of Experimental Psychology*, 1971, **89**, 75–80.
- Shepard, R. N., Kilpatrick, D. W., & Cunningham, J. P. The internal representation of numbers. *Cognitive Psychology*, 1975, **7**, 82–138.
- Shipley, W. C., Coffin, J. I., & Hadsell, K. C. Affective distance and other factors determining reaction time in judgments of color preference. *Journal of Experimental Psychology*, 1945, **35**, 206–215.
- Shipley, W. C., Norris, E. D., & Roberts, M. L. The effect of changed polarity of set on decision time of affective judgments. *Journal of Experimental Psychology*, 1946, **36**, 237–243.
- Sternberg, R. J. Component processes in analogical reasoning. *Psychological Review*, 1977, **84**, 353–378.
- Thurstone, L. L. Psychophysical analysis. *American Journal of Psychology*, 1927, **38**, 368–389.
- 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, NJ: Erlbaum, 1975.
- Wallis, C. P., & Audley, R. J. Response instructions and the speed of relative judgments. II. Pitch discrimination. *British Journal of Psychology*, 1964, **55**, 121–132.
- Woocher, F. D., Glass, A. L., & Holyoak, K. J. Positional discriminability in linear orderings. *Memory and Cognition*, 1978, in press.
- Woodworth, R. S., & Schlosberg, H. *Experimental psychology*. New York: Holt, Rinehart & Winston, 1950.

REFERENCE NOTES

1. Holyoak, K. J. *Symbolic processes in mental comparisons*. Ph.D. thesis, Stanford University, 1976.
2. Baum, D. R. *Cognitive maps: The mental representation of geographic distance*. Ph.D. thesis, University of Michigan, 1977.