Modeling Imageless Thought:
The Relative Judgment Theory of Numerical Comparisons

STEPHEN LINK
McMaster University

The mental process of comparing numbers is shown to follow the principles of Relative Judgment Theory. Martingale predictions of the theory are confirmed in an extensive study of numerical comparisons using the Method of Symmetric Differences. The theory is then used as a tool to reveal the Psychonumeric Function relating psychological difference to numerical difference. This function transforms a difference between numbers into a sextal valued (0, 1, 2, 3, 4, 5) number with zero as a unit psychonumeric value. The Psychonumeric Function combined with the theory's prediction of response times gives a close account of the numerical comparison Chronometric Function.

INTRODUCTION

For many early experimental psychologists introspection seemed as natural a method for viewing mental phenomena as looking toward the heavens seemed an obvious means of viewing the stars. The author of the first English book on the new psychology, Sully (1884), described introspection as a method for directing our attention to what is going on in our mind at the time of its occurrence, or afterwards. We have the power of turning the attention inwards on the phenomenon of the mind. Thus I can attend to a particular feeling say, admiration for a beautiful object, in order to see what its nature is, of what elementary parts it consists, how it is affected by the circumstances of the moment, and so on.

Later critics, however, were quick to seize upon examples of mental processes that defied introspective analysis; in particular Ach (1905) who used the mental addition or subtraction of digits in studies of "imageless thought" (Boring, 1957; Woodworth, 1915, 1938). Another form of imageless thought that may precede the operation of mental arithmetic (Groen, 1967; Restle, 1970), is the mental comparison of numbers. The number three is easily judged to be smaller than seven, but no amount of searching the mind illuminates the mechanism, or elements of mind, that give rise to the judgment. The failure of introspective analysis to shed light on imageless thought proved to be a pitfall for introspectionism, but a positive increment in the drift of experimental psychology toward a more behavioristic ideology. Meanwhile, the mechanism for imageless thought lost theoretical significance, as behavior, rather than mind, became the newer psychology's principal interest.

Reprint requests should be addressed to Dr. Stephen Link, McMaster University Hamilton, Ontario, Canada L8S 4K1. This research is supported by the Natural Sciences and Engineering Research Council of Canada.

0022-2496/90 $3.00
Copyright © 1990 by Academic Press, Inc.
All rights of reproduction in any form reserved.
Today, the use of behavioral measurements combined with mathematical models of thought processes permits deeper analysis of psychological mechanisms than those afforded by either behaviorism or introspective analysis. These powerful "approches nouvelles" lead toward greater understanding of the mechanisms of thought and, in this paper, provide a description of a mental mechanism engaged to form a comparative judgment about numbers.

Because human judgments of quantity are inexact and variable the use of numbers is a practical everyday necessity. The precision we attribute to these numbers leads us to believe that our mental comparison of numbers is equally precise. We know that seven is larger than three and certainly smaller than nine. Thus it comes as a surprise to discover, as did Moyer and Landauer (1967), that such mental judgments are prone to systematic error and that the time required to judge whether one number is larger than another depends upon the size of the difference between them.

Results from other paired comparison studies of digits (Sekuler, Robin, & Armstrong, 1971; Moyer & Landauer, 1973; Buckley & Gillman, 1974) demonstrated that as the difference between digits increased, the response time and/or the probability of responding in error decreased; although a frequently encountered finding was that the number 1 produced a peculiar result and seemed to be treated in a fashion different from that for other digits. In some experiments response errors were somehow ignored, and not reported, on the presumption that errors represent only random effects rather than a consequence of the comparison process itself (cf. Sekuler et al., 1971). However, an exceptional experiment investigating comparisons of two-digit numbers, by Hinrichs, Yurko, and Hu (1981), and a subsequent experiment on multidigit number comparison (Hinrichs, Berie, & Mosell, 1982) did summarize error probabilities as a function of the difference (Δn) between a fixed standard and a variable comparison. Their results replicate the decline in the probability of an error, as Δn deviates from zero, first shown by Moyer and Landauer (1967). In a comprehensive doctoral dissertation on the application of Relative Judgment Theory to digit comparisons, Foltz (1982) found the theory to yield a close account of results from three experiments. These remarkably complete reports of experimental results provided a starting point for the experimental studies reported here.

**CORROBORATIVE INVESTIGATIONS: EXPERIMENTS 1 AND 2**

The phenomena of numerical comparisons are illustrated by a serial paired comparison experiment using two-digit numbers ranging from 11 to 99. The experimental procedure required each of four university students to compare a

---

1. Dr. Stephen Poltrock supervised Dr. Foltz's dissertation.
2. Miss Karen Rankin (1984) performed Experiments 1 and 2 as part of a B.Sc. thesis project directed by the author in his laboratory and assisted by Mr. Bodo Bilazewski.
randomly selected, visually displayed number, such as 37, against the previously presented number, say 85. In order to begin a self-paced trial and display a comparison number on a computer controlled X–Y display screen, a subject used the index fingers of each hand to depress and then release together two choice response keys. The subject depressed a response key labelled "SMALLER" to indicate that 37 was smaller than 85, or "LARGER" to indicate that 37 was (erroneously) thought to be larger than 85. The response terminated the visual display and stopped a clock which measured response times to an accuracy of 1 msec. No feedback was given. The reported response times are the durations from the first display of the comparison stimulus to the moment a response occurred. Subjects were seated 1 m from the X–Y display screen (Tektronix 602, white on black display, P4 phosphor). Single digits were 2 cm high and 1½ cm wide.

Each subject participated in five consecutive daily sessions containing three blocks separated by 5-min rest periods. Each block began with a standard of 55 followed by 18 practice numbers (each digit presented in the tens and units position twice), followed without pause by four randomizations of 88 different stimuli. Thus, three blocks yielded 1056 experimental trials. The first two sessions familiarized the subject with the task and do not enter into the analysis of data. The remaining three sessions contributed 1056 experimental trials each day for each of four subjects for a total of 12,672 trials. This Serial Method of Paired Comparisons provides for a broad range of differences between the comparison stimulus and standard to be obtained without using the 7,744 trials per subject required for a complete paired comparison design.

Shown in Fig. 1 are mean response times (correct and error responses included) as a function of \( \Delta n \), the difference between the comparison and variable standard. This function, which was defined previously as the Chronometric Function (Link, 1984), shows clearly that response time declines dramatically as the difference between numbers increases.

Shown in Fig. 2 are the proportions of error responses for the Serial Method of Paired Comparisons. These proportions are computed for half-decade intervals of \( \Delta n \) ranging from –88 to 88. As the value of \( \Delta n \) increases from –88 the proportion of “Larger” responses also continually increases. “Larger” responses to stimuli smaller than the standard are error responses. As \( \Delta n \) increases beyond zero the proportion of “Smaller” responses, errors in judging a larger comparison to be smaller, steadily decreases. The proportion of errors is greatest near \( \Delta n = 0 \). The error proportion is surprisingly large for numbers quite near each other. When the numbers differ by \( \pm 1 \) to 4, the proportion of errors is 0.14. When the numbers differ by \( \pm 5 \) to 9 the proportion of errors is 0.08 and when they differ by \( \pm 10 \) to 14 the proportion declines to 0.056.

For purposes of comparison, a second study using the same apparatus as in Experiment 1 employed the Method of Constant Stimuli. The numbers 11–99, exclusive of 55, were randomly permuted and then serially presented to subjects, who again self-paced the visual, response-terminated presentations. In contrast to Experiment 1, these subjects maintained in memory a fixed standard of 55. Five
FIG. 1. Mean response time as a function of the difference between comparison and standard numbers for two numerical comparison experiments.

FIG. 2. The proportion of errors as a function of Δn.
university students engaged in five consecutive daily sessions. A session of 1320 experimental trials contained three blocks, each consisting of 18 practice trials (each digit presented in the units and tens positions twice) followed, without pause, by trials formed from five randomized sequences of the 88 numbers. Again, two response keys were used, one key indicating a "Larger" and the other a "Smaller" judgment. The comparison number remained visible until a response occurred. The first three sessions and all practice trials were used as familiarization trials and were excluded from this analysis.

The mean response times for this experiment (shown in Fig. 1) and the proportion of errors (in Fig. 2) support and extend the findings of Hinrichs, Yurko, and Hu (1981) and demonstrate how mean response time and the relative frequency of an error decline as $\Delta n$ diverges from zero. Although the range of $\Delta n$ is considerably smaller in this experiment than that afforded by the Method of Serial Paired Comparisons, the decline in response time and error rate as $\Delta n$ deviates from zero is substantial.

Results from both experiments appear to be in keeping with predictions of the Relative Judgment Theory of the Psychometric Function (Link, 1975; Link, 1978a). The simplest of many predicted empirical laws, relating Chronometric and Psychometric (Urban, 1908) Functions, tested this hypothesis. In the case of data produced by the Method of Constant Stimuli (Experiment 2), good support for the theory of comparisons was obtained. The data from the Method of Serial Paired Comparisons (Experiment 1) yielded a decidedly poor fit of the data to predicted empirical relations.

Further analysis of data from Experiment 1 produced a familiar finding (Sekuler et al., 1971; Link, 1984) shown in Fig. 3. The figure shows that for responses conditioned on the decade of the standard, the mean response time for "Smaller" responses is greater than for "Larger" responses when the decade of the standard is less than fifty. This relation reverses as the decade of the standard increases beyond fifty.

Response proportions, also presented in Fig. 3, show a marked change as a function of the decade of the standard. The proportion of "Larger" responses steadily decreases as the decade of the standard increases to 90 from 10. Naturally, when the standard is small there is a high probability that the comparison will be larger than the standard. Thus the most prevalent correct response is "Larger." When the standard is large there is a high probability that comparison stimuli are smaller, resulting in a smaller number of "Larger" responses. The linear decline in the probability of a "Larger" response, as the decade of the standard increases, is a feature of the paired comparison design.

This experimental design ensures that the a priori probability of a standard in the 10's decade being followed by a larger comparison stimulus is 727/755, or nearly 1.0. A subject can perceive, from experience with previous trials, that a small standard is likely to be followed by a larger comparison. When this bias is taken into account, a much better fit of the data to the predictions of this theory are obtained, not only in this case, but also in results reported by Poltrock (1989).
Yet methodological problems inherent in the Method of Paired Comparisons are particularly evident in this study of numerical differences. The first problem derives from the influence of the size of a standard on response bias. The amount of response bias may remain constant for a fixed standard, but each comparison against the fixed standard produces a different value of $\Delta n$. In the most extreme cases, at the two off-diagonal corners of the design matrix, exist the maximum values for $\Delta n$. These allow for the maximum influence of standard-induced response bias and, of all the $\Delta n$'s, produce the smallest number of observations. At these extremes, estimates of effects due to $\Delta n$ become particularly difficult to interpret because extreme values of $\Delta n$ and response bias jointly influence performance. For some theories, separating these confounded effects of response bias and $\Delta n$ is difficult. For experiments which investigate these theories, simultaneous stimulus presentation thwarts the analysis of these confounded effects.

The influence of such confounding on response times is quite evident. In particular, when response times for the “Larger” and “Smaller” responses are viewed as a function of the magnitude of the standard, as shown in Fig. 3, the response time is fast whenever the response proportion is large. When the standard is within the 10’s decade the most likely response is “Larger,” because most comparison stimuli are, in fact, larger than the standard. Bias toward a “Larger” response reduces the mean response time, as might be expected. However, response time is
also affected by $\Delta n$. For “Larger” comparisons the average $\Delta n$ is greatest when the standard is in the 10's decade but continually diminishes as the decade of the standard increases. As $\Delta n$ increases response time decreases, as shown in Fig 1. Thus, two influences on response time both suggest a decrease in “Larger” response time as the decade of the standard decreases. How can these influences be successfully measured using a paired comparison design that confounds them?

The effect of these two factors on response time is evident in Fig. 4. Here, two-digit stimuli are grouped with respect to decade values. This perspective view of the mean response time surface for the Method of Serial Paired Comparisons shows how response time declines as the magnitude of $\Delta n$ increases. Comparisons within decades, which correspond to the small $\Delta n$'s occurring on the major diagonal, require greater response time than comparisons between decades. Even within decades, however, the shortest response times occur when the standard is extreme, in the 10's or 90's decade. The fastest response times also occur for these standards when they are combined with comparisons at the opposite extremes of 90 (408 msec) or 10 (388 msec), respectively. Note that response time is a maximum (610 msec) for comparisons within the 50's decade. Comparisons within the 10's and 90's decades require equal times of 514 msec. A quite noteworthy feature of this image is the symmetry of these mean response times with respect to the major diagonal.

The second methodological problem is that response bias and numerical difference both influence response probability. Thus, in the case of a standard stimulus in the 10's decade, both a bias toward responding “Larger” and the large average value for $\Delta n$ combine to produce an exceptionally high probability of a correct response when a larger comparison stimulus is presented. The probability of a
correct "Smaller" response is diminished for the few smaller comparison stimuli. Consequently, Psychometric or Chronometric Functions derived from the Method of Paired Comparisons require a correction for response bias in order to measure the unperturbed effect of stimulus difference. Third, with respect to a theory such as Relative Judgment Theory, a paired comparison design renders the estimation of parameters particularly difficult, due to the confounding of response bias and \( \Delta n \). More specifically, parameters are estimated easily when there are two stimulus conditions producing equal but opposite effects, such as comparing the numbers 70 to 30 or 30 to 70. However, for a paired comparison design matrix, such pairs have different standards and, therefore, potentially different response biases that invalidate the assumption of a fixed bias that leads to a simple method of obtaining parameter estimates. Within the middle of a paired comparison matrix it is possible to compare, say, 51 or 57 to a fixed standard of 54 in a manner consistent with the assumption of equal but opposite effects and fixed response bias. Yet, in this range the influence of response bias may be small anyway, and, as the standard diverges increasingly from the median value of 55, this method becomes less useful as the possible number of such comparisons dwindles to zero.

**Numerical Representations**

Much previous analysis of the process of numerical comparison assumes that the standard and comparison numbers are transformed into an analog quantity (e.g., Moyer & Landauer, 1967). A more specific translation specifies that numbers are represented logarithmically (Buckley & Gillman, 1974). As Shepard, Kilpatrick, and Cunningham (1975) also argued:

"... it is not unreasonable to suppose that the psychological spacing of numbers is essentially logarithmic in view of (a) the observation that the recovered separation between adjacent numbers was generally greater for small than for large numbers in the present Figs. 13-16, (b) the prevalence of the Weber–Fechner effect for magnitude continua of all sorts, and (c) Moyer and Landauer's (1967) finding that reaction times were more closely related to ratios than to differences between numbers.

The assumption that Weber's Law applies to numbers leads to the further assumption that Fechner's Law also applies. It is Fechner's Law which yields a logarithmic representation for numbers. If Weber's Law does not apply, then the resulting logarithmic assumption is called into question.

When it comes to judgments of numerical difference the evidence in support of the logarithmic assumption is slim indeed. For example, the results of Shepard et al. (1975) are based on similarity judgments, not judgments of difference. Buckley and Gillman (1974) claimed support for the assumption that mean response time for digit comparisons is a linear function of the reciprocal of the difference between logarithmically transformed numerical values. The support weakens, however, when their methodology is reviewed in detail. First, the data derive from digit comparisons that were forced to be error free by re-running all error trials until a
correct response occurred. Second, the two digits, $d_1$ and $d_2$, were presented side by side and the correct times to $(d_1, d_2)$ and $(d_2, d_1)$ were averaged, thus obscuring, but not eliminating, effects due to digit position. Third, the mean response times were ordered and the ordinal values, not actual values, were analyzed by a multidimensional scaling program, MD-SCAL, which recovered a two-dimensional stimulus representation.

Looking to Wald's (1947) sequential analysis model to account for their forced-correct response times, Buckley and Gillman presumably computed mean response times based on the general prediction from bounded random walk theory that the mean response time to a stimulus (that is, correct and error times included) equals the average distance travelled by the walk divided by the rate of drift. By defining the bounds of the random walk space to be equally distant from the starting position (which expresses the idea of unbiased responding), by assuming that the walk travelled only toward the correct response threshold (so there were no errors), and by assuming that the rate of drift equalled the difference between logarithms of $d_1$ and $d_2$ they generated several sets of mean response times for various values of the random walk boundaries (response thresholds). They then analyzed these means using the MD-SCAL program. For certain boundary positions the recovered stimulus representation was two-dimensional and appeared to be similar to that obtained from the ordinal data. On this basis the authors conclude that the logarithmic assumption is supported by the data.

There is a disturbing gap between the theory advanced by Wald and the application of it to numerical comparisons suggested by Buckley and Gillman. Most troubling is the fact that no predictions based on the theory are tested with the data. Also troubling is the idea that removing errors produces data in better correspondence with the theory. Major predictions from random walk theories concern errors. By specifically deleting them the theory is stripped of its distinctive predictions. The authors do comment that “Error rate was less than 3% and was positively correlated with correct time ($r = .68$, $p < .001$); there were too few errors for more systematic analysis.”

If anything this observation concurs with the more general psychophysical literature on comparative judgment. Apparently, the authors omitted error responses that provide support for some kind of random walk theory. For such theories, as the difference between stimuli decreases the rate of drift decreases and error rate increases. As the difference between stimuli decreases the rate of drift decreases and the mean response time increases, thereby producing a positive correlation between error rate and mean response time. But Wald's theory, in particular, predicts that error and correct times, conditioned on the response made, must be equal. This prediction is seldom observed in either choice or discrimination response times, and probably not in the Buckley and Gillman experiment.

Producing a random walk theory of comparative judgment that avoided this awkward prediction was one reason the author introduced a new random walk theory in 1972/73 that appeared in publication in 1975 (Link & Heath, 1975; Link, 1975). The sequential theory adopted by Buckley and Gillman attempted to apply these
random walk ideas, but without a clear understanding of the relationship between theory and data. By artificially restricting the results only to correct performance, the error data needed to show that the theory requires a logarithmic transformation of numerical value to account correctly for errors are lost. As far as the mean response times are concerned, even a linear transformation might do equally well (Link, 1978a).

Weber's Law predicts that discriminability declines for a fixed Δn but increasing values of the standard. From numerous psychophysical studies where Weber's Law does hold (Luce, 1986; Welford, 1980), response time, for fixed Δn, is expected to increase as values of the standard increase. This does not occur in Experiment 1. Instead, the mean response time for comparisons within the 10's decade, which are based on small values of Δn, equals 514 msec. The mean response time for comparisons within the 90's decade, which are also based on small Δn, also equals 514 msec. Further analysis of the serial paired comparison experiment reported above shows no tendency for mean response time to increase for a fixed Δn = 1 as the magnitude of the numbers being compared increases. In fact, as seen in Fig. 4, the mean response times for small Δn's, which correspond to the within-decade comparisons shown on the major diagonal, decrease as the value of the standard deviates in either direction from its mid-range—a result consistent with the effect of response bias. Thus, the assumption that Weber's Law applies to numerical comparisons receives no support from these studies.

Similar conclusions might be drawn from previous paired comparison studies were it not for the undesirable amalgamation of response times across the main diagonal of the paired comparison design matrix. This method of averaging data combines response times for pairs such as (1, 9) and (9, 1), thereby obscuring the influence of the standard. In yet other studies the two numbers were presented simultaneously, thereby successfully removing the possibility of ever determining whether one stimulus acted as a standard that promoted response bias, and another as the comparison.

None of these various results provides support for the assumption that Weber's Law applies to the comparison of numbers. Furthermore, the often observed decline in response time as a function of stimulus difference, that is, the Chronometric Function, is easily accounted for by the assumption of a linear transformation of stimulus difference, even in psychophysical studies (Link, 1975, 1978a). Thus the decline in response time as a function of numerical difference does not require the assumption of a logarithmic transformation for a successful account of the Chronometric Function.

Some researchers suppose that the representation of number is invariant across different tasks. Others, such as Banks and Coleman (1981), suppose the scale of numbers depends upon the task. Weissmann, Hollingsworth and Baird (1975) imply "that numbers are processed in the same manner as other physical attributes and, consequently, can be catalogued according to measures such as difference limen, d', and information transmitted" (Baird, 1975). Baird (1975) postulates that "all continua are transformed by characteristic combinations of base systems
operating in the manner described for numbers' (Noma & Baird, 1975)." But Noma and Baird (1975), investigating digit, base, and quarter models of the generation of numbers greater than an experimenter-defined lower bound, conclude that "a combination of models seems necessary to adequately predict number generation," and presumably number representation, as well. Thus, many assumptions are made concerning the form of a transformation of a number into an internal representation used in a numerical comparison task, but there is still substantial uncertainty regarding what that transformation may be, and whether a single transformation applies to all numerical tasks.

Furthermore, studies of numerical comparisons have yet to establish that whatever transformation does occur is based upon the individual numbers rather than the difference between them. The importance of obtaining an internal representation for individual stimuli derives from the belief, commonly held among psychophysicists, that a comparative difference is derived from the internal representations of individual stimuli. In the case of numerical comparisons, there is no evidence to show this to be the case. The difference between numbers may be determined first, then the comparison process acts on the transformed difference.

These various considerations directed the present investigation toward a less restrictive assumption—that for a fixed standard the effect of a positive discrepancy is equal but opposite to the effect of a negative discrepancy of equal absolute value. No assumption is made about the nature of the numerical representation. If the theory accounts for results obtained under this weak assumption, then the way in which performance changes, as a function of the magnitude of the discrepancy, can be explored.

The study of comparative judgments based on equal but opposite stimulus differences is facilitated by a new experimental design called the Method of Symmetric Differences. This design removes the confounding of response bias and stimulus difference and gives a stringent test for the theory that describes how these judgments are made. Before this design is introduced, the theory and its implications are presented.

**THE THEORY OF RELATIVE JUDGMENT**

Relative Judgment Theory (RJT) describes a sequential sampling model of the sensory and psychological processes used to form mental judgments. The theory postulates that a comparison stimulus, $S_c$, is transformed into a waveform having an amplitude at time $t$, $s_c(t)$, which is compared against an internal psychophysical referent represented by another waveform, $s_r(t)$. In the case of numerical comparisons the standard, $N_c$, is transformed into a waveform $n_c(t)$. The comparison number, $N_r$, is transformed into another waveform $n_r(t)$. The comparison between $N_c$ and $N_r$ must produce a waveform $dn(t)$ that contains information about the difference between $N_c$ and $N_r$. These waveforms characterize the electrical signals representing the stimuli and their difference.
For comparisons between numbers the information in $dn(t)$ is extracted by a mental algorithm, $\alpha$, which operates on $dn(t)$ to produce a comparative value. These values, $\alpha[dn(t)]$, are accumulated over time during a trial either until enough positive values produce an accumulated total equal to, or larger than, the response threshold for a "Larger" response, $R_L$, or until sufficient negative values accrue to produce a "Smaller" response, $R_S$. The process of accumulation is illustrated in Fig. 5, where the starting position of an accumulation of comparisons is labelled $C$ and the two possible terminal values are labelled $A$ and $-A$. During each unit of time a new value of the waveform $dn(t)$ is sampled and the value of $\alpha[dn(t)]$ is added to previously accumulated values. Note that these values are variable; thus the sum performs a random walk, with drift, toward thresholds, $A$ or $-A$, and produces stochastic paths such as the one illustrated.

The values of $A$, $-A$, and $C$ are assumed to remain constant during an experimental trial. Across trials these values may change to meet experimenter-imposed task demands. The value of $A$ can be reduced to meet response time deadline restrictions. For smaller values of $A$, the subject emits faster responses and is more responsive than for larger values of $A$. Thus, the reciprocal of $A$ is an index of responsiveness.

For the purpose of discussion and derivation the standard, $N_s$, is assumed to be a fixed number such as 55, while the comparison, $N_e$, varies. Many experimental results are conditioned upon a given value for the standard, in accordance with this assumption. For results averaged across standards the expected value calculations treat the standard as a variable quantity without doing any injustice to theoretical derivations.

**TWO-THRESHOLD RANDOM WALK MODEL**

![Diagram of a stochastic path showing accumulation of comparative difference](image)

**FIG. 5.** A stochastic path shown to absorb at the value $A$ after a number of accruals of comparative difference.
Obvious qualitative predictions derive from this theory. For example, when the physical difference between a comparison stimulus and standard is positive, the expected value of $E@[\eta N_c, N_r]$ should be positive. For the present purposes, the mental algorithm, $@$, is assumed to generate a constant expected value for the duration of the judgment process. This provides that for any value of $t$, the mean or expected value, $E@[\eta N_c, N_r]$, equals $\mu$. This is the rate at which the sum of differences drifts toward the “Larger” response threshold positioned at the value $A$. Thus, large values of $\mu$ lead to quick termination at the response threshold, while smaller values result in longer termination times. The termination time for reaching the response threshold is the decision time and forms the component of response time that is of major interest.

A second qualitative feature depends upon the effect of an initial predisposition toward one response at the expense of the other. This predisposition is measured by the amount of response bias, $C$. When the bias assumes a positive value, representing a predisposition to favor the response “Larger,” a shorter time will be needed to reach the “Larger” response threshold than when there is no response bias at all and the value of $C$ equals zero. When the response bias is negative, the response “Smaller” is favored and the time required to reach the “Larger” response threshold will increase beyond the time required when the response bias is zero. Other factors remaining constant, when response bias changes from a value of $+C$ toward a value of $-C$, the time to correctly decide “Larger” increases.

A third qualitative prediction derives from the stochastic nature of the process of comparison and accumulation. Because the values $@[\eta t]$ are continually sampled during the decision process, and because these values are variable, there will be instances when a stochastic path with positive expected value terminates at the $-A$ threshold, resulting in an erroneous response. That is, in spite of a positive rate of drift, the response made on a trial can be “Smaller” rather than “Larger.” Other things being equal, such errors are more likely to occur when there is substantial bias toward responding “Smaller,” and to decline steadily as the response bias reaches zero and then turns positive, indicating a predisposition to favor the response “Larger.” Also, for comparisons between quite similar numbers, that is, for values of $\mu$ near zero, there should be a higher error rate than when $\mu$ deviates substantially, either positively or negatively, from zero. The error rate is also influenced by variability in the values accumulated. When the drift rate is near zero, the influence of variability in these values is more significant than when there is a non-zero drift rate that pushes the process to completion quickly.

The agreement between the data in Figs. 1, 2, and 3 and these qualitative predictions provides a solid base for launching a more formal investigation of theoretical predictions. This investigation requires a new experimental technique in order to test important predictions concerning changes in response time and response probability, which result from variations in theoretical parameters. Understanding the advantages of this new experimental technique is facilitated by discussing the more formal properties of the theory.

Theoretical predictions are easily derived when an ideal pair of stimuli have
Let \( N_{r+i} \) be a comparison stimulus \( di \) units greater than the standard \( N_r \). “Symmetric effects” means that when \( N_{r+i} \) is judged against a standard, \( N_r \), the comparison created is the mirror image, with respect to zero, of the comparison when \( N_{r-i} \) is presented. That is, \( \hat{\omega}[dn(t)] \) is a reflection, with respect to zero, of \( \hat{\omega}[-dn(t)] \).

For a fixed standard, the mean comparison values, \( \mu_i \) and \( \mu_{-i} \), are equal but opposite in sign, so that \( \mu_i = -\mu_{-i} \).

While a previous analysis of this process (Link & Heath, 1975) used the mathematical approach of random walk theory, the weaker assumptions used here permit the application of the Theory of Martingales (Doob, 1953; Hall, 1969; Shiryayev, 1984). Assuming stationarity of \( \hat{\omega}[\pm dn(t)] \), there exist two Martingales, a sub-Martingale and a super-Martingale, with parameters \( \theta_{i} = -\theta_{-i} \). These satisfy two Wald Identities (Wald, 1947):

\[
E\{\exp - \hat{\omega}[dn(t)] \theta_{i}\} = 1
\]

\[
E\{\exp - \hat{\omega}[-dn(t)] \theta_{-i}\} = 1.
\]

The value of \( \theta_{i} \), like that of \( \mu_{i} \), generally increases with stimulus difference. These values of \( \theta \) play an important supporting role in the theory, because response probabilities depend upon them directly. For example, a number that is larger than the standard produces a correct response, \( R_L \), with probability

\[
P_{L,i} = \Pr(R_L, \text{"Larger"} | \text{Stimuli } N_{r+i}, N_r \text{, and Bias } C) = \frac{\exp(\theta_{i}A) - \exp(-\theta_{i}C)}{\exp(\theta_{i}A) - \exp(-\theta_{i}A)}. \tag{1}
\]

The probability of response \( R_L \), given the mirror-image stimulus \( N_{r-i} \), is the probability of responding in error to a stimulus smaller than the standard. This probability is

\[
P_{L,-i} = \Pr(R_L, \text{"Larger"} | \text{Stimulus } N_{r-i}, N_r \text{, and Bias } C) = \frac{\exp(\theta_{-i}A) - \exp(-\theta_{-i}C)}{\exp(\theta_{-i}A) - \exp(-\theta_{-i}A)}. \tag{2}
\]

These results from ideal symmetric stimuli yield as estimators of the distance from the common starting position, \( C \), to the common response thresholds at \( A \) and \( -A \), the terms \( \theta_{i}(A - C) \) and \( \theta_{i}(A + C) \).

\[
\ln[P_{L,i}/P_{L,-i}] = [\theta_{i}(A - C)] \tag{3}
\]

and

\[
\ln[(1 - P_{L,-i})/(1 - P_{L,i})] = [\theta_{i}(A + C)]. \tag{4}
\]

Equations (3) and (4) add to give \( 2\theta_{i}A \), while the difference between (4) and (3) yields \( 2\theta_{i}C \). In this way, the previously unknown response threshold values, \( A \) and
and the amount of response bias, \( C \), are estimated from response proportions up to a multiplicative constant, \( \theta \), which depends upon the difference between the stimuli.

Changes in the values of \( \theta \), \( A \), and \( C \) each cause changes in response probabilities. First, increases in the value of \( \theta \), which typically occur when the magnitude of the difference between stimuli increases, will cause increases in the values of \( \theta A \) and \( \theta C \) even though \( A \) and \( C \) remain constant. An increase in the value of \( \theta A \) is often, correctly, interpreted as an increase in discriminability as measured by \( \theta \). Even so, a change in \( \theta A \) cannot, in its own right, be distinguished from a change in \( \theta A \) due to a change in the value of the response threshold, \( A \), used to reach a judgment. By comparison, in Signal Detection Theory (Swets, Tanner & Birdsall, 1955), there is a single measure of discriminability, \( d' \), for which estimates can be made from equations identical to (3) and (4) (Noreen, 1979; Ogilvie & Creelman, 1968). Within Signal Detection Theory an increase in \( d' \) can only be attributed to an increase in sensitivity. Changes in the response threshold value, \( A \), which alter the subject's responsiveness would be misinterpreted as a change in sensitivity. To avoid this possible misinterpretation, an experimental design that shows the separate influences of \( \theta \) and \( A \) is desirable (Link, 1978b).

Another important but unexpected effect is that as the value of \( \theta \) increases the estimated values of response bias, \( \theta C \), must also increase, although the actual bias, \( C \), remains constant. For small values of \( \theta \), the value of \( \theta C \) will remain near zero, but as the value of \( \theta \) increases the estimate of \( \theta C \) must also be seen to increase. There is no similar prediction in the Theory of Signal Detection (Green & Swets, 1961), where such a change would be attributed to an inexplicable change in response bias as a function of discriminability. A carefully designed experiment that tested this prediction in the context of memory judgments, thought to occur by the process described here, demonstrates quite clearly this apparent increase in response bias as \( \theta \) increases (Noreen, 1970).

**Response Time Predictions**

Random walk theory is simplified by Martingale analysis. In particular, response time predictions, response probability derivations, and relations between response
times and response probabilities follow from Wald's Identity (Wald, 1947) which follows from Martingale Theory (Doob, 1953; Shiryayev, 1984). While the results are the same as those obtained from the stationary random walk analysis at the heart of Relative Judgment Theory, a greater degree of generality applies, especially to the predictions concerning response times. The prediction that the expected time to reach one or the other of the two response thresholds equals the average size of the accumulated discrepancies divided by the rate of accumulation remains unchanged. Thus nothing is lost while the Martingale context suggests new techniques for theoretical analysis.

When number \( N_{r,i} \) is presented, the rate at which discrepancies accrue, commonly referred to as \( \mu_i \), is equal to the expected value of \( @[dn(t)] \). This average decision time is one component of the observed response time which is postulated to be the sum of the decision time and a mean time representing all other non-decision response time components. Thus, the equation for the time taken to respond to number \( N_{r,i} \) is

\[
\text{RT}_i = \text{Decision Time to } N_{r,i} + \text{Non-decision Time.} \tag{5}
\]

The average response time can be written in a form that shows its theoretical origins, that is, as

\[
\bar{\text{RT}}_i = \frac{\text{Average Amount of Accumulated Difference}}{\text{Rate of Accumulating Differences}} + K
\]

\[
= \frac{P_{L,i}(A - C) + (1 - P_{L,i})[-(A + C)]}{\text{Rate of Accumulating Differences}} + K
\]

\[
= \frac{A(2P_{L,i} - 1) - C}{\mu_i} + K \tag{6}
\]

where \( P_{L,i} = \text{Pr}(R_{L_i} \text{ "Larger" } \mid N_{r,i}, N_r, \text{ and } C) \), and \( K \) is the mean non-decision component of response time. It is easily seen that with no response bias the value of \( C \) is 0 and the response time then becomes

\[
\bar{\text{RT}}_i = \frac{A(2P_{L,i} - 1)}{\mu_i} + K. \tag{7}
\]

Many previous psychophysical experiments confirm this relation between response time and response probability. An exemplary experiment is that by Kellogg (1931), who measured the response times and response proportions of well-practiced subjects who discriminated differences in luminance. For seven luminance differences, he discovered that the response time became longer as the difference between luminances decreased. Using this assumption that the value of \( \mu_i \) is a linear function of the deviations between such luminances, Link (1978a) showed the empirical relation described by Eq. (7) to obtain good support from Kellogg's data.
Further evidence is provided by data from a classic experiment on line-length discrimination reported by Shallice and Vickers (1964), as well as from later experiments by Link and Tindall (1971).

When the assumption of a known relation between physical stimulus differences and psychophysical differences is undesirable or an impossibility, as it is for stimuli having no physical measurements, Relation (6) can be rewritten to provide a separate method of testing the relationship between response time and response probability. This alternative form permits evaluation of theoretical predictions in experimental paradigms that commonly employ two stimuli having the property of symmetry. Multiplying both the numerator and denominator of (6) by the unknown $\theta_i$ gives the equations

\[
\bar{RT}_i = \frac{\theta_i A(2P_{L,i} - 1) - \theta_i C}{\theta_i \mu_i} + K
\]

\[
= (\theta_i \mu_i)^{-1} [\theta_i A(2P_{L,i} - 1) - \theta_i C] + K
\]

\[
= (\theta_i \mu_i)^{-1} [Z_i] + K
\]

where the value $Z_i$ can be estimated from observed response proportions and the application of (3) and (4).

For a pair of stimuli, $N_r$ and $N_{r+i}$, the values of both $\theta_i$ and $\mu_i$ remain constant, while changes in $A$ and $C$ produce changes in response times and response probabilities. When plotted against $Z_i$, values of response time should exhibit a linear increase with slope equal to the reciprocal of $\theta_i \mu_i$ and an intercept that estimates the unknown value of the mean non-decision time. This is the well-documented $RT$ vs $Z$ relation characteristic of random walk theories and supported by many experiments (Link, 1975; Noreen, 1979; Link, 1978b).

A similar equation for response time applies to presentations of the symmetric comparison $N_{r-i}$. In this case, by replacing in (8a) $\theta_i$ with $\theta_{-i}$, $\mu_i$ with $\mu_{-i}$, $P_{L,i}$ with $P_{L,-i}$, and then noting that $\theta_{-i} = -\theta_{+i}$ and $\mu_{-i} = -\mu_{+i}$, we find that the mean time to respond to $N_{r-i}$ given $N_r$ and bias $C$ equals

\[
\bar{RT}_{-i} = \frac{A(1 - 2P_{S,-i}) - C}{\mu_{-i}} + K
\]

\[
= \frac{A(2P_{S,-i} - 1) + C}{\mu_{-i}} + K
\]

\[
= (\theta_{i} \mu_{i})^{-1} [\theta_{i} A(2P_{S,-i} - 1) + \theta_{i} C] + K
\]

\[
= (\theta_{i} \mu_{i})^{-1} [Z_{-i}] + K
\]

where $P_{S,-i}$ is the probability of response "Smaller" given stimulus $N_{r,-i}$.

As an investigative tool, these analyses can isolate causes of changes in response time and response probability and permit response times to be corrected for the influence of response bias. Although the correction for response bias is implicit in
previous papers on Relative Judgment Theory, the present circumstance provides an opportunity to illustrate the use of this correction in detail. Note that the difference between (8b) and (9b) gives

$$[Z_i - Z_{-i}] = \theta_i \mu_i [\bar{RT}_i - \bar{RT}_{-i}]$$

which gives a direct estimate of

$$\theta_i \mu_i = [Z_i - Z_{-i}] / [\bar{RT}_i - \bar{RT}_{-i}]. \quad (10)$$

Alternatively, $\theta_i \mu_i$ is estimated better when several experimental conditions provide data for a regression estimate of the slope of this similarity transformation. Because $\theta_i \mu_i$ is known and because $\theta_i A$ and $\theta_i C$ are available from (3) and (4) it is possible to estimate $K$ in either (8b) or (9b). Correcting the observed response time for response bias follows from computing the theoretical value for the response times, with the parameter $C$ set equal to zero. Note, however, that $\theta C$ is also a component of the response probability which enters into the equation for decision time. Thus, a procedure that merely subtracts a component of response time, such as $C/\mu_i$, is not sufficient to correct for response bias. More directly, an additive model of the effect of response bias on response time is not adequate because the response probabilities also influence the response time through the multiplication of $A/\mu_i$.

Previously, the measure of discriminability, $d'$, used in the Theory of Signal Detection had no comparable measure obtained from response time data. How, however, it is apparent from (8a) and (9a), relating response time and response probability, that a component of response time due to bias depends upon $C/\mu_i$ and another component due to discriminability depends on $A/\mu_i$. These considerations suggest a new measure of relative discriminability equal to $\mu_i/A$, that is, the expected difference divided by the amount of discrepancy needed to generate a response. Because (10) yields an estimate of $\mu_i \theta_i$ and (3) and (4) give a value of $\theta_i A$, the estimate of the unknown relative discriminability is

$$\frac{\mu_i}{A} = \left[ \frac{Z_i - Z_{-i}}{\bar{RT}_i - \bar{RT}_{-i}} \right] / 2 \ln \left( \frac{P_{L,i} P_{S,-i}}{P_{L,-i} P_{S,i}} \right). \quad (11)$$

The values of $\mu_i/A$ are relative psychological differences that depend upon the difference between a standard, $N_i$, and a comparison, $N_{i+1}$.

This analysis reveals that decision times are largely a consequence of three factors: the effect of stimulus difference as measured by $\theta$ or $\mu$; responsiveness, $1/A$; and response bias, $C$. The fact that these various parameters are estimated from a pair of responses, to stimuli having symmetric effects, suggests immediately that any direct evidence regarding the application of the theory to judgments of differences between numbers will require an experimental design based upon the idea of stimulus symmetry. Moreover, an experimental design that treats the effects of discriminability and response bias as orthogonal will permit straightforward estimation of the amount of influence of each of these factors.
While the Method of Constant Stimuli used in Experiment 2 retains the very useful feature of presenting comparison stimuli that are both larger and smaller than a fixed comparison stimulus it provides no way to estimate any effects of response bias. Extending the Method of Constant Stimuli, by adding a design feature that allows for the occurrence and estimation of response bias, permits the response bias to become apparent.

**Experimental Design: The Method of Symmetric Differences**

To achieve this end, each standard stimulus is followed by a comparison stimulus that is larger or smaller than the standard, with probability 0.50. In contrast to the standard presented in the Method of Paired Comparisons, the standard now conveys no information concerning the possible direction of a stimulus difference. The potential for manipulating or measuring the existence of response biases is provided for by permitting the standard to vary across a range of values, so that information other than the value of the standard may be introduced by the experimenter, or used by the subject, in order to bias performance. For example, payoffs may heavily reward particular stimulus–response combinations. Given the predictions of Relative Judgment Theory, this influence is measurable for each value of the standard and for each symmetric stimulus difference.

In the parlance of experimental designs, the Method of Symmetric Differences is a novel factorial arrangement of standards and stimulus differences that generates, by necessity, a rectangle with as many columns as there are standards and as many rows as there are positively and negatively valued deviations from the standards. The particular design of the experiment described below employs standards ranging from 33 to 77 (including 55) and stimulus differences, $\Delta n$, ranging from $-22$ to $+22$ (excluding the case of zero) for a total of 1936 trials. A selected standard and value of stimulus difference, $\Delta n$, generate a comparison stimulus value equal to the value of the standard plus the value of $\Delta n$.

Figure 6 illustrates the features of the experimental design used in Experiment 3. The Method of Symmetric Differences includes many useful features of the Method of Constant Stimuli, while avoiding the features of paired comparison design matrices that confound the influences of response bias and stimulus difference. This is accomplished by allowing the range of the comparison stimuli to depend upon the value of the standard in such a way that, for any standard, the probability of a larger comparison stimulus equals 0.50. At the same time, the range of possible comparison stimuli remains constant (the matrix will be called “complete” for this reason). A consequence of the “complete” design is that the probability distribution for the comparison stimuli is triangular, unlike the uniform probability distributions for comparisons characteristic of the Methods of Constant Stimuli and Paired Comparisons.

An implication of a design matrix that produces a marginal probability of a

---

4 Miss Mary Ann Azzarello (1985) performed Experiment 3 as part of a B.Sc. thesis project directed by the author in his laboratory and assisted by Mr. Bodo Bilazewski.
Method of Symmetric Differences

Design Matrix

<table>
<thead>
<tr>
<th>Standards</th>
<th>Comparison Values</th>
</tr>
</thead>
<tbody>
<tr>
<td>34</td>
<td>12 13 ... 55 56</td>
</tr>
<tr>
<td>33</td>
<td>11 12 ... 54 55</td>
</tr>
</tbody>
</table>

Probability Distribution of Comparisons

Fig. 6. The design matrix for the Method of Symmetric Differences produces uniform distributions of standards (rows) and Δn’s (columns). Entries are values of comparison stimuli obtained by adding values for the corresponding row and column.

Larger comparison equal to 0.50 is that the waiting time for a repetition of a positive difference is asymptotically negative binomially distributed. Thus, no information regarding which correct response will be required can be drawn from the value of the standard presented to the subject, or from the preceding sequence of trials.

Experimental Procedure

The design matrix illustrated in Fig. 6 contains 1936 possible pairs of standard and comparison stimuli. Presenting all these combinations requires more than a single 1-hr subject session. To meet various procedural demands, the matrix was partitioned into blocks of trials requiring no more than 1 hr to complete. Each within-subject block consisted of 22 practice trials followed, without pause, by 462 experimental trials. The practice trials were drawn from the 88 elements that define the two diagonals of the design matrix, so that only during a series of practice trials did the subject ever see a comparison number equal to the average value of all comparison numbers, 55. Within each series of practice trials the comparison value 55 occurred 11 times, while the remaining comparison values were drawn at random from the alternative diagonal entries (11, 13, ..., 53, 57, ..., 97, 99). This selection procedure also ensured that the 88 practice trials contained one instance of each value of the standard and two instances of each value of Δn. During a subject session of 1 hr, two different blocks containing half the randomly-selected trials of the design matrix were used. By combining four different blocks, a complete replication was obtained in two 1-hr sessions which occurred on successive days.

Each of five subjects participated in this experiment for a payment of $35 for 7 days. During the first 2 days the complete design matrix occurred once. These
data were treated as familiarization and do not enter into the analysis of data. The next 4 days yielded two replications of the design matrix and $2 \times (1936 - 88) = 3696$ experimental trials per subject, for a total of 18,480 trials. The last day was reserved to present several blocks of the 88 practice trials in order to examine the influence of stimulus difference and magnitude of the standard by using familiar stimuli presented with a greater frequency than before. These data are not a part of analysis reported here but will be reported elsewhere.

For each experimental session the subject sat in a darkened room 1 m from a Tektronix 602 (P4 phosphor; white on a black background) display screen (8 cm in height and 10 cm wide), surrounded by a large (71 cm wide by 61 cm height) flat black screen that prevented the seated subject from looking behind the display screen. The experimenter instructed each subject to “decide if the second number presented is numerically larger or smaller than the first... Make your response as quickly and as accurately as possible.”

To begin a trial, the subject depressed and then released two response keys by using the index fingers of each hand. Thereafter, the standard stimulus, a two-digit numeral, with single digits approximately 2 cm wide and $1\frac{1}{3}$ cm high, remained visible for 1000 msec followed by an empty, dark display screen for 500 msec, followed by the continuous display of the two-digit comparison number until the subject responded. A Digital Equipment PDP-11/34 computer programmed to control the experiment also recorded the response choice and response time, accurate to 1 msec. Subjects received no indication of whether a response was correct or in error.

**Experimental Results**

To simplify the data, the design matrix was collapsed into $\Delta n$ ranges of plus or minus 1–3, 4–6, 7–9, 10–12, 13–15, 16–18, and 19–22 (a copy of the complete matrix of results is available upon request). A similar reduction of data across the decades of the standard (30, 40, 50, 60, and 70) yielded the results in Table 1. The entries in this table are based upon data obtained from all five subjects across 4 days. The entries in each cell of Table 1 correspond to the total number of “Smaller” or “Larger” responses and the mean response times for “Larger” and “Smaller” responses. From cell to cell there may be differing numbers of observations. The variation is due to the differing numbers of standard-$\Delta n$ pairs, the reductions to decades of standards, and the loss of eight trials due to a computer memory malfunction. There are a total of $5 \times 2 \times (44 \times 44 - 88) - 8 = 18,472$ trials.

---

5 Eight trials were lost through a memory failure that duplicated the results of eight contiguous trials as if they were normal experimental trials. The eight trials that were so replaced were permanently lost, thereby reducing the total to 18,472. One bizarre response time of 8700 msec for Subject 5 in responding “Larger” to a standard of 62 and a comparison of 80 was replaced by the subject's second datum value of 629 msec.
<table>
<thead>
<tr>
<th>$\Delta n$</th>
<th>33–39</th>
<th>40–49</th>
<th>50–59</th>
<th>60–69</th>
<th>70–77</th>
<th>Average</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Smaller</td>
<td>Larger</td>
<td>Smaller</td>
<td>Larger</td>
<td>Smaller</td>
<td>Larger</td>
</tr>
<tr>
<td>$N$</td>
<td>5</td>
<td>235</td>
<td>13</td>
<td>387</td>
<td>5</td>
<td>355</td>
</tr>
<tr>
<td>MEAN</td>
<td>990.4</td>
<td>709.9</td>
<td>983.3</td>
<td>705.1</td>
<td>820.4</td>
<td>689.5</td>
</tr>
<tr>
<td>$N$</td>
<td>238</td>
<td>1</td>
<td>399</td>
<td>1</td>
<td>356</td>
<td>4</td>
</tr>
<tr>
<td>MEAN</td>
<td>179</td>
<td>1</td>
<td>296</td>
<td>4</td>
<td>268</td>
<td>2</td>
</tr>
<tr>
<td>$N$</td>
<td>670.4</td>
<td>1316</td>
<td>643.2</td>
<td>987</td>
<td>689.5</td>
<td>1045.7</td>
</tr>
<tr>
<td>MEAN</td>
<td>820.4</td>
<td>689.5</td>
<td>695.0</td>
<td>730.4</td>
<td>711.5</td>
<td>682.5</td>
</tr>
<tr>
<td>$N$</td>
<td>8172</td>
<td>3</td>
<td>297</td>
<td>2</td>
<td>268</td>
<td>2</td>
</tr>
<tr>
<td>MEAN</td>
<td>866.7</td>
<td>778.9</td>
<td>699.7</td>
<td>640.0</td>
<td>726.4</td>
<td>728.0</td>
</tr>
<tr>
<td>$N$</td>
<td>179</td>
<td>1</td>
<td>296</td>
<td>4</td>
<td>268</td>
<td>2</td>
</tr>
<tr>
<td>MEAN</td>
<td>650.4</td>
<td>734</td>
<td>699.7</td>
<td>640.0</td>
<td>726.4</td>
<td>728.0</td>
</tr>
<tr>
<td>$N$</td>
<td>7</td>
<td>203</td>
<td>7</td>
<td>263</td>
<td>10</td>
<td>260</td>
</tr>
<tr>
<td>MEAN</td>
<td>1066.4</td>
<td>711.9</td>
<td>817.8</td>
<td>687.9</td>
<td>939.0</td>
<td>682.7</td>
</tr>
<tr>
<td>$N$</td>
<td>210</td>
<td>0</td>
<td>267</td>
<td>3</td>
<td>265</td>
<td>5</td>
</tr>
<tr>
<td>MEAN</td>
<td>673.0</td>
<td>711.9</td>
<td>817.8</td>
<td>687.9</td>
<td>939.0</td>
<td>682.7</td>
</tr>
<tr>
<td>$N$</td>
<td>7</td>
<td>203</td>
<td>7</td>
<td>263</td>
<td>10</td>
<td>260</td>
</tr>
<tr>
<td>MEAN</td>
<td>800.7</td>
<td>722.4</td>
<td>815.0</td>
<td>704.0</td>
<td>703.0</td>
<td>706.9</td>
</tr>
<tr>
<td>$N$</td>
<td>210</td>
<td>0</td>
<td>270</td>
<td>0</td>
<td>266</td>
<td>4</td>
</tr>
<tr>
<td>MEAN</td>
<td>640.6</td>
<td>722.4</td>
<td>815.0</td>
<td>704.0</td>
<td>703.0</td>
<td>706.9</td>
</tr>
<tr>
<td>$N$</td>
<td>8</td>
<td>202</td>
<td>6</td>
<td>264</td>
<td>6</td>
<td>264</td>
</tr>
<tr>
<td>MEAN</td>
<td>756.0</td>
<td>734.8</td>
<td>908.0</td>
<td>712.1</td>
<td>1422.8</td>
<td>728.0</td>
</tr>
<tr>
<td>$N$</td>
<td>207</td>
<td>2</td>
<td>268</td>
<td>2</td>
<td>267</td>
<td>3</td>
</tr>
<tr>
<td>MEAN</td>
<td>668.0</td>
<td>546.0</td>
<td>685.1</td>
<td>463.0</td>
<td>724.9</td>
<td>965.0</td>
</tr>
<tr>
<td>$N$</td>
<td>12</td>
<td>198</td>
<td>6</td>
<td>284</td>
<td>6</td>
<td>234</td>
</tr>
<tr>
<td>MEAN</td>
<td>768.8</td>
<td>743.7</td>
<td>1068.5</td>
<td>730.9</td>
<td>1474.6</td>
<td>719.4</td>
</tr>
<tr>
<td>$N$</td>
<td>210</td>
<td>0</td>
<td>285</td>
<td>5</td>
<td>231</td>
<td>9</td>
</tr>
<tr>
<td>MEAN</td>
<td>696.3</td>
<td>726.5</td>
<td>734.6</td>
<td>757.6</td>
<td>1222.1</td>
<td>783.7</td>
</tr>
<tr>
<td>$N$</td>
<td>7</td>
<td>203</td>
<td>8</td>
<td>291</td>
<td>10</td>
<td>200</td>
</tr>
<tr>
<td>MEAN</td>
<td>901.2</td>
<td>789.5</td>
<td>828.7</td>
<td>741.2</td>
<td>773.9</td>
<td>755.1</td>
</tr>
<tr>
<td>$N$</td>
<td>206</td>
<td>4</td>
<td>293</td>
<td>7</td>
<td>293</td>
<td>7</td>
</tr>
<tr>
<td>MEAN</td>
<td>777.2</td>
<td>1061.3</td>
<td>751.7</td>
<td>665.0</td>
<td>762.9</td>
<td>1128.0</td>
</tr>
<tr>
<td>$N$</td>
<td>54</td>
<td>1416</td>
<td>50</td>
<td>2049</td>
<td>44</td>
<td>1845</td>
</tr>
<tr>
<td>MEAN</td>
<td>861.8</td>
<td>742.9</td>
<td>890.3</td>
<td>718.7</td>
<td>955.5</td>
<td>711.6</td>
</tr>
<tr>
<td>$N$</td>
<td>1460</td>
<td>8</td>
<td>2078</td>
<td>22</td>
<td>1857</td>
<td>33</td>
</tr>
<tr>
<td>MEAN</td>
<td>682.5</td>
<td>933.5</td>
<td>690.0</td>
<td>801.3</td>
<td>716.8</td>
<td>1005.5</td>
</tr>
</tbody>
</table>

**Table 1**

Number of Observations and Mean Response Times for Smaller and Larger Responses for Symmetric $\Delta n$ Values across the Standard Stimulus Decades.
Fig. 7 Three Chronometric Functions illustrate the influence of response bias on mean response time. The base response time is 640 msec, which equals the mean response for \(-7\) to \(-9\) \(A_n\) for standards in the 30's decade.

Mean response times as a function of \(A_n\), but averaged within ranges of the standard, are shown in Fig. 7. These Chronometric Functions are similar to those obtained using the Methods of Constant Stimuli and Paired Comparisons. For the standard in the range 50–59, the symmetry with respect to \(A_n = 0\) is obvious. Also shown in Fig. 7 are the conditional Chronometric Functions based on the two extreme ranges of the standards, those ranging between 33 and 39 and those between 70 and 77, inclusive. The influence of the decade of the standard on the mean response time is readily apparent. When the standard is drawn from the 30's decade, mean response time is fast for negative values of \(A_n\), that is, for values of the comparison number smaller than the standard, but is substantially slower for symmetric positive values of \(A_n\). At the other extreme, when the standard is in the 70's decade, a negative value for \(A_n\) produces a slow response time when compared with the response times for corresponding positive values of \(A_n\). The difference between the Chronometric Functions for the 30's or 70's decade and the 50's decade, where there is no response bias, measures the effect of bias on response time.

The Psychometric Error Functions conditioned on the 30's, 50's, and 70's decades of the standard are not illustrated, but also show the influence of the decade of the standard on response errors. When the standard is drawn from the 30's decade, error response proportions are low for negative values of \(A_n\), but when the standard is from the 70's decade the error response proportions are lower for the positive values of \(A_n\). For the 50's decade, the Psychometric Error Function reveals little tendency for either positive or negative \(A_n\)'s to generate more errors.

Error response proportions are summarized in Fig. 8. For the 30's decade the "Larger" error response proportion is 0.005. This proportion rises to .035 at the 70's
The low, but informative, error rates for "Smaller" and "Larger" responses depend in part upon the magnitude of the standard. Conversely, the "Smaller" error response proportion equals .036 for the 30's decade and declines to .009 in the 70's decade. The marginal proportion of errors is only 0.02. This small error rate is characteristic of a sequential mechanism that operates quickly with a high probability of success but suffers an occasional error. Typically, sequential sampling devices require fewer samples of \( \pm [dn(t)] \) to reach a decision than do fixed sample size systems operating at the same error rate. Of course, when error rates are low, large numbers of trials are required for one to observe that the predictions about errors are accurate.

In spite of the design considerations for the Method of Symmetric Differences, these results suggest that the value of the standard still promotes response bias. That substantial bias does exist is seen in Fig. 9, where mean response times for larger and smaller comparison stimuli are shown for two similar experiments. The paired comparison experiment, whose Chronometric Function is shown in Fig. 1, has mean response times for larger and smaller comparison stimuli that also vary as a function of the decade of the standard but in a manner exactly opposite to those obtained from the Method of Symmetric Differences. Paired comparison method response times for comparison stimuli larger than a standard are fast for small values of the standard and increase with increases in the value of the standard. For the Method of Symmetric Differences, the mean response times decline for the larger comparisons, as the value of the standard increases. A similar reversal is
MEAN RESPONSE TIMES
LARGER AND SMALLER COMPARISONS

METHOD OF SYMMETRIC DIFFERENCES

METHOD OF PAIRED COMPARISONS

FIG. 9. Mean response times (including correct and error responses) for larger and smaller comparison stimuli in the Method of Paired Comparisons and the Method of Symmetric Differences, shown as a function of the decade of the variable standard stimulus.

evident for the smaller comparison stimuli. Because results from the Method of Symmetric Differences are different from those for the Method of Paired Comparisons, it is apparent that previous conclusions drawn from the Method of Paired Comparisons need careful interpretation.

THEORETICAL ANALYSIS

Three parameters are the basis for the variations in response proportions and times characteristic of numerical comparisons. The purpose of this section is to estimate these parameters, or functions thereof, and show how they change as a function of the factors of the Method of Symmetric Differences design matrix. Equation (3) and (4) were used within each combination of $\Delta n$ level and decade of the standard to obtain the estimates of $\theta A$ and $\theta C$ reported in Table 2. When marginal results are referred to, they are the averages of these estimators, rather than estimates based on marginal averages of the data, although the two methods of estimation lead to similar numerical values.

The parameter $\theta$, like $\mu$, is determined by numerical difference. Although the function relating $\theta$ to $\Delta n$ is unknown, generally, as $\Delta n$ increases in value, so
TABLE 2
Estimates of $\theta A$ and $\theta C$ Based on Data within Intervals of $\Delta n$ and Decades of the Standards

<table>
<thead>
<tr>
<th>Range of standard</th>
<th>1-3</th>
<th>4-6</th>
<th>7-9</th>
<th>10-12</th>
<th>13-15</th>
<th>16-18</th>
<th>19-22</th>
<th>Average</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\langle \theta A \rangle$</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>33–39</td>
<td>3.65</td>
<td>4.42</td>
<td>3.93</td>
<td>4.70</td>
<td>4.70</td>
<td>4.13</td>
<td>4.66</td>
<td>4.32</td>
</tr>
<tr>
<td>40–49</td>
<td>3.66</td>
<td>3.95</td>
<td>4.34</td>
<td>4.96</td>
<td>4.06</td>
<td>4.45</td>
<td>4.69</td>
<td>4.30</td>
</tr>
<tr>
<td>50–59</td>
<td>3.22</td>
<td>3.45</td>
<td>4.14</td>
<td>4.08</td>
<td>3.61</td>
<td>4.90</td>
<td>4.38</td>
<td>3.97</td>
</tr>
<tr>
<td>60–69</td>
<td>3.66</td>
<td>3.67</td>
<td>3.87</td>
<td>5.04</td>
<td>4.03</td>
<td>4.65</td>
<td>4.11</td>
<td>4.15</td>
</tr>
<tr>
<td>70–77</td>
<td>3.40</td>
<td>4.57</td>
<td>3.78</td>
<td>5.01</td>
<td>4.21</td>
<td>4.29</td>
<td>4.33</td>
<td>4.23</td>
</tr>
<tr>
<td>Average</td>
<td>3.52</td>
<td>4.01</td>
<td>4.01</td>
<td>4.76</td>
<td>4.12</td>
<td>4.48</td>
<td>4.43</td>
<td>4.19</td>
</tr>
</tbody>
</table>

| $\langle \theta C \rangle$ |     |     |     |       |       |       |       |         |
| 33–39             | -0.27| -1.56| -0.68| -1.30 | -1.30 | -1.02 | -0.79 | -0.99   |
| 40–49             | -0.07| -0.09| -0.54| -1.31 | -0.42 | 0.14  | -1.27 | -0.51   |
| 50–59             | -0.29| 0.20 | -0.34| -0.11 | -0.34 | 0.00  | -0.11 | -0.14   |
| 60–69             | -0.06| 0.14 | 0.59 | 1.23  | 0.48  | -0.34 | 0.08  | 0.30    |
| 70–77             | 0.42 | -0.20| 0.95 | 1.14  | 1.18  | 0.34  | 1.26  | 0.73    |
| Average           | -0.05| -0.30| 0.00 | -0.07 | -0.08 | -0.18 | -0.17 | -0.12   |

should $\theta$. When there is no difference, and $\Delta n = 0$, the value of $\theta$ must also equal zero. To examine how $\theta$ varies, two different views of the joint value, $\theta A$, were obtained. The changes in $\theta A$ as a function of $\Delta n$, viewed in Fig. 10, shows that as $\Delta n$ increases, the value of $\theta A$ also increases steadily. This must be the case for the theory to have any application to psychophysical studies. The apparently small increase in $\theta A$ over the full range of $\Delta n$ is, in fact, a full log unit change. This rate of change promotes a large reduction in the percentage of error responses as $\Delta n$ increases.

One could argue that this increase in $\theta A$ is due to simultaneous changes in $\theta$ and $A$, and not to $\theta$ alone. Therefore, to show that $A$ does not change, a second view of $\theta A$ is obtained by viewing $\theta A$ as a function of the magnitude of the standard for the five decades: 30's, 40's, 50's, 60's and 70's, where $\theta$ characterizes average performance obtained by averaging across values of $\Delta n$. As seen in Fig. 11, the value of $\theta A$ remains fairly constant. The constancy also suggests that discriminability, as measured by $\theta$, does not decline as the standard increases in magnitude. Hence, unlike peripheral stimuli that for a fixed $\Delta n$ often generate poor performance as the standard increases, these results show no evidence of a Weber effect.

Response bias is determined by the value of $C$. For large positive values of $C$ performance is biased toward the response “Larger.” As $C$ shifts toward zero and becomes increasingly negative, performance reflects a bias toward the response “Smaller.” Although $C$ determines the amount and direction of response bias only
**Fig. 10.** Marginal estimates of $\theta A$ and $\theta C$ as a function of the differences between comparison and standard numbers.

**Fig. 11.** Marginal values of $\theta A$ and $\theta C$ as a function of the decade of the standard.
θ times C, that is θC, can be determined from equations 3 and 4. To simplify the analysis of changes in θC the data were averaged across values of Δn to produce only a single value of θ representative of the averaged data. For these data, values of θC are also shown in Fig. 11.

Somewhat surprisingly, θC shows a linear increase as the standard increases to the 70's from the 30's decade. This change in response bias affects the Chronometric and Psychometric Functions. On the average the value of θC is nearly zero, indicating no bias at all, and for the 50's decade the value of θC is also nearly zero. This unbiased performance is reflected in the Chronometric Function for the 50's decade shown in Fig. 7. One unit of response bias toward the “Smaller” response threshold is evident in the 30's decade and the Chronometric Function for the 30's decade illustrates the influence of this bias. The mean response times for negative values of Δn, that is, for comparison numbers smaller than the standard, are less than times for responses to comparisons larger than the standard. Equally important, the mean “Smaller” response times in the 30's decade are also smaller than corresponding times in the 50's decade where performance is unbiased. For positive values of Δn mean response times are longer in the 30's decade than in the 50's decade. When the response bias changes to favor the “Larger” response, as in the 70's decade, the mean response time for positive values of Δn is lower than for negative values of Δn. Also, the positive values of Δn generate mean response times less than those in the 50's decade, where performance is unbiased. The large mean response times for negative Δn's are less than corresponding results for the unbiased performance of the 50's decade. These results are consistent with theoretical predictions.

A major prediction of this theory relates measures of response time and response probability. The theory predicts the existence of a linear function having an intercept that measures non-decision aspects of the discrimination task, while the slope is a measure of discriminability, (8b) and (9b). To analyze this prediction, averages across larger or smaller comparison stimuli were used to generate 10 mean response times, two for each of the five decades of the standard. For each of these 10 cases, the corresponding marginal values for θA and θC were computed and used to determine the values of Z in (8b) and (9b). The obtained relation is shown in Fig. 12, where there is little question about the predicted linearity. The correlation between RT and Z proved to be 0.88, which is quite satisfactory for ten data points obtained by averaging the data across all values of Δn.

To investigate the RT vs Z relation further, values of Δn were collapsed into three levels, 1–7, 8–14, 15–22, and the same RT versus Z relation was calculated for each of the three levels. The correlations between RT and Z are of particular interest here, because for small values of discriminability the values of μ and θ are small, leading to long response times but a restricted range of values for RT and Z. A strong correlation is more easily shown when the ranges of RT and Z are large, as they are for higher versus lower discriminability. For these three levels (1–7, 8–14, and 15–22), the ranges of RT and Z became less restricted as the values of Δn increased and, as expected, the correlations increased from 0.37 for the 1–7 range to 0.86 and 0.91 for the 8–14 and 15–22 ranges, respectively.
STEPHEN LINK

METHOD OF SYMMETRIC DIFFERENCES

Fig. 12. The linear prediction for RT vs Z in (8b) and (9b) is shown to fit larger and smaller comparisons for five decades of the standard.

The RT versus Z relation is but one of many predictions leading to the conclusion that Relative Judgment Theory describes accurately the mental process activated when visually-presented two-digit numbers are compared. Systematic changes in parameter estimates also show the Method of Symmetric Differences to separate the influences of stimulus difference and response bias. The action of these two experimenter-controlled variables in promoting changes in response time, seen through the conditional Chronometric Functions in Fig. 7, illustrates again how response bias plays its role in determining response time.

More important, a theory that describes the precise way in which response bias affects response time also provides a method of correcting response times for response bias. For those researchers who readily measure response time, rather than response proportions, this theory provides an answer to the question of how an RT bias measure can be computed. The values of $\theta C$ give a direct measure of response bias, but when they are combined with estimates of $\mu 0$, $\theta A$, and $K$, all components of the response time equations are available and the influence of response bias can be removed.

THE NUMERICAL TRANSFORMATION $[@dn(t)]$

The final question to be treated here is: How does a difference between numbers give rise to the differences in mean response time that are so clearly evident in the
Chronometric Functions shown in Fig. 13? The upper function shows results for all five subjects. Analyses of individual performances discovered that the last of five subjects was far more variable in performance than the others. This variability had a volatile influence on the Chronometric Function and explains, in part, the large difference in response time between the two sets of results shown in Fig. 9. Removing this subject from the analysis, in order to reduce variability in the Chronometric Function, did not alter conclusions concerning the role of response bias, or other conclusions drawn from the previous analyses. However, the mean response time based on four subjects was substantially reduced, as seen in Fig. 13.

In order to further refine the Chronometric Function, response times to symmetric differences were averaged. This average response time, which is half the sum of (8a) and (9a), removes the unknown value of $C/\mu$ from this mean. The mean response time for an absolute numerical difference of size $i$ is

$$
\overline{RT}_{|i|} = \left[ \frac{P_{L,i} + P_{S,-i} - 1}{\mu_i} \right] + K.
$$

Note that both $P_{L,i}$ and $P_{S,-i}$ still contain the effect of response bias, but, because they can be estimated from the experimental data, the mean response time can be plotted as a function of the term in brackets. When the value of $\mu_i$ is substituted the values of $\overline{RT}_{|i|}$ and the term in brackets should provide a linear function.
The average Chronometric Function is closely fit by the theoretical function based on a seximal number system.

The Chronometric Function based on (12) is shown in Fig. 14. The figure suggests that there are definite periods in the Chronometric Function. There is a steady decline in response time until a numerical difference of 6 occurs; then response time increases, only to once again begin a decline that is halted by an increase in response time, followed by another decline, and so forth. Similar results are evident in Fig. 1 and in the figures of Hinrichs et al. (1981, 1982). In previous studies of numerical comparisons the value of $\mu$ was assumed to derive from a monotonic function. On the basis of (12), a clear view of the value of $\mu$, uncontaminated by response bias, gives the first inkling that $\mu$ is not a simple transformation of numerical difference.

These systematic declines followed by increases in RT suggest a mechanism for determining the value of $\mu$ quite distinct from a similarity transformation of a difference that is the basis for many simple psychophysical judgments.

Rather, the Chronometric Function suggests that the difference between the numbers $N_{t+i}$ and $N_{t}$, $dn(t)$, is acted upon by a mental algorithm, $@[dn(t)]$, to produce a two-digit number based on the values (0, 1, 2, 3, 4 and 5), that is, a base-6 or seximal number. At most, two base-6 places are required to describe the small range (1-22) of differences used here. These will be designated by the symbols [ ] for the most significant and $<$ > for the least. The decimal numbers 1 through 5 correspond to their seximal equivalents. The decimal difference of 6 equals a seximal value of 10, i.e., $6_{10} = 10_{6} = [1] * 6^1 + <0> * 6^0$. 

---

**Fig. 14.** The average Chronometric Function is closely fit by the theoretical function based on a seximal number system.
The equation used above weights each digit by its appropriate power of 6 for an unambiguous monotonic transformation of a difference base 10 into a difference base 6. However, the weights attached to the sextal digits extracted from the numerical difference are not necessarily those of the sextal number system. They are weights that depend upon the importance of each digit to the subject making numerical comparisons. Of the total weight, a proportion \( p \) is applied to the most significant digit, and a proportion \( (1 - p) \) to the least. This algorithm produces a value of \( \mu \) equal to \( p[dn(t)] + (1 - p)<dn(t)> \), as shown in Fig. 16.

A last feature of numerical comparisons concerns the meaning and representation of zero. For sensory stimuli, the absolute threshold stimulus produces a psychophysical representation value equal to zero. The stimulus magnitude at absolute zero is not zero but some physical value. The actual presentation of no stimulation produces, theoretically, an infinitely large negative sensation. For numbers, however, the presentation of zero is hardly the same as the presentation of nothing. Psychologically, zero is something and therefore must be represented as a value. The numerical difference adds to the representation for zero. Although the value representing zero is not known, it can be used as a unit of numerical representation having a value equal to 1.0. For this reason the results of a numerical comparison, \( \mu \), are added to the value 1.

These considerations lead to the determination of \( \mu \) from the equation

\[
\mu = p[dn(t)] + (1 - p)<dn(t)> + 1
\]

\[
= p\{[dn(t)] + 1\} + (1 - p)\{[dn(t)] + 1\}.
\]

(13)

For example, when \( \Delta n = 6 \), the value of \( \mu \) is

\[
\mu = p[1] + (1 - p)<0> + 1
\]

\[
= 1 + p.
\]

When \( \Delta n = 5 \) the value of \( \mu \) is

\[
\mu = p[0] + (1 - p)<5> + 1
\]

\[
= 6 - 5p.
\]

Therefore, if \( p \) is less than \( 5/6 \), as would seem natural for weights applied to digits whose importance depends upon their position or order of extraction, then the effective comparative difference for a numerical difference of 6 is smaller than that for a numerical difference of 5. Although \( \mu \) does not increase monotonically, values of \( \mu \) increase in unit steps for numerical differences of 7, 14, 21, and so forth.

To determine a suitable value for \( p \), the \( RT_{ij} \) vs \( Z_{ij} \) relation of (12) was used with values of \( \mu_{ij} \) substituted into the denominator. The linear relation shown in Fig. 15 resulted from a value of \( p = .78 \). Ninety-five percent of variance in the 22 mean response times was accounted for by this linear relation. The visual fit of theory and data shows that (13) is very sensitive to \( p \). Values of \( p \) of .75 or .80 produced
obviously poor fits to the data. For purposes of comparison, Fig. 15 also shows the relation obtained if \( \mu \) is taken to be the best fitting similarity transformation, \( \mu = a \times \Delta n \), of the numerical difference. There is little doubt that the similarity transformation, although simple, is quite inferior to the values of \( \mu \) obtained from (13).

A convenient way to illustrate how numerical differences produce a Chronometric Function is to treat the value of \( \mu \) as a function of numerical difference. In order to distinguish this function from the Psychophysical Function, often applied to sensory continua, this function of numerical difference will be called the Psychonomic Function. Its values are illustrated in Fig. 16. The linear increases occur within sextades, that is, intervals containing the digits 0, 1, 2, 3, 4 and 5, of the sextal transformation of a base-10 numerical difference. Declines in the value of \( \mu \) occur at the beginning of each sextade. The failure of \( \mu \) to increase linearly is due to the weights applied to the most and least significant digits of the sextal transformation. Unit increases in \( \mu \) occur for decimal differences equal to 0, 7, 14, 21, and so forth. Averaging data over several values of \( \Delta n \), as is often done to increase the number of observations per mean response time, would tend to reduce, if not eliminate, these discontinuities. The result would yield a nearly linear relation between \( \mu \) and \( \Delta n \).

The best linear fit in Fig. 15 generated the predicted Chronometric Function in Fig. 14. This two-parameter fit to the decision component of response time (the
MODELING IMAGELESS THOUGHT

FIG. 16. The Psychonumeric Function shows how numerical differences are transformed into rates of drift for the numerical comparison process.

number base is usually not considered as part of the goodness of fit) provides an excellent account of performance. The rate of descent of mean response time depends on the value of $p$, while the discontinuities occur at the beginning of each sextade. The small perturbation at the numerical difference of 17, rather than 18, may be due to a faulty determination of a relatively large difference between a standard and comparison. Or it may follow directly from a recursive application of the algorithm for 0–11, to the numbers 12–23, etc. This deviation is worthy of further study, but is small compared to the overall performance of this theory of numerical transformation and comparison.

CONCLUSION

The discussion of the theory of numerical comparisons focuses on four questions. First, what is learned from the theoretical account of changes in response probability and response time? Second, in what way does the theory's description of numerical representation change previously held views regarding the conception of number? Third, to what extent do these results generalize to other psychophysical tasks? Last, what are the implications of these results for the teaching of arithmetic skills?
With regard to the theoretical account of performance measures, such as response time and probability, there are substantial changes that concur with the general finding in psychological research, that the greater the separation between stimuli the faster the time to distinguish between them. However, on the basis of Fechner's interpretation of Weber's Law, experiments using sensory stimuli should show that as the magnitude of a standard stimulus increases in value, a fixed difference becomes more difficult to judge. This is clearly not the case with two-digit numbers, as is shown by estimates of $\theta A$. This measure of discriminability remains constant across changes in the decade value of the standard, which ranges from 30 to 70, more than a two-fold increase in magnitude. Thus, discriminability does not decline as the magnitude of the standard increases.

The kind of error changes as a function of the magnitude of the standard. For small standards there is a higher probability of an error in responding "Smaller" to a larger comparison. This is due to a bias toward responding "Smaller" when the standard is small. A concomitant reduction in response time for smaller comparisons, and an increase in response time for larger comparisons, is predicted to, and does, occur. When the standard is large, the response bias is toward the "Larger" response. The probability of an error to smaller comparisons is relatively high. Furthermore, the response time to smaller comparisons is slow when compared to responses to larger comparisons.

Similar changes in response times were characterized by Banks, Fujii, and Kayra-Stuart (1976) as due to a "semantic congruity effect." Their account does not explain the concomitant changes in response proportions. The "semantic congruity effect" seems to come and go, depending upon the circumstance of the experiment. For the Method of Paired Comparisons, where the bias is opposite to that obtained using the Method of Symmetric Differences, consistency forces one to argue that there is a "semantic incongruity effect." There seems to be little value in creating yet another name to describe experimental results due to response bias.

There is a legitimate question regarding the change in response bias found in numerical comparisons performed using the Method of Paired Comparisons and that found using the Method of Symmetric Differences. For paired comparisons the reason for response bias is readily apparent—there is a high probability that a small (large) standard will be followed by a large (small) comparison. The subject will perform better by taking advantage of this information.

For symmetric differences the response bias reverses. The reversal may be due to the association that occurs between a comparison stimulus $N_{r+i}$, and the subsequent numerical difference, $\Delta n = N_{r+i} - N_r$. The expected value of this difference, conditioned on a fixed value of the comparison number,

$$E[\Delta n | N_{r+i}] = \sum_r P[\Delta n | N_{r+i}] \Delta n.$$

increases linearly from the smallest to the largest comparison stimuli. For example, for a comparison of 11 this expected difference is $-22$, while for the largest com-
parison of 99 the conditional expected difference is 22. The subject may associate the magnitude of a difference with the value of the comparison stimulus. Later, when a standard is presented, this information is applied to the value of the standard, resulting in a bias toward responding “Smaller” when the value of the standard is small and “Larger” when the standard is large.

The difference between numerical stimuli determines the value of \( \mu \) which is critical to the prediction of response times. The three major components of the representation are: a base-6, or sextal, transformation of the base-10 difference between the comparison and the standard stimulus; a differential weighting of the most and least significant place values for the difference; and a representation for zero that characterizes it as a non-zero value. From this mental algorithm comes a general formula for the Chronometric Function,

\[
\frac{RT}{A[2P_c - 1] - C} = \frac{p[An_6] + (1 - p)\langle An_6 \rangle + 1}{1} + K,
\]

where \( An_6 \) is the sextal value of the base-10 difference

- \( A \) is the response threshold
- \( C \) is the response bias
- \( P_c \) is the probability of a correct response
- \( An \) is the base-6 numerical difference
- \( p \) is the weighting of importance given to the most significant digit, \( [An] \),
- \( (1 - p) \) is the weight given to the least significant digit, \( \langle An \rangle \), and
- \( K \) is the non-decision time component of \( RT \).

The psychophysical representation of these numerical differences depends upon a base-6 transformation. This base system has digit values 0, 1, 2, 3, 4, and 5, the same as the Chinese abacus and the numerical system of Finno-Ugrian. More obviously it is also the numerical base used for counting on the fingers and thumb of one hand. In the psychophysical production of numbers, studied extensively by Baird (1975), various base systems proved to be the basis for performance. In these results the base-6 system is the clear choice.

Of particular significance is the role played by the representation of zero. From a theoretical point of view every waveform, \( s(t) \), must have values in order for it to characterize a stimulus. If the mental representation of zero is allowed to equal nothing, then there is not even a waveform for its characterization. This enigma is settled by giving a non-zero value to the representation of zero.

Previous analyses of numerical comparison judgments showed that the number one required special treatment. On the basis of the present results, the psychophysical value of one is still intermediate between zero and two, but the starting point for the numerical representations is a unit, not zero. Earlier results may profit from re-examination under the assumption that zero must be given a non-zero value. The
need for a non-zero psychological representation for nothing seems clear, even in
the case of such a well-known anomaly as, “Every set contains the empty set, which
contains nothing except the empty set.” Obviously, the empty set must be a mental
something that cannot equal a mathematical nothing.

The results from this study of numerical comparisons do not prove that the
standard and comparison stimuli are first converted to a sextal representation and
then compared. At this point, we can only conclude that the difference between
stimuli is converted. The path from the presentation of the stimulus into its charac-
teristic waveform, followed by its use in the judgment process, is quite likely a series
of queued mental algorithms operating on waveforms carrying information about
the stimuli being compared. The algorithms and their use are the focus of simple
experiments which are now underway.

The implications of these results for the teaching of numerical skills requires
special mention. Mathematicians hold the common belief that the brain is an
orderly device, if not an algebraic machine, that is subject to error because the
student is lax in his or her studies. They are uniformly astonished to discover that
the Chronometric Function is not flat, but shows instead that different times are
required to judge whether 47 is smaller than 59 or whether 47 is smaller than 51.

The arithmetic of the schoolbook is not the calculation of the mind, as the many
studies of numerical comparisons bear out. An important step forward in the
instruction of children about numerical processes is to separate out errors due to
the comparison process itself and errors due to misconceptions about the mental
algorithm that should be applied. The recognition by mathematicians and
educators that the brain in not an arithmetic device, needing only to be teased, if
not reinforced, into accurate performance, can lead to improvements in the quality
of mathematical training.

The results reported here show that the same process of comparison is at work
as was previously shown (e.g., Link, 1975; Link & Heath, 1975; Noreen, 1979) to
account for the psychophysical discrimination of difference. The process of judg-
ment depends on information extracted from the waveform which characterizes a
difference between numbers. Sometimes an error occurs. Such errors are predictable
in terms of both response time and frequency. Because they are predictable they are
not the random intrusions of experimental error so often assumed by 19th century
theoreticians, and systematically ignored by some accounts of numerical compari-
sions. Rather, errors are inherent part of the process of making judgments.

Response bias plays a significant role in limiting the accuracy of performance, yet
reduces the average time taken to respond. The form of the response bias shown by
these subjects is compatible with the “semantic congruity effect,” where smaller
standards generate a bias to respond “Smaller” and larger standards lead to a bias
to respond “Larger.” Judging from the reports of Banks et al. (1976), Banks and
Flora (1977), and the more psychophysical investigations of Jamieson and Petrusic
(1975), the semantic congruity effect is simply a manifestation of response bias
rather than a result of a linguistic theory that neither predicts response time nor
response errors. According to Poltrock’s analysis of paired comparisons (1989),
response bias helps account for paired comparison results when the direction of response bias is opposite to that expected by the semantic congruity effect.

The process of comparing numbers is not directly observable through the mind's eye. Thus, speculation about the processes giving rise to these "imageless thoughts" is often difficult to disconfirm. Yet, the methods for analyzing "imageless thought" illustrated here show how elusive thought processes may be brought under control in a laboratory setting, as Külpe argued many years ago, and how their basis can be profitably explored through the application of modern psychophysical theory.

REFERENCES


**RECEIVED:** November 24, 1986