Journal of Experimental Psychology:
Human Perception and Performance
1990, Vol. 16, No. 3, 626–641

0096-1523/90/$00.75

Is Numerical Comparison Digital? Analogical and Symbolic Effects in Two-Digit Number Comparison

Stanislas Dehaene
Institut National de la Santé et de la Recherche Médicale
Laboratoire de Sciences Cognitives et Psycholinguistique,
Paris

Emmanuel Dupoux and Jacques Mehler
Centre National de la Recherche Scientifique
Ecole des Hautes Etudes en Science Sociales
Laboratoire de Sciences Cognitives et Psycholinguistique,
Paris

Do Ss compare multidigit numbers digit by digit (symbolic model) or do they compute the whole magnitude of the numbers before comparing them (holistic model)? In 4 experiments of timed 2-digit number comparisons with a fixed standard, the findings of Hinrichs, Yurko, and Hu (1981) were extended with French Ss. Reaction times (RTs) decreased with target-standard distance, with discontinuities at the boundaries of the standard's decade appearing only with standards 55 and 66 but not with 65. The data are compatible with the holistic model. A symbolic interference model that posits the simultaneous comparison of decades and units can also account for the results. To separate the 2 models, the decades and units digits of target numbers were presented asynchronously in Experiment 4. Contrary to the prediction of the interference model, presenting the units before the decades did not change the influence of units on RTs. Pros and cons of the holistic model are discussed.

Moyer and Landauer (1967) showed that reaction times for deciding which of two digits is the largest decrease as the numerical distance between the two increases. This finding, called the *distance effect*, was previously found in perceptual comparisons of various materials, for example, the length of bars (Johnson, 1939). Since then, it has been reproduced many times with miscellaneous materials: digits (Banks, Fujii, & Kayra-Stuart, 1976; Buckley & Gilman, 1974; Parkman, 1971; Sekuler & Mierkiewicz, 1977; Sekuler, Rubin, & Armstrong, 1971; see also Restle, 1970), two-digit numbers (Hinrichs, Yurko, & Hu, 1981), dot arrays compared for numerosity (Buckley & Gilman, 1974), objects indicated by name and compared for size (Holyoak, 1977; Kosslyn, Murphy, Bemesderfer, & Feinstein, 1977; Moyer, 1973), and abstract orderings with no physical counterpart (Woocher, Glass, & Holyoak, 1978). One particularly compelling experiment (Buckley & Gilman, 1974) should be noted, in which the same subjects were tested in two different paradigms: comparison of numerosities and comparison of digits. The results were essentially identical for the two tasks, as assessed by a multidimensional scaling procedure.

The theoretical interpretation of the distance effect in numerical comparison has been the matter of some debate. To some researchers, the continuous decrease of comparison times with numerical distance, a finding similar to that observed in psychophysical comparisons, suggests that digits may be encoded analogically on a mental map called *number line* (Buckley & Gilman, 1974; Moyer, 1973; Moyer & Landauer, 1967; Restle, 1970). The distance effect was easily accounted for in analogical models by hypothesizing that small distances impair the encoding or the retrieval of positions of objects on the number line. Several successful analogical models have been proposed along these lines. The most successful to date is probably Jamieson and Petrusic's (1975) reference point model, but random walk models seem to be adequate, too, especially when predicting error patterns (Buckley & Gilman, 1974; Poltrock, 1989).

Other researchers point out that the distance effect can be explained without resorting to analogical encoding. Banks et al. (1976) proposed a semantic-coding model of comparison in which the objects to be compared are initially labeled only as "large" or "small." If the numbers bear different labels, a response can be given. If the labels are the same, a supplementary (presumably constant) amount of time has to be spent for a response to be reached. Because the labeling process is probabilistic (the boundary between large and small varies randomly), the closer the two objects, the more likely they are to fall under the same label, and thus the longer the response time. In addition to the distance effect, Banks' semantic-coding model accounts for other findings of comparison tasks, such as the congruity effect (Banks et al., 1976; Banks & Root, 1979; Jamieson & Petrusic, 1975).

The analogical–propositional debate has remained controversial, owing to the lack of a precise definition of analogical representations. In particular, several researchers have failed to acknowledge the fact that no model of numerical comparison can be purely analogical. Obviously, the visual input in a numerical comparison task is always encoded into Arabic numerals or some other symbolic coding system. Thus, any "analogical" model requires a symbolic encoding device that computes the magnitude of the number from its symbolic

We thank Jean-Luc Aucouturier, Yves Barbin, Laurent Cohen, Susana Franck, Peter Jusczyk, Steven Poltrock, and two anonymous reviewers for their help at various stages of this research. This work was made possible with grants from the Centre National de la Recherche Scientifique (ASP Processus cognitifs en jeu dans la production et la compréhension du langage) and the Fondation pour la Recherche Médicale.

Correspondence concerning this article should be addressed to Stanislas Dehaene, Laboratoire de Sciences Cognitives et Psycholinguistique, 54 Bd Raspail, F-75270 Paris Cédex 06, France.

appearance. In our view, the real issue at stake then becomes the existence of such an encoding device and the level at which the comparison takes place: In propositional models, no computation of magnitude is necessary, only symbols are compared; in contrast, analogical models assume that comparison uses an internal representation of magnitude obtained after some preprocessing of the symbolic input.

Thus rephrased, the analogical–propositional dichotomy becomes somewhat fuzzy. There are no generic analogical or propositional models to be opposed. Rather, models differ in the processing steps that are assumed to take place between the digital input and the motor response and in the stage of processing at which comparison is supposed to take place.

Comparing Two-Digit Numbers

As pointed out by Hinrichs et al. (1981), the comparison of two-digit numbers offers an opportunity to oppose two well-defined processing models. According to a *lexicographic model*, subjects would first extract only the decades digits of the two numbers and compare them; thereafter, they would resort to comparing the units digits only if the two decades digits are equal. Alternatively, according to a *holistic model*, the comparison would not take place at the digit level. Rather, the symbolic input would be processed into a representation of the magnitudes of the two numbers; only then would the comparison take place.

The two models yield diverging predictions about the effect of units on reaction times (RTs). According to the holistic model, units should have a significant influence within decades, because they contribute to the difference in magnitude between the target and the standard. In contrast, according to the lexicographic model, units should have no effect on RTs when the two numbers belong to separate decades; in that case, only the distance between the decades digits of the two numbers should matter.

To examine these diverging predictions, Hinrichs et al. (1981) used a classification task: Target numbers between 11 and 99, presented serially every 4 s, had to be compared to a fixed standard of 55. Reaction times to determine whether the target was smaller or larger than the standard were recorded via two response keys. To a large extent, the results can be said to support the holistic model. Reaction times were a quasisymmetrical logarithmic function of the distance separating the target and the standard (Log D function). The units had a significant influence within decades. However, there was a surprising symbolic effect that rendered the RT curve discontinuous contrary to what the holistic model predicts: Within the decade of the standard, response times were globally slower, so that two discontinuities in RTs appeared at the boundaries of the decade of the standard (between 49–50 and 59–60).

Hinrichs et al. (1981) proposed an explanation for the discontinuities within a hybrid model. They assumed that the discontinuities are the result of a shifting of the subject's attention from the target number compared as a whole to a comparison with the rightmost digit only. This shift would occur only when the decades digit is equal to the decades digit of the standard. Yet Hinrichs et al. (1981) were plainly conscious of the unsatisfactoriness of this ad hoc hypothesis. Questions remain open as to why the subject should stop the holistic comparison when the target is within the decade of the standard, and why precisely at the decade boundary. Recall that at the level at which holistic comparison is supposed to take place, only magnitudes are relevant, so the exact location of decades is supposedly not available anymore.

Of course, discontinuities and other symbolic effects can, in principle, be explained at the encoding stage of the holistic model. The transformation from a symbolic code to a magnitude estimate need not be fast and independent of the number processed. It may be that target numbers sharing the decades digit of the standard require a more thorough encoding than the other numbers in order to achieve the same precision in the representation of magnitude.

However, the interpretation of the results within a holistic model is not fully convincing. The literature contains several examples in which the lexicographic model is better supported by the data. Thus, the comparison of three- to six-digit numbers (Hinrichs, Berie, & Mosell, 1982; Poltrock & Schwartz, 1984) appears to depend on the symbolic appearance of the stimuli: The two variables governing comparison times are (a) whether the two numbers have the same number of digits or not and (b) the position of the first set of differing digits when the numbers are scanned from left to right. Even in two-digit number comparison, when the standard is 50 and the targets range from 40 to 60, the units digit of the target number does not influence comparison times (Hinrichs et al., 1981, Experiment 2). These observations run contrary to what a purely holistic model predicts.

The purpose of the following experiments is to reexamine the data on two-digit number comparison. Experiments 1, 2, and 3 are particularly directed at examining the conditions of appearance of discontinuities in two-digit number comparison and at determining whether these discontinuities are central to the comparison process or whether they arise only from a putative encoding stage in an otherwise holistic process. Experiment 4 further attempts to separate the lexicographic and the holistic models by using an asynchronous presentation of the decades and units digits of a two-digit target number.

Experiment 1

Hinrichs et al.'s (1981) findings of a mixture of holistic and symbolic effects in two-digit number comparison, which was predicted neither by holistic nor by symbolic models, are sufficiently surprising as to motivate a replication. A replication is also desirable because, in the following experiments, we will be using French subjects, while Hinrichs et al. used English subjects. The French labeling system for numbers differs from the English system in a number of ways. In particular, in French, some of the names for the decades do not derive from the names of the units. Thus while the name for sixty (*soixante*) derives from *six* (six), the name for seventy (*soixante-dix*) does not come from *sept* (seven); rather, it means literally "sixty plus ten." Irregular names are also given to eighty (*quatre-vingt*, i.e., four times twenty) and to ninety (*quatre-vingt-dix*, i.e., four times twenty plus ten).

In Experiment 1, we simply attempted to replicate the results of Hinrichs et al. (1981) by using subjects from a different linguistic background. The possible effect of linguistic notation on comparison times, and particularly on discontinuities, is further examined in Experiment 2.

Method

Subjects. Thirty-five French right-handed volunteers were tested individually. Their sex was not recorded. Their age varied from 20 to 40 years.

Procedure. Subjects were seated in a dark room at about 50 cm from a monochrome cathode-ray tube. They were told that two-digit numbers, distributed around 55, would appear on the screen. They were asked to press the right-hand response key if the number was larger than 55, or the left-hand key if the number was smaller than 55. Instructions emphasized the necessity to respond as fast as possible while keeping errors at a minimum.

The experiment was controlled by a Solar computer that measured reaction times with a ±5 ms accuracy. The stimuli were two-digit numbers, approximately 3 cm high, drawn on a Hewlett-Packard 1321A graphic screen. Each number was displayed for 2 s, followed by a blank screen for 2 s, so that stimuli were presented at a 4-s rate. All numbers from 11 to 99, except the standard 55, were presented. Numbers ranging from 41 to 69 were presented four times; numbers outside this interval were presented twice. A pseudorandom list was constituted, fulfilling the additional constraints that (a) the same number never be presented twice in a row, and (b) subjects never press the same key more than three times in a row. Half the subjects received the list in direct order, and half in reverse order. Before the beginning of the experiment, a training list of 10 numbers was presented. Data from these 10 trials were not included in the analyses. The experimental sessions lasted about half an hour, during which 242 numbers were presented.

Results

Reaction times from erroneous responses (1.3% of all responses) were not analyzed. More errors occurred when the target was close to the standard, the error rate increasing from 0.2% at the extremities to 6.8% in the fifties. A regression analysis of the percentage of errors with Log D, the natural logarithm of the absolute distance between the target number and the standard 55, yielded a significant correlation ($r = .83$, $p < .001$).

An analysis of variance (ANOVA) was performed on reaction times with target-standard distance and response type ("larger" or "smaller") as within-subjects conditions. Distance had an important effect, $F(43, 1462) = 38.9$, $p < .001$, and all polynomial contrasts up to degree 6 were significant, $p < .001$. Response type had no direct effect, as response times were globally identical for larger and smaller responses, $F(1, 34) = 0.69$. But the two conditions interacted, $F(43, 1462) = 3.47$, $p < .001$: Response times were identical for both types of response for targets far from the standard, but they were longer for smaller than for larger responses for targets close to the standard. This pattern of asymmetries permits the rejection of the Welford (1960) function

$$\mathrm{RT} \ \alpha \ \mathrm{Log} \frac{L}{L - S}, \qquad (1)$$

where L and S are respectively the larger and the smaller of the digits to be compared. Although this function correlates well with RTs ($r = .89$), it predicts a symmetrical function close to the standard as well as shorter RTs to smaller than to larger numbers away from the standard, a pattern opposite to what was observed.

Figure 1 (upper panel) shows the mean response time for each target number. Response time correlated well with Log D ($r = .91$, $p < .001$). In multiple regression, the three conditions—Log D, response type (a dummy variable: −1 if response was smaller, +1 if response was larger), and their product—were all significant, confirming the asymmetry of slopes of the distance effect for larger and smaller responses. To study the contribution of units to the distance effect, two multiple regressions were used. In the first one, two variables were included: *LogDiz*, giving the logarithm of the number of decades between target and standard; and *Dunit*, measuring the contribution of units to the target-standard distance.[1] Both variables were significant at $p < .01$, showing that units indeed influenced RTs. In the second multiple regression, RTs to targets outside the decade of the standard were regressed with Log D, response type, their product, and the variable Dunit. The absence of a significant contribution of this last variable ($p = .80$) shows that the unit effect totally reduces to a continuous effect of the target-standard distance, at least outside the decade of the standard.

Figure 2 (top panel) summarizes the influence of units on RTs. First, we subtracted from each target's response time the mean response time of the corresponding decade. Difference scores from targets ending with the same digit were then averaged across decades (excluding the fifties). Data from decades above and below the standard were joined by pairing units symmetrically with respect to 5: We paired 5 below standard with 5 above, 4 below with 6 above, and so forth. In this process, numbers ending with a zero cannot be paired. Conventionally, values from numbers ending with zero and larger than the standard were given the label 10 as "ones-digit." The resulting curve summarizing the effect of units within decades is shown in Figure 2 (top panel). In global linear regression, the estimated slope (2.57 ± 1.50) closely approximates the value of 3.02, which is predicted from the slope of the global Log D regression by supposing RTs follow a strictly continuous logarithmic curve. The proportion of variance accounted for (34%) can be increased up to 62% if the regression excludes data points 0 and 10 (corresponding to numbers ending with zero). The mean difference scores for these numbers differ significantly from the prediction of regression on points 1 to 9. They are not different from zero, which means that the mean response time to a number ending with zero is not different from the mean response time of the decade.

---

[1] Let D_s, U_s, D_t, and U_t be respectively the decades and units digits of the standard and the target. LogDiz is the logarithm of $1 + |D_s - D_t|$. Dunit equals zero for targets within the standard's decade (i.e., when $D_s = D_t$). Outside the standard's decade, Dunit equals $U_t - 4.5$ for targets smaller than the standard and $4.5 - U_t$ for targets larger than the standard.

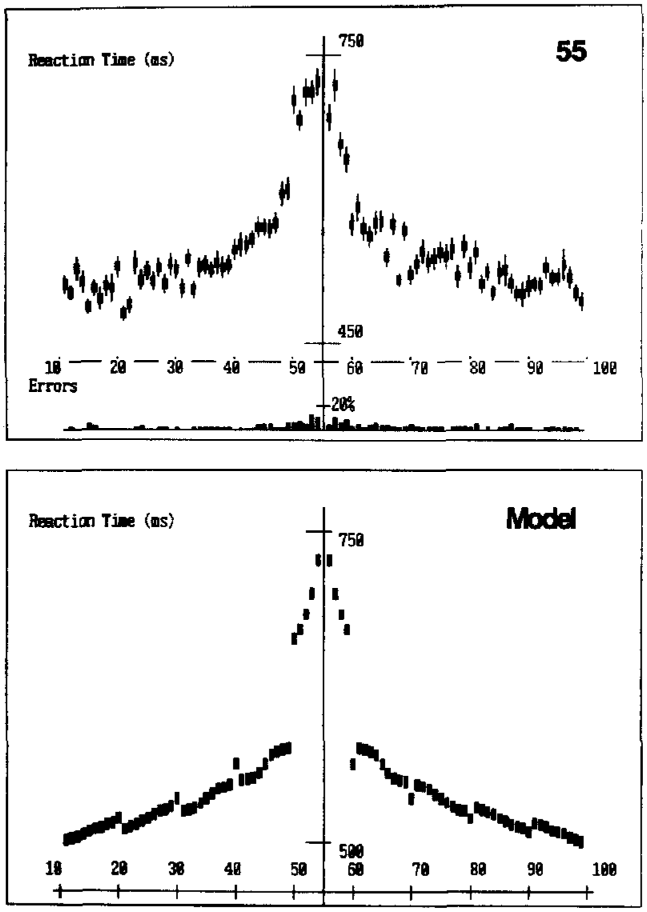

Figure 1. Reaction time and errors in two-digit number comparison with standard 55. Upper panel: Data obtained in Experiment 1. Lower panel: Reaction times predicted from a symbolic interference model.

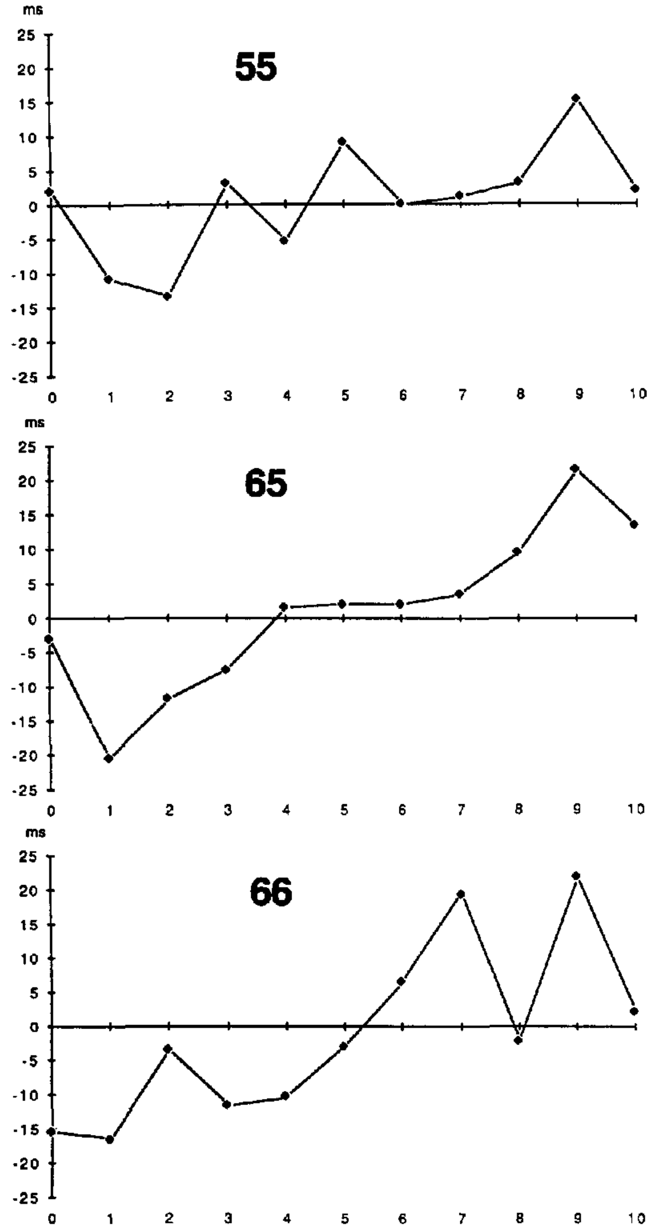

Figure 2. Influence of the ones-digit within a decade in Experiments 1, 2, and 3. (The mean reaction time [RT] of the corresponding decade was subtracted from each RT. For numbers smaller than the standard, these difference scores were averaged across numbers ending with the same digit. Data from numbers larger than the standard were also included for points 1–9 by pairing the ones-digits symmetrically with respect to 5 [4 with 6, 3 with 7, etc. . . .]. Finally, point 0 gives the average difference score for target numbers ending with 0 and smaller than the standard, and point 10 for target numbers ending with 0 and larger than the standard. In all experiments, difference scores increase with the abscissa, which shows that the ones-digits significantly contribute to the distance effect. Points 0 and 10 are deviant, meaning that RTs to numbers ending with zero are abnormally close to the mean RT of their decade [unit zero effect].)

Finally, we assessed the presence of discontinuities in the RT curve, which Hinrichs et al. (1981) found at the 49–50 and 59–60 decade boundaries. We compared the values of the observed differences between consecutive RTs on the curve with the theoretical values predicted by the Log D regression curve. Only the consecutive differences for 49–50 and 59–60 had significant z scores (respectively $p < .025$ and $p < .05$). Another test consisted of comparing the consecutive difference of interest, say, RT(49) − RT(50), with its neighbors. If a, b, c, d, e, and f are consecutive numbers and c/d is the point where a discontinuity is expected, one may compute for each subject the value of

$$D_{c,d} - \frac{D_{a,b} + D_{b,c} + D_{d,e} + D_{e,f}}{4}, \qquad (2)$$

where $D_{i,j} = \mathrm{RT}(i) - \mathrm{RT}(j)$.

This value should not be different from zero on a t test if the slope of the RT curve varies slowly at the point considered. Applied at 49–50, this test revealed a discontinuity: The observed difference of 84 ms was significantly larger than the mean consecutive difference of 13 ms observed around 49–

50 ($p < .01$). At 59–60, the observed difference of 65 ms was not far from significance against the mean background difference of 22 ms ($p = .138$ on a two-tailed t test; $p = .096$ on a nonparametric ranks test).

On these grounds, the RTs were refitted with a symmetrical Log D curve outside the fifties, and two separate Log D curves respectively for numbers in the 50–54 and the 56–59 intervals. This description of the data significantly improved the proportion of variance accounted for (the standard error of estimate falls from 23.2 ms to 13.4 ms, $p < .01$). The slopes of these three regression lines did not differ, but the intercepts did: For numbers 50 to 54, the intercept was 30 ms higher than for numbers 56 to 59, which in turn was 60 ms higher than for the other numbers.

Discussion

The results of Experiment 1 with French subjects are in striking agreement with those of Hinrichs et al. (1981) for English subjects. Not only did reaction times follow a Log D curve with a significant influence of the ones-digits, but even features that seemed less reliable (the local asymmetry close to the standard, the discontinuities at the decade boundaries 49-50 and 59-60) were reproduced in every detail. This suggests that the representation of numbers used in comparison tasks is largely language-independent.

The finding of a continuous distance effect that extends over the ones-digits favors a holistic model of two-digit comparison. The slope of the units curve can be derived from the slope of the global Log D regression. This shows that for numbers outside the fifties, the effect on RTs of a switch from the last number in a decade to the first one in the adjacent decade is similar to that of a change of one unit within a decade. This suggests that the comparison algorithm normally has no access to the symbolic representation of the numbers that it is compared with, but only has access to their magnitudes. However, as mentioned in the introduction, discontinuities around the decade of the standard do not readily fit the holistic explanation. Discontinuities will be examined again in Experiments 2 and 3.

Experiment 2

In Experiment 1, some subjects reported a tendency to verbalize the targets either mentally or in a low voice; they were especially likely to do so after an error. Could parallel verbal processing of the numbers be reflected in the reaction times to the main comparison task? Verbal processing, if it is involved, may result in spurious symbolic effects that would be superimposed over the main distance effect. Experiment 2 tests whether the observed discontinuities with 55 as standard can be attributed to an interference with linguistic procedures. The underlying hypothesis is that discontinuities in linguistic representations for numbers may be reflected in discontinuities in the RT curve.

In French, at some decade boundaries, the linguistic discontinuities are different from the discontinuities in Arabic representations of numbers. For example, between *soixante-neuf* (69) and *soixante-dix* (70) there is no real linguistic discontinuity, whereas in Arabic numerals the decades digit suddenly changes. In Experiment 2, we chose 65 as the standard of comparison. This way, the 69–70 boundary is made to coincide with one of the decade boundaries of the standard. If the discontinuities observed with 55 are really dependent on the visual appearance of the number and its subsequent treatment in the comparison module, then discontinuities should still be observed at 59–60 and 69–70. However, if they are the consequence of an interaction with a linguistic module, discontinuities should appear at 59–60 but not at 69–70. Finally, there is the possibility that discontinuities are a consequence of the choice of the standard itself. Some property of the number 55—for example, the repetition of the digit 5—may draw subjects' attention to numbers in the fifties. In that case, discontinuities are an artifact of the comparison with 55, and should simply disappear with 65. These alternatives are evaluated here.

Experiment 2 also incidentally investigates the influence of the spatial organization of the responses on RTs. In the experiment by Hinrichs et al. (1981), subjects were randomly distributed with respect to the "response-side" factor: Half the subjects answered with the right hand when the target was larger than the standard, and half when it was smaller. In our Experiment 1, all subjects answered "larger" with the right hand. In Experiment 2, two groups are compared: one responds "larger-right" and the other "larger-left."

Method

Task. Instructions were similar to those in Experiment 1, except for a new standard of comparison fixed at 65. One group of subjects had to respond by pressing the right-hand key with the right hand when the target number was larger than the standard of 65 (larger-right or LR group). The second group (larger-left or LL group) responded "larger" by pressing the left-hand key with the left hand.

Subjects. Forty-two French students who had not participated in previous comparison experiments were tested individually. Their ages ranged from 16 to 25 years ($M = 20$). Twenty-two students (13 men and 9 women), among whom 2 were left-handers, served as subjects in the LR group, and 20 students (13 men and 7 women), among whom 5 were left-handers, served as subjects in the LL group.

Procedure. Target numbers ranged from 31 to 99. Each number was presented four times (except the standard 65). A total of 282 numbers were presented, including a practice list of 10 numbers. Four random lists were constituted with the same constraints as in Experiment 1. Subjects were exposed to one list randomly chosen among the four.

Results

Two subjects in the larger-right group were eliminated because of excessive errors (22% and 18%). Error rate for the remaining 20 subjects of each group did not exceed 10%. Here again, the number of errors decreased with distance from the standard and was highly correlated with Log D ($r = -.76$ for LR, $r = -.87$ for LL; $p < .001$ in both cases).

Response times from left- and right-handed subjects were pooled together after separate analyses revealed no clear influence of handedness. An ANOVA was performed, with target-standard distance and response type (larger or smaller) as within-subject conditions and response side (larger-right or larger-left) as a between-subjects condition. The first two conditions were significant at $p < .001$: Response times decreased with target-standard distance and were slightly longer for smaller responses. As in Experiment 1, the two conditions interacted, $F(33, 1254) = 2.32$, $p < .001$, because the asymmetry between larger and smaller responses was restricted to targets close to the standard (Figure 3, upper panel). Finally, response side almost reached significance, $F(1, 38) = 3.72$, $p < .10$, and interacted with distance, $F(33, 1254) = 1.62$, $p < .025$: Responses were slightly slower and the distance effect was more pronounced for the LL group than for the LR group.

Mean response times are plotted on Figure 3 as a function of the target number. RTs were again highly correlated with Log D. In a multiple regression, distance, response side, and also their product were found significant, confirming the asymmetry of the slopes of the distance effect for smaller and larger responses.

The influence of units again appeared significant. As in Experiment 1, two multiple regressions were used. In the first one, both LogDiz and Dunit (measuring the respective contributions of decades and units to the distance effect) were significant at $p < .002$. In the second multiple regression, performed only for targets outside the decade of the standard, the variable Log D was introduced, and consequently the variable Dunit lost its significance ($p = .51$); thus, the distance effect reduces to a smooth continuous decrease of RTs with distance. Figure 2 (middle panel) summarizes the effect of units in Experiment 2 obtained by the pairing procedure described earlier. The curve increases significantly with the units. For both groups of subjects, the slope is well predicted by the global regression with Log D (LR group: observed slope 3.80 ± 1.46, predicted 4.07; LL group: observed 4.56 ± 1.26, predicted 5.37). The unit zero effect is clearly visible: A better linear regression ($r^2 = 89\%$) is obtained when points 0 and 10 are excluded than with global regression ($r^2 = 73\%$). These points deviate significantly from the regression on points 1 to 9.

Last but not least, no discontinuities were apparent in the results. The only consecutive difference in RTs that was significantly above the value predicted from the Log D regression was for 71–72 ($p < .05$), but it did not correspond to a decade boundary. Three other consecutive differences deviated to a lesser degree ($p < .10$). At 49–50 and 50–51, this reflected the fact that responses to 50 were abnormally short. At 66–67, response times were almost equal, whereas the distance effect predicted a large difference. None of these differences, which were only weak, could be interpreted as discontinuities. The second way to study discontinuities—comparing the critical consecutive difference with the neighboring differences—confirms this diagnosis. Neither at 59–60 (critical difference = 21.7 ms; neighboring difference = 23.1 ms; $p = .96$) nor at 49–50 (critical difference = 65.1 ms; neighboring difference = 23.3 ms; $p = .16$ on a t test, $p = .096$ on a nonparametric ranks test) does a significant break in slope appear. Thus, although the variability was comparable with that in Experiment 1, no discontinuities emerged with 65 as the standard.

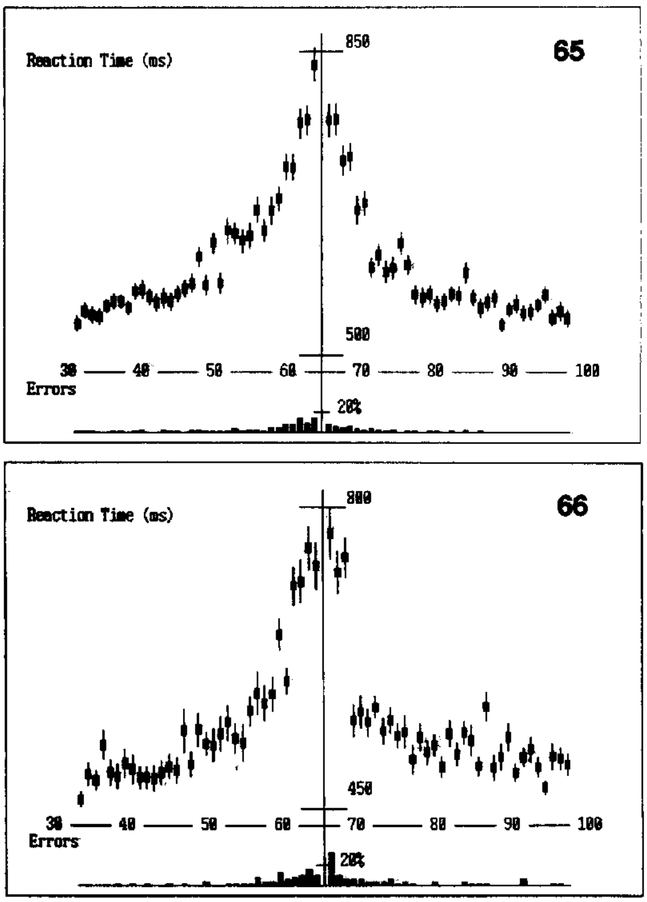

Figure 3. Reaction time and errors in two-digit number comparison with standard 65 (Experiment 2; upper panel) and with standard 66 (Experiment 3; lower panel).

Discussion

The results of Experiment 2 essentially replicated the main features of Experiment 1: the distance effect, the significant influence of units within decades, and the unit zero effect. As before, the observed asymmetries around the standard do not favor the Welford function. In addition, a response-side effect has been revealed: Subjects are slower, and the distance effect steeper, in the larger-left than in the larger-right condition. This will be discussed in greater detail in the General Discussion.

The most striking result is the disappearance of all discontinuities at the boundaries of the standard's decade. This finding was not predicted either by the two-stage model of Hinrichs et al. (1981) or by a model that assumes that linguistic properties affect numerical comparison. The former predicted two discontinuities (at 59–60 and 69–70), whereas the latter with French suggested a discontinuity at 59–60 but none at 69–70. It is unlikely, however, that discontinuities were not observed with 65 merely because of a lack of statistical power: A total of 40 subjects were tested, and the variability was lower than in the studies with a 55 standard (Experiment 1) or 66 standard (see Experiment 3).

Two-digit number comparison can thus be performed without discontinuities in RTs at the decade boundaries. This demonstrates that discontinuities—an effect of the symbolic appearance of numbers—are not central to the comparison process. This is not to say that their occurrence in the results of Experiment 1 was merely accidental, because the results with 55 faithfully replicated those of Hinrichs et al. (1981). One remaining logical possibility is that some standards, like 55, induce discontinuities in RTs while others (65) do not. What distinguishes these standards? One possibility is the presence of a repeated digit: Standards like 55 or 66 may induce subjects to pay more attention to numbers starting with the same digit, and this may produce a difference in processing within the decade of the standard. Experiment 3 tests this hypothesis by using a comparison task with standard 66.

Experiment 3

Method

Subjects. Twelve French subjects aged between 22 and 35 years were tested individually.

Procedure. The task was similar to the one used in Experiments 1 and 2, but with 66 for standard. All the subjects were tested in the larger-right condition. The experiment was controlled by a PC-compatible Olivetti M-24 computer with a standard monochrome screen. Numbers appeared at the center of the screen for 1,300 ms, followed by a 1,200-ms blank screen. Response keys were directly connected into the computer, which measured RTs with a 1-ms accuracy. Each number from 33 to 99, except 66, was presented five times in a pseudorandom list that was different for each subject and that did not allow for the same target twice in a row. Fifteen numbers served as an initial training list. The experiment lasted about 15 min. A total of 345 numbers were presented.

Results

Mean response times as a function of target appear in Figure 3 (lower panel). The correlation with Log D was very significant ($r = .91$, $p < .0001$). In a multiple regression, all three conditions—Log D, response side, and their product—were significant at $p < .05$, revealing the same asymmetries as before (the ANOVA showed a distance effect and Distance × Response Side interaction, $p < .001$). The unit effect was also reproduced: On the one hand, the variables LogDiz and Dunit, which measured the respective contribution of decades and units to the distance effect, were both significant in a multiple regression ($p < .03$); on the other hand, introducing the variable Log D in the multiple regression suppressed the significance of the factor Dunit. Thus, a holistic distance effect sufficed to account for the distance effect.

Figure 2 (bottom panel) summarizes the effect of units in Experiment 3. The units curve increased significantly with units ($r^2 = 57.8\%$) and the regression improved if points 0 and 10 were excluded ($r^2 = 65.0\%$). However, only point 10 deviated significantly from the regression on points 1–9.

To study the existence of discontinuities at points 59–60 and 69–70, consecutive differences in RTs were compared either with the values expected from the Log D regression or with the consecutive differences observed at neighboring points. At 69–70, both techniques revealed a huge discontinuity of 170 ms ($p < .001$). At 59–60, only the first technique revealed a marginal discontinuity of 69 ms ($p < .05$, unilateral). The other technique failed at $p = .34$. In addition, an unexpectedly short mean RT to 61 caused a 109-ms "discontinuity" to appear between 61 and 62 ($p < .01$).

Discussion

The results again reproduced the distance effect, the units effect, and the slight asymmetry of responses close to the standard. But the most striking difference between the comparison times with standard 65 and standard 66 is the reoccurrence of discontinuities in RTs at the decade boundaries of the standard. The discontinuity is extremely clear at 69–70, but less visible at 59–60. One may attribute this difference in magnitude to the eccentric position of the standard 66 in the decade: Because of the distance effect, responses are faster to target 60 than to target 69, leaving less room for a discontinuity to occur at 59–60 than at 69–70.

The fact that discontinuities can appear or disappear with only a minimal change in the magnitude of the standard suggests that discontinuities are an effect of the digital representation of the standard, not an effect of its magnitude. The particular standards tested in Experiments 1, 2, and 3 indicate that the repetition of a digit in the digital representation of the standard may induce a specific processing for target numbers starting with this particular digit.[2]

Evaluation of the Holistic Model

The holistic model is well supported by the data. Hinrichs et al. (1981) had previously demonstrated not only an effect of units on RTs, which was more compatible with the holistic model, but also the presence of discontinuities, which were more compatible with the lexicographic model. Our research showed that the effect of units can effectively be reproduced and that the appearance of discontinuities depends on the choice of the standard and is thus not central to the comparison process; it can be explained at the encoding stage of the holistic model. Hence, the holistic model can account for both effects, whereas the lexicographic model does not predict the effect of units within decades.

[2] It may be argued that we never directly tested the statistical significance of the influence of the standard on discontinuities. An ANOVA was performed on the data of Experiments 1, 2, and 3. Only RTs to numbers next to the boundary of the standard's decade were included. There was one between-groups factor (repetition of a digit in the standard or not) and two within-subjects factors: side of target number (smaller or larger than the standard) and location (within or outside the standard's decade). A significant interaction of location and group was found, $F(1, 85) = 4.13$, $p < .05$; thus, crossing the boundaries of the standard's decade indeed has a different effect in the two groups.

What about the other aspects of the response time curve? The asymmetry of RTs close to the standard fit nicely with a holistic model. Dehaene (1989) has shown that the asymmetries may be amplified, diminished, or even reversed, depending on the location of the target in the range of numbers tested. Thus in a comparison experiment with standard 75 and targets ranging from 20 to 99, the asymmetry increases: Close to the standard, smaller responses are much slower than larger responses. The asymmetry reverses with standard 35. The crucial variable determining the relative speed of the two responses is the location of the standard relative to the two extremes of targets tested.

In the present study the standard was always chosen at the numerical center of the range of targets. Thus, the persistence of an asymmetry in RTs may be tentatively interpreted as a nonlinearity in the internal representation of magnitude (Dehaene, 1989). The direction of the asymmetry suggests that for equal *numerical* distance, 55 would stand closer to 99 than to 11 on an *internal* continuum. Thus, the internal representation of magnitude would obey Fechner's law. The hypothesis of an internal compression of numerical magnitudes has been independently reached by several researchers on the basis of very different experimental paradigms (Banks & Hill, 1974; Curtis, Attneave, & Harrington, 1968; Curtis & Fox, 1969; Ekman, 1964; Ekman & Hosman, 1965; Rule, 1969; Schneider, Parker, Ostrosky, Stein, & Kanow, 1974). In the case of numerical comparison, Dehaene (1989) showed how Jamieson and Petrusic's (1975) reference point model—a holistic model—may be formalized with Fechner's law to accurately predict comparison data.

How do the other findings fit with the holistic model? We have outlined how the discontinuities may be explained at the encoding level. The last effect that remains to be accounted for is the unit zero effect. This refers to the finding that a number ending with a zero prompts the same reaction time as a number ending with five and belonging to the same decade, that is, a reaction time close to the mean of the decade. Note that this effect does not amount to a simple increase or decrease in RTs for numbers ending with zero. Rather, it represents a slowing down of smaller responses but an acceleration of larger responses to numbers ending with zero.

The unit zero effect is clearly a symbolic effect, and as such, within a holistic model, it can only be attributed to the encoding stage. Its complexity precludes any simple account; the only admittedly ad hoc hypothesis that we could formulate is the following: At the encoding stage, numbers ending with zero may initially be only grossly encoded in the correct decade and receive a more precise encoding after some delay. Evidently, reaction times to numbers ending with zero would then approach the mean RT of their decade. There are two indications that this hypothesis is not that farfetched. First, everyday use of numbers does not usually require access to the exact quantity that a number ending with zero represents. Rather, numbers ending with zero usually provide orders of magnitude rather than precise quantities. Thus when we say that "this car weighs 700 kilograms," we usually mean that the precision of the measure was ± 100 kg, not ± 1 kg. It is thus possible that the default representation for a number ending with zero is vague and that a slower, specialized procedure is used when more precision is needed. Slower and more difficult processing of numbers ending with zero is also likely given the lateness of appearance of the concept of zero in the otherwise scientifically advanced Babylonian, Indian, and Mayan civilizations (Ifrah, 1981) and given the difficulty of its acquisition by children.

In the comparison task, estimating an order of magnitude permits a response to all numbers ending with zero, with the exception of 50. The prediction that responses to 50 should be slower (because finer encoding is necessary) is upheld by the data, but it is indistinguishable from the discontinuity in RTs observed between 49 and 50.[3]

In summary, the holistic model can offer an explanation for all the effects that were observed in two-digit number comparison tasks. Yet, a clear weakness of the holistic hypothesis is the necessary attribution of all observed symbolic effects (discontinuities and the zero unit effect) to an ill-specified encoding stage, which is hardly accessible to experimentation. Were it not for the effect of units within decades, the holistic model would be uncalled for. Is there an alternative hypothesis that may account for the effect of units without assuming a holistic encoding? The *interference model*, examined in the next section, appears as an excellent potential alternative to the holistic one.

The Interference Model

Hinrichs et al. (1981) pointed out that a variant of lexicographic comparison might explain the influence of the units on RTs in a symbolic framework. Imagine that subjects simultaneously compare both the decades digits and the units digits of the operands. Outside the decade of the standard, the response would be selected according to the result of the decades comparison only. However, the result of the units comparison, if available, might interfere with the main task in a kind of "Stroop effect." Thus comparing 33 to 55 would be faster than comparing 37 to 55, because in 33, both digits are smaller than 5, whereas in 37, the ones-digit 7 is larger than 5 and the decades digit 3 is smaller. We call this model the *interference* model.

Obviously, in the model, interference can play a role only when the units comparison finishes before the decades comparison; as long as the result of the units comparison is not available, it cannot bias the subject in any direction. Assume that a constant time increment or decrement adds to the mean RT every time the units comparison reaches an end

[3] An intriguing possibility is that discontinuities are a mere consequence of the unit zero effect: The reaction time to 50 is increased to the mean RT in the fifties, giving rise to an artificially large jump between 49 and 50. Similarly, the reaction time to 60 is decreased to the mean RT in the sixties, so that the difference in RTs between 59 and 60 is unusually large. This explanation of discontinuities is rejected by the results of Experiment 2, where discontinuities are not observed while the unit zero effect is still present.

before the decades comparison.[4] In the mean RT then, the size of the interference effect should be proportional to the probability that the units comparison finishes before the decades comparison.

This property may explain the gradedness of the units curve as an effect of the relative speed of the decades and units comparisons. In agreement with the literature on the comparison of single digits, one may suppose that both the units and the decades comparisons obey the distance effect. Then within any given decade, the farther apart the units of the target and the standard, the more likely the units comparison is to terminate before the decades comparison, and thus the larger the interference effect. For example, in the comparison with standard 55, the amount of interference would be larger with 69 than with 64, because the comparison of 9 and 5 is much faster than the comparison of 4 and 5. The smooth increase of the units curves in Figure 2, which was found favorable for the holistic model, is thus perfectly compatible with a symbolic interference model.[5]

Without any supplementary assumptions, the interference model may also account for the unit zero effect. Data from the comparison of pairs of digits including zero (Parkman, 1971) show that comparisons involving zero are always very slow. Thus with targets ending with zero, the units comparison would always be slower than the decades comparison and hence would not have enough time to interfere. Units digits zero and five would both be neutral with respect to the decades comparison; this would explain why numbers ending with zero or with five yield comparable RTs that approach the mean RT of the decade.

To assess the quantitative fit of the interference model with the data, we formalized it with the following equations. Let D_s, U_s, D_t, and U_t be, respectively, the decades and units digits of the standard and the target. Outside the decade of the standard (for $D_s \neq D_t$), the following equation applies:

$$RT_{outside} = a - b \operatorname{Log}|D_s - D_t| + c \operatorname{sign}[(D_s - D_t)(U_t - U_s)] P_{s,t}. \quad (3)$$

The first two terms represent the time to compare the two decades digits. In the third term (the interference term), the sign function yields a minus sign (i.e., a decrease in RT) if the result of the units comparison is congruent with the result of the decades comparison and a plus sign (i.e., an increase in RT); otherwise, the value of the sign function is assumed to be zero (i.e., no interference) when the units digits of the target and the standard are equal.

The size of the interference effect in Equation (3) is assumed to be proportional to the probability $P_{s,t}$ that the units comparison finishes before the decades comparison. $P_{s,t}$ was computed with the assumption that both comparison times follow the same Gaussian distribution with standard deviation e and mean RT equal to $a - b \operatorname{Log}|x - y|$ (for the comparison of digit x with digit y). The only exceptions were (a) the comparison of zero with any other digit was assumed to be very long, so that in effect $P_{s,t}$ was zero whenever the standard or the target ended with zero, and (b) $P_{s,t}$ was also assumed to be zero when U_s was equal to U_t, that is, when the comparison of units was inconclusive.

A distinct equation had to be chosen for targets within the decade of the standard ($D_s = D_t$). The best results were obtained by assuming that in that case, subjects have to restart the units comparison from scratch. (The alternative assumption that subjects simply consult the result of the already performed—or almost achieved—comparison of units was incompatible with the observation of a distance effect within the decade of the standard.) This hypothesis yields the following equation:

$$RT_{within} = d + a - b \operatorname{Log}|U_s - U_t|, \quad (4)$$

where d is the average time it takes to decide whether the decades digits of the target and the standard are equal.

We attempted to fit the interference model to the results of Experiment 1 (comparison with 55). Of the five free parameters a to e, only parameter e cannot be fitted easily with multiple regression techniques. We chose as a reasonable estimate for e the average standard deviation of the observed RTs, averaged over all possible targets; at any rate, the fitness of the model was found very insensitive to values of e ranging from 50 to 300 ms. The remaining parameters a, b, c, and d were estimated by using a multiple regression. Figure 1 (lower panel) shows the predicted RT curve, which was obtained for $a = 563$, $b = 39.7$, $c = 14.6$, $d = 166$, and $e = 90$. The interference model accounted for 91.7% of the variance, as compared with 82.8% for the two-parameter regression with Log D.

Separating the Holistic and Interference Models

The interference model clearly captures the essential features of the comparison data: the distance effect, the influence of units, and the discontinuities. It is not clear how it would account for the effect of different standards on discontinuities or for the asymmetries in the RT curve. Despite these shortcomings, the interference model represents a plausible alternative to the holistic model.

Which experiments can be run to choose among the models? The interference model puts a strong emphasis on the relative processing speed of the units and decades digits. Altering this speed should have a predictable effect on RTs. Suppose we were able to present the decades and units digits of the target disjointly, with an arbitrary positive or negative asynchrony. According to the interference model, presenting the decades digit well before the ones-digit should reduce or suppress the influence of units on RTs. Conversely, presenting the ones-digit well before the decades digit should increase the influence of units; at the extreme, if the comparison of

---

[4] The assumption that interference results in a constant time increment or decrement is not central to the interference model. It was chosen merely for mathematical simplicity: The size of the interference effect becomes directly proportional to the probability that the units comparison finishes first. Under weaker hypotheses, only monotonicity is predicted.

[5] We thank S. Poltrock for pointing out to us the consequences of the relative speed hypothesis in the interference model.

units always ended before the comparison of decades, $P_{s,t}$ would reach the ceiling value of 1, and the units curve would become a step function, not a linear one. In both conditions of asynchrony, however, responses to targets should remain time locked to the appearance of the decades digit, which carries the information necessary for responding.

In conditions of decades–units onset asynchrony, what would the holistic model predict? Because the subjects are requested to respond as fast as possible, it is plausible that when presented with the decades digits first, the subjects would shift to a comparison of decades only and would not wait for the units digit to appear. This is all the more plausible the larger the onset asynchrony. Thus for decades-first trials, the predictions of the holistic model and the interference model may not differ.

This is however not the case for units-first trials. Units alone are not informative, so the holistic model predicts that the subjects will wait until the decades digit appears. Given that at that time the full two-digit number would be present, it is likely that it would then be treated holistically exactly as in the synchronous case.

In short, the critical difference between the interference and the holistic model is the following: On trials in which the ones-digit of the target would appear before the decades digit, the holistic model predicts no change with respect to the synchronous condition, whereas the interference model predicts an increase in the amplitude of the units curve, which may become similar to a step function. These diverging predictions are examined in Experiment 4.

Experiment 4

In Experiment 4, subjects were requested to compare two-digit numbers, ranging from 31 to 79, to a fixed standard 55. In each trial, the two digits of the target could either appear synchronously or one could lead the other by 50 ms. The critical experimental trials for the interference hypothesis were those in which the ones-digit preceded the decades digit (units-first trials), but we also included trials in the reversed condition (decades-first trials).

Obviously, the asynchronous conditions might disrupt the normal processing of two-digit numbers and thus might not be directly comparable with the synchronous condition. To limit alterations in processing, the subjects were not told of the three conditions. Furthermore, each digit presentation was preceded by a masking pattern, which largely prevented the subjects from noticing the variations in onset asynchrony.

Method

Subjects. Twenty subjects, 10 men and 10 women, were tested individually. Their ages ranged from 20 to 53 years. Two of the subjects were left-handed.

Instructions. The instructions were identical to Experiment 1 (comparison with 55 in the larger-right condition). No mention was made of the asynchrony in the onset of the decades and units digits in some trials. The subjects were simply told to respond "as soon as the number appeared."

Procedure. The stimuli were presented at the center of the plasma screen of a portable IBM-compatible Toshiba T-2100 computer. Target numbers ranged from 31 to 79. Each number was presented six times: twice with synchronous onset of the decades and units digits (synchronous trials), twice with units leading by 50 ms (units-first trials), and twice with decades leading by 50 ms (decades-first trials). To familiarize the subjects with the display, we presented five training trials as an initial separate block; three more training trials were provided, unknown to the subject, just before the actual experimental list of 288 trials. The training trials and the order of the list were randomized differently for each subject.

Each trial started with the presentation of a mask that consisted, at both the decades and the units location, of the superposition of the digits 0–9. After 300 ms, depending on the type of trial, one or two digits of the target number replaced the mask at the appropriate location. In synchronous trials, the full mask was erased and replaced by the two-digit target. In units-first trials, only the right half of the mask, standing at the units location, was replaced by the units digit of the target; the left half of the mask, standing at the decades location, was not erased. Finally, the reverse was done in decades-first trials: The right half of the mask was preserved, and the left half was replaced by the decades digit of the target. In asynchronous trials, the remaining digit of the target number appeared 50 ms later. Conventionally, response time was always measured from the onset of the decades digit of the target (300 ms after the mask appeared in synchronous and decades-first trials, but 350 ms after in units-first trials). The target number remained on for 1,000 ms, during which the subject's response was recorded. The display was then blanked for 1,500 ms before the next trial started.

Results

An ANOVA was performed on the correct response times, with target-standard distance, response type (larger or smaller), and trial type (synchronous, units-first, and decades-first) as within-subject conditions. Distance had a very significant effect, $F(23, 437) = 60.2$, $p < .001$. As in the earlier experiments in this study, larger responses were slightly faster than smaller responses, $F(1, 19) = 12.5$, $p < .005$; this effect was restricted to targets close to the standard, as revealed by a Distance × Response Type interaction, $F(23, 437) = 4.86$, $p < .001$. None of these effects interacted with trial type; the only effect of this variable was global: Units-first trials were significantly faster than synchronous trials, and synchronous trials were faster than decades-first trials (respectively 541, 554, and 568 ms; $p < .001$ for all pairwise differences).

A similar ANOVA was performed on the error rates, with distance introduced only as a dichotomous variable (within or outside the decade of standard). Only distance had a significant effect, $F(1, 19) = 15.1$, $p < .001$. In particular, either inside or outside the decade of the standard, the error rates did not differ significantly for synchronous trials (respectively 7.0% and 0.8%), units-first trials (8.6% and 1.1%), or decades-first trials (5.3% and 1.4%).

Figure 4 shows mean response time as a function of target for the three types of trials. The three curves do not differ significantly in slope or asymmetry pattern, and show similar discontinuities at the boundaries of the standard's decade. To assess this statistically, an ANOVA was performed on individual subjects' discontinuity scores, computed according to Equation (2), with discontinuity location (49–50 or 59–60) and

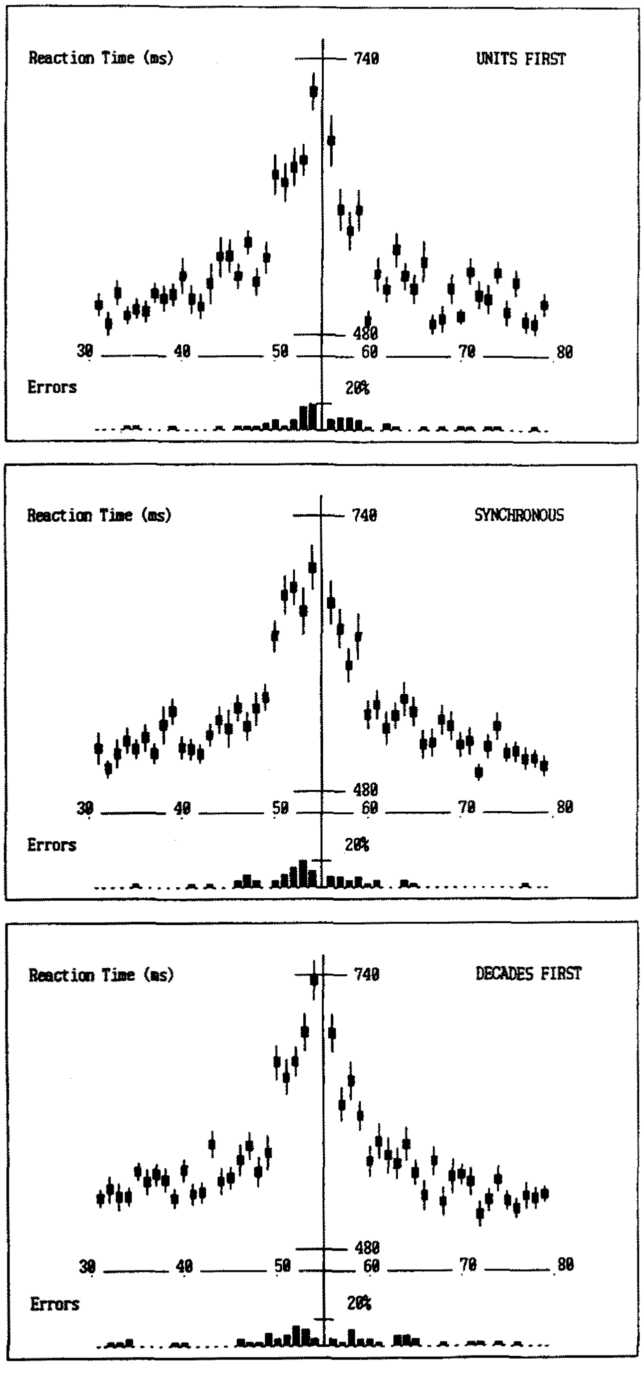

Figure 4. Reaction time and errors in two-digit number comparison with standard 55, plotted separately for synchronous, units-first, and decades-first trials (Experiment 4).

trial type (synchronous, units-first, or decades-first) as within-subject conditions. Neither the conditions nor their interaction had any significant effect.

Most important with respect to the purpose of the experiment is the possible effect of trial type on the units curves. Average curves for the effect of units were computed with the usual technique for each of the three trial types (Figure 5). The three regressions on the interval 1–9 (excluding target numbers ending with 0) were all significant, with comparable slopes (3.67 for both the synchronous and the units-first trials; 2.91 for decades-first trials). An ANOVA on individual subjects' units curves, with units and trial type as within-subject conditions, confirmed the lack of effect of trial type on the units curve: Trial type did not interact with a linear contrast for the effect of units, $F(2, 38) = 0.368$.

As mentioned earlier, depending on the parameters, the interference model could also predict a change in the shape of the units curve, which would become more sigmoidal in the units-first condition. Such an effect was not apparent (Figure 5). To assess it rigorously, we computed from the individual subject's unit curve the average effect of units 1–4 and of units 6–9. These scores were submitted to an ANOVA with interval (1–4 or 6–9) and trial type as within-subject conditions. Interval had a significant effect, $F(1, 19) = 64$, $p < .001$, which simply confirmed the influence of units on RTs. If the units curve was more sigmoidal for units-first trials, the difference between the scores for the 1–4 and 6–9 intervals should have been larger than in synchronous trials. Yet the interaction of trial type with interval was not significant, $F(2, 38) = 0.03$.

Discussion

Presenting the units and decades digits of a two-digit number asynchronously had no effect whatsoever on the amplitude of the influence of units on RTs. This goes against the prediction of the interference model. There was a nonsignificant trend for the units curve to be slightly flatter in the decades-first condition than in the synchronous conditions, but such a flattening is compatible with both the interference and the holistic model. As far as the critical difference is concerned (units-first vs. synchronous conditions), the slopes of the effect of units on RT were virtually identical, and the shapes did not differ significantly either.

It is possible that the asynchrony value of 50 ms was too small for an effect to emerge. To evaluate this possibility, we simulated the effect of asynchronous presentation in the interference model. First, Equations (3) and (4) were fitted to the data from synchronous trials. We obtained the following estimate of parameters: $a = 543$, $b = 29.4$, $c = 16.0$, $d = 130$, and $e = 90$. With these parameters, we then computed new values of $P_{s,t}$ (the probability that the units comparison finishes before the decades comparison) for each value of the target, in two new conditions: Either assuming that the decades comparison always started 50 ms before the units comparison or the reverse. In this way, we obtained predicted RT curves for the decades-first and the units-first conditions and computed the predicted units curve for each condition. The latter are shown in Figure 5. Clearly, even with only a 50-ms asynchrony, a large variation in the slopes of the units curve was predicted by the interference model; the ratios of the slopes in the decades-first, synchronous, and units-first conditions should have been about 1:2:3. Yet, no such variation of slopes was observed experimentally.

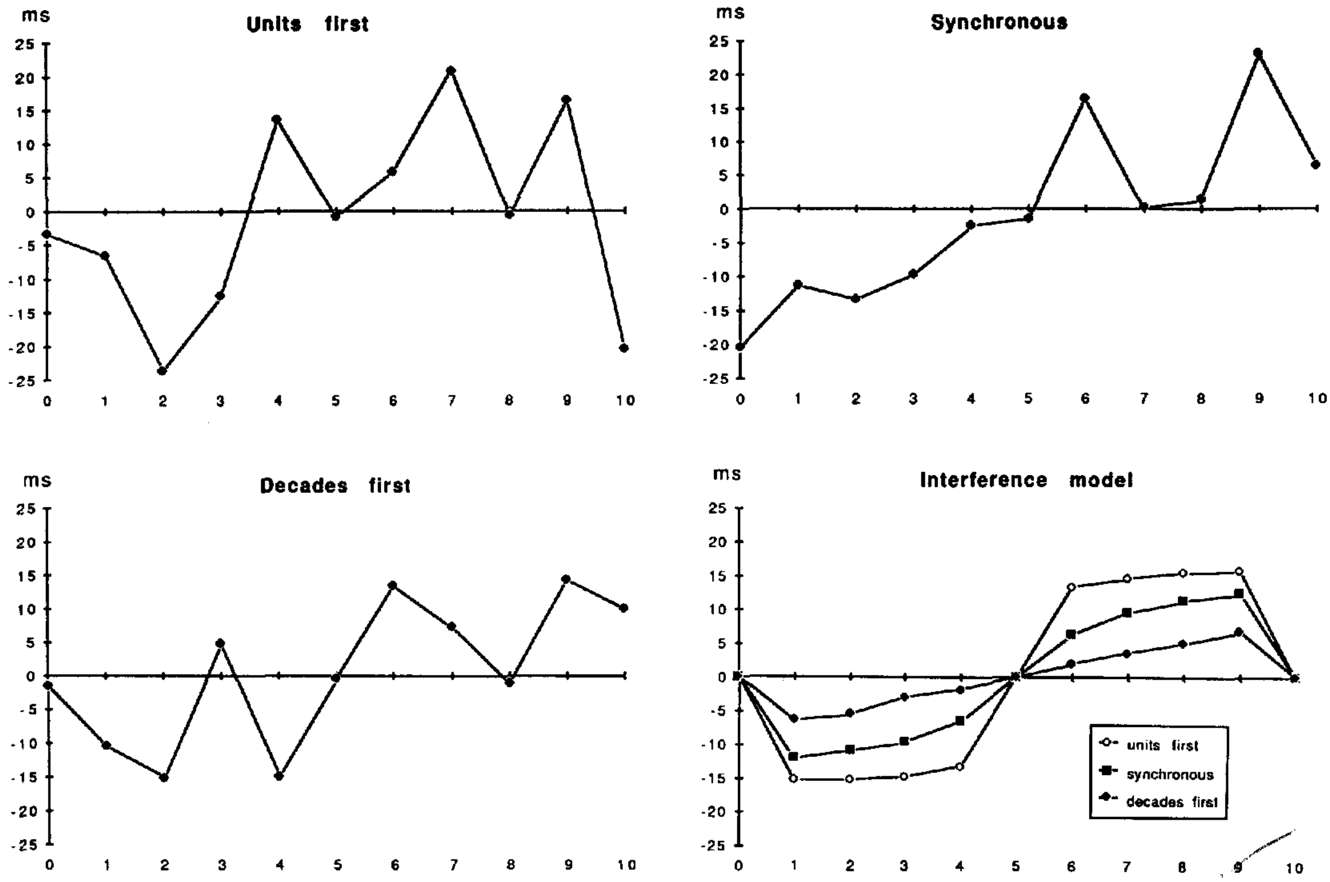

Figure 5. Observed units curves for synchronous, units-first, and decades-first trials, as compared with the predictions of the symbolic interference model. (The predicted variation in the slopes of the units curves is not supported by the data.)

Another indication that a 50-ms asynchrony is sufficient comes from the examination of mean response times and error rates (Table 1). Mean RTs were significantly affected by trial type ($p < .001$). The responses were not fully time locked to the onset of decades. Rather, they were faster (as measured from decades onset) when the units appeared first. This may be imputed to the instructions that required subjects to respond "as soon as the number appeared"; it resulted in slightly more errors. Conversely, the responses were slightly slower when the decades appeared first. This shows that subjects tended to wait for the units to appear even when the information provided by the decades was sufficient to respond; it may explain why the subjects were less error-prone in decades-first trials.

Table 1
Effect of Asynchrony Level on Mean Response Time and Accuracy in Experiment 4

	Response time (ms)			Error rate (%)	
Trial type	Globally	Within	Outside	Within	Outside
Units first	541	631	521	8.7	1.6
Synchronous	554	647	533	7.3	0.8
Decades first	568	660	547	5.3	1.4

Note. Conventionally, response times were measured starting from the onset of the decades digit. Within = within the decade of the standard (target numbers 50–59); Outside = outside the decade of the standard (target numbers 31–49 and 60–79).

Coherent as it may look, this pattern of performance is incompatible with the interference model, which predicted that trial type would affect performance only within the decade of the standard: Both error rates and RTs should have been larger for decades-first trials and smaller for units-first trials. Outside the decade of the standard, the facilitating or inhibiting effect of units should have canceled in the mean, so performance should have been independent of trial type. Two findings contradict these predictions. First, trial type affected RTs to the same degree for targets within or outside the fifties: There was no interaction of trial type with target-standard distance. Second, error rates tended not to vary in parallel with RTs. If anything, error rates in the fifties tended to be higher for units-first trials and lower for decades-first trials, a trend which goes against the above predictions.

General Discussion

Summary of Results

The four experiments described in this study extend the results of Hinrichs et al. (1981) on two-digit number comparison and help to clarify several points.

First, language does not influence the time it takes to compare numbers. In Experiment 1, two-digit numbers were compared with a standard of 55, and the pattern of RTs of French subjects was identical to the data reported by Hinrichs et al. (1981) with English subjects. The aim of Experiment 2 was to establish whether the discontinuities in RTs found at the boundaries of the decade of the standard originated from interference with a linguistic module. A special feature of the French numerical system, namely the absence of linguistic discontinuities at some decade boundaries, for example, *soixante-neuf* and *soixante-dix*, enabled us to test this hypothesis by choosing 65 as the standard in Experiment 2. However, discontinuities totally disappeared from the RTs.

Experiment 3 clarified this result: Numerical comparison with standard 66 made discontinuities reappear, suggesting that they originate from the repetition of a given digit in the standard. There is no need to suppose, as Hinrichs et al. (1981) have, that within the decade of the standard, subjects only compare the rightmost digits. Rather, discontinuities are not intrinsic to the comparison algorithm and appear as an accessory effect.

The influence of units on RTs, which was repeatedly found in Experiments 1, 2, and 3, is a natural consequence of holistic models, but it can also be explained by an interference model, which assumes a simultaneous but distinct treatment for decades and units. Experiment 4 tested more directly the holistic hypothesis by presenting the units and the decades digits of each target asynchronously. Even in this artificial condition, the shape of the reaction time curve was not affected. The influence of units on RTs was not differentially amplified or decreased, as was predicted by the interference model.

In summary, our findings are not compatible with a sequential comparison of decades and units (lexicographic model) or with two separate comparisons in parallel (interference model). Holistic processing, which assumes that decades and units are initially combined into a magnitude code before comparison per se, seems more compatible with the data.

Outline of a Holistic Model of Numerical Comparison

A schematic working hypothesis for numerical comparison tasks might thus be the following: First, the digital code of numbers is converted into an internal magnitude code on an analogical medium termed *number line*. This encoding stage is fast and independent of which particular number is coded. However, two minor effects may originate at this stage: (a) Numbers ending with zero are initially only grossly encoded within the correct decade and are given full precision (if necessary) only later, giving rise to the unit zero effect; and (b) when the standard has repeated digits, such as 55 or 66, encoding is slower for numbers starting with this digit, resulting in discontinuities at the boundaries of the standard's decade.

The second stage is comparison per se. This stage is assumed to be purely analogical, that is, without access to the digital appearance of the numbers. The only variables that play a role at this stage are analogical distances on the continuum. Dehaene (1989) has recently shown that comparison times can be adequately described by a model with two points of reference (Jamieson & Petrusic, 1975; Marks, 1972). These are two anchor locations on the continuum. The distance between these two anchors and the two operands is computed, and response time is a logarithmic function of the ratio of these distances. This function successfully models not only the distance effect, but also the magnitude (or "minimum") and congruity effects; the latter are explained as effects of distance from the reference points. Finally, the slight asymmetry, omnipresent in the above experiments, between larger and smaller responses to numbers close to the standard, appears in the model as a consequence of Fechnerian encoding: Although the standard is numerically centered in the range of target numbers, it is nevertheless internally closer to the larger reference point than to the smaller one, thus yielding a small asymmetry (for details see Dehaene, 1989).

Finally, in the last stage, the analogical comparison algorithm triggers a response buffer to make one of two discrete responses, namely, larger or smaller.

The holistic model just outlined does not suffice to account for all the results of numerical comparison. Two exceptions have to be considered. First, in two-digit number comparison with a standard ending with zero (e.g., 50), the subjects may notice that a comparison of decades only suffices and may thus respond without actually encoding the full magnitude of the two-digit target (Hinrichs et al., 1981); such strategic shifts are especially expected if the range of targets is narrow (e.g., targets in 40–59 with standard 50), in which case the comparison task reduces to a simple visual discrimination of the digits 4 and 5. Second, it is clear that the comparison of very large numbers, with more than three digits, is performed in a lexicographic fashion (Poltrock & Schwartz, 1984). Thus, fast encoding of magnitude may be restricted to small numbers with less than two digits, with which an educated adult is reasonably familiar.

Pointers Toward Further Research

The hypothesis that holistic comparison depends on familiarity with the range of targets has two important consequences. First, it should be possible to demonstrate (a) lexicographic comparison of two-digit numbers in children who are a priori not familiar with them, and (b) a progressive shift toward holistic comparison in the course of development. Second, one can predict that training adult subjects intensively in the comparison task should not disrupt holistic processing. Many models, in particular models that assume the progressive automatization of a direct mapping from the symbolic representation to the appropriate response, would predict just the contrary—that training should progressively suppress the effect of units and the distance effect.

To examine these predictions, we ran a pilot experiment and trained 3 subjects in the comparison with 55 condition, with numbers ranging from 11 to 99. The details of the method and the results appear in the Appendix. To summarize, the mean RTs decreased in the course of training, but the correlation coefficients of RTs with Log D, which measure the distance effect, remained fairly constant. Furthermore, the effect of units remained significant even after training. By contrast, the discontinuities at the boundary of the decade of the standard and the unit zero effect, both of which were effects of the symbolic appearance of the stimuli, weakened or disappeared with training. These results are only indicative, because we could not find devoted subjects who would volunteer for longer testing. However, we feel confident that the distance effect and the effect of units were extremely stable with training.

The pilot experiment also confirmed an important and as yet undiscussed phenomenon. In Experiment 2, we had found that the group of subjects who responded "larger" with the right-hand button were faster than the group of subjects who responded "larger" with the left-hand button. The pilot training experiment replicated this finding within single subjects: All three subjects were trained alternatively in the larger-right and larger-left paradigms, and all of them showed a significant saw-toothed pattern of response times with training. Initial performance was affected by as much as 100 ms by the side of response; that is, whether the larger response was assigned to the right-hand key or the left-hand key. This finding clearly classifies the response-side effect as a major, yet unexplored, effect of numerical comparison tasks.

The response-side effect can be described as a kind of association between "large" and "right." In the pilot training experiment, inconsistent mapping of response codes onto response sides had the effect of progressively reducing this initial association. But what caused such a pairing of two unrelated semantic concepts? The results from a small group of left-handers in Experiment 2 suggested that the effect was not linked to handedness: All 5 left-handers in the larger-left group had longer mean reaction times and steeper slopes of regression with Log D than the other 2 left-handers in the larger-right group. Even with very few subjects, the difference almost reached significance (mean reaction times: 588 ms for LR vs. 633 ms for LL, $p = .12$ one-tailed; slopes of regression with Log D: $p < .05$, one-tailed).

If handedness can be eliminated—this would clearly require a larger sample of subjects—we are then left with the following few possibilities. Perhaps the direction of the subjects' writing system (from left to right in French) is the cause of the effect. When a series of numbers is written, larger numbers appear to the right of smaller numbers. However, it is also true that for written numbers, hundreds appear to the left of decades, which, in turn, are located to the left of units. Furthermore, an informal questioning of Arab subjects, who write from right to left, shows that they have the same intuition as the French: They prefer to "see" larger numbers on the right. Another factor might be the degree of scientific education of the subjects. In mathematics and other sciences, graphs are plotted with smaller numbers in the lower left corner. This habit may have shaped the association of large numbers with the right. A last intriguing possibility is that this association is indeed innate. Testing children may allow us to assess this hypothesis. At any rate, the response-side effect deserves exploring. Documenting the psychological relations between numerical and spatial concepts may help to better formulate and test experimentally the hypothesis of an analogical representation of numbers.

References

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

Banks, W. P., & Hill, D. K. (1974). The apparent magnitude of number scaled by random production. *Journal of Experimental Psychology* (Monograph Series), *102*, 353–376.

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

Buckley, P. B., & Gilman, C. B. (1974). Comparison of digits and dot patterns. *Journal of Experimental Psychology, 103*, 1131–1136.

Curtis, D. W., Attneave, F., & Harrington, T. L. (1968). A test of a two-stage model of magnitude judgment. *Perception & Psychophysics, 3*, 25–31.

Curtis, D. W., & Fox, B. E. (1969). Direct quantitative judgments of sums and a two-stage model for psychophysical judgments. *Perception & Psychophysics, 5*, 89–93.

Dehaene, S. (1989). The psychophysics of numerical comparison: A reexamination of apparently incompatible data. *Perception & Psychophysics, 45*, 557–566.

Ekman, G. (1964). Is the power law a special case of Fechner's law? *Perceptual and Motor Skills, 19*, 730.

Ekman, G., & Hosman, B. (1965). Note on subjective scales of numbers. *Perceptual and Motor Skills, 21*, 101–102.

Hinrichs, J. V., Berie, J. L., & Mosell, M. K. (1982). Place information in multidigit number comparison. *Memory & Cognition, 10*, 487–495.

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

Holyoak, K. J. (1977). The form of analog size information in memory. *Cognitive Psychology, 9*, 31–51.

Ifrah, G. (1981). *Histoire universelle des chiffres* [The universal history of numbers]. Paris, France: Seghers.

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

Johnson, D. M. (1939). Confidence and speed in the two-category judgment. *Archives of Psychology, 241*, 1–52.

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

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

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

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

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

Poltrock, S. E. (1989). A random walk model of digit comparison. *Journal of Mathematical Psychology, 33*, 131–162.

Poltrock, S. E., & Schwartz, D. R. (1984). Comparative judgments of

multidigit numbers. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 10*, 32–45.
Restle, F. (1970). Speed of adding and comparing numbers. *Journal of Experimental Psychology, 83*, 274–278.
Rule, S. J. (1969). Equal discriminability scale of number. *Journal of Experimental Psychology, 79*, 35–38.
Schneider, B., Parker, S., Ostrosky, D., Stein, D., & Kanow, G. (1974). A scale for the psychological magnitude of numbers. *Perception & Psychophysics, 16*, 43–46.
Sekuler, R., & Mierkiewicz, D. (1977). Children's judgments of numerical inequality. *Child Development, 48*, 630–633.
Sekuler, R., Rubin, E., & Armstrong, R. (1971). Processing numerical information: A choice time analysis. *Journal of Experimental Psychology, 90*, 75–80.
Welford, A. T. (1960). The measurement of sensory-motor performance: Survey and reappraisal of twelve years' progress. *Ergonomics, 3*, 189–230.
Woocher, F. D., Glass, A. L., & Holyoak, K. J. (1978). Positional discriminability in linear orderings. *Memory & Cognition, 6*, 165–173.

Appendix

A Pilot Training Experiment

Method

Subjects. Two women (SSF and RBB) and 1 man (JYD), all right-handed adults aged between 30 and 50 years and associated with the Laboratoire de Sciences Cognitives et Psycholinguistique, were tested twice a week for approximately 1 month. They had not been subjects in the previous experiments in this study, although they were aware of the obtained results.

Procedure. The subjects were tested with standard 55 alternatively in the larger-right and the larger-left paradigms. One subject (JYD) started with LR and was tested nine times. The other 2 subjects started with LL and were tested, respectively, seven (SSF) and four (RBB) times. The same hardware and temporal presentation as in Experiment 3 were used. Numbers from 21 to 89 (except 55) were presented four times each. A new random list was generated for each experiment, with the constraint that the same number never be presented twice in a row. Including the initial training list of 20 numbers, a total of 292 numbers were presented in a 15-min session.

Results

Table A1 gives the mean RTs and correlation coefficients of RTs with Log D, computed separately for each test. The coefficients ranged from .48 to .81. The correlation was always significant ($p < .001$) within a single subject in a single test. No tendency to drop or improve was visible even after nine tests.

Each session was divided in eight arbitrary intervals of 34 numbers; this gave us eight measures per subject per session and permitted us to perform repeated-measures ANOVAS separately for each subject. Mean RT appeared to be influenced by response side ($p < .01$) and by training ($p < .005$). Mean RTs smoothly decreased with amount of experience. This general decrease was accompanied by a decrease in the response-side effect, as shown by a significant interaction of response-side and training ($p < .001$ for JYD and RBB; $p = .24$ for SSF). However, for each subject, at the time training was stopped, the response-side effect was still present, as assessed by a significant t test on the mean RTs from the last two sessions.

Data from the last two trials of subjects JYD and SSF and from the last trial of subject RBB were combined to analyze performance after training. The usual multiple regression analysis with Log D, response side, and their product revealed the same asymmetries as in Experiment 1: RTs were slightly slower, and the distance effect was steeper, for smaller than for larger responses. There was only weak evidence of the presence of discontinuities at the fifties decade boundary: Points that deviated more than two standard errors of estimate from the regression curve, sorted according to the magnitude of the deviation, were points 50, 51, 53, 57, and 61, indicating a bad fit of the global Log D curve in the fifties. The differences for 50–51, 51–52, 53–54, 56–57, and 57–58 were significantly larger than expected at the .10 level. The differences of interest, 49–50 and 59–60, also deviated to a lesser degree. However, although discontinuities disappeared with training, one cannot exclude the possibility that the larger variability simply limited their significance.

We performed the same multiple regression analyses as before to study the influence of units. The two variables measuring the respective contributions of decades and units to the distance effect were both significant ($p < .0001$ and $p < .004$). However, when the Log D

Table A1
Mean Reaction Times (RTs) and Correlation Coefficients of RTs With Log D During Alternated Training in the Larger-Left (LL) and Larger-Right (LR) Tasks

	Task								
Subject	LR	LL	LR	LL	LR	LL	LR	LL	LR
					Mean RTs				
JYD	460	618	449	565	432	473	427	486	397
RBB[a]	—	556	441	454	427	—	—	—	—
SSF[b]	—	693	559	573	527	580	514	540	—
					Correlation coefficients				
JYD	−.73	−.48	−.78	−.72	−.63	−.61	−.81	−.72	−.78
RBB[a]	—	−.73	−.74	−.75	−.72	—	—	—	—
SSF[b]	—	−.68	−.74	−.71	−.75	−.65	−.66	−.68	—

[a] Subject RBB started in the LL task and could only be tested 4 times.
[b] Subject SSF started in the LL task and was tested 7 times.

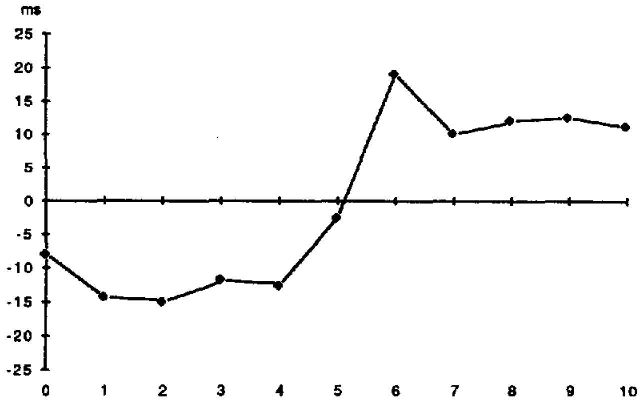

Figure A1. Influence of the ones-digit within a decade after training for 3 pilot subjects.

variable was entered (measuring the continuous target-standard distance), the part of the distance effect imputable to units still remained almost significant ($p = .127$ or .060 depending on whether a variable for the zero effect was introduced or not). Thus, the effect of units was slightly larger than expected from the global Log D curve. Figure A1 shows the influence of units within decades after training, computed with the procedure described earlier. The correlation with the units was very significant ($r^2 = 69\%$, $p < .002$). The slope of the regression was consistent with the estimate from the global Log D curve under the holistic hypothesis (predicted value: 3.85; observed value: 4.40 ± 2.88). The unit zero effect was weakened: Difference scores for points 0 and 10 differed from 0, and only the scores for point 0 significantly differed from the value predicted in regression with a step function.

Received May 10, 1988
Revision received June 30, 1989
Accepted August 25, 1989 ■