The Effect of External Representations On Numeric Tasks

Jiajie Zhang and Hongbin Wang

School of Health Information Sciences

University of Texas Health Science Center at Houston

Address Correspondence to:

Jiajie Zhang

School of Health Information Sciences

University of Texas Health Science Center at Houston

7000 Fannin, Suite 600

Houston, TX 77030

Phone: (713)500-3922

Fax: (713)500-3929

E-mail: Jiajie.Zhang@uth.tmc.edu

Running Title: External Representations in Number Comparison

ABSTRACT

This article explores the effect of external representations on numeric tasks. Through several minor modifications on the previously reported two-digit number comparison task, we obtained different results. Rather than a holistic comparison, we found parallel comparison. We argue that this difference was a reflection of different representational forms: the comparison was based on internal representations in previous studies but on external representations in our present study. This representational effect was discussed under a framework of distributed number representations. We propose that in numerical tasks involving external representations, numbers should be considered as distributed representations and the behavior in these tasks should be considered as the interactive processing of internal and external information through the interplay of perceptual and cognitive processes. We suggest that theories of number representations and process models of numerical tasks should consider external representations as an essential component.

Many numerical tasks involve external representations. In these tasks, people need to process information distributed across internal and external representations in an interactive manner (see Zhang & Norman, 1995; Zhang, 1997). For example, to multiply 735 by 278 using paper and pencil, we need to process not just the information in internal representations (e.g., the value of each individual symbol, the addition and multiplication tables, arithmetic procedures, etc.) but also the information in external representations (e.g., the visual and spatial properties of the symbols, the spatial relations of the partial products, etc.).

There are at least two different views on the roles of external representations in numerical tasks. One view is that external representations are merely input and stimuli to the internal mind. In this view, external representations first have to be re-represented as internal representations through some encoding processes. Only after this encoding stage does real numerical cognition take place: mathematical thinking processes operate upon the internal abstract number representations. This is the view reflected in many studies on numerical cognition (see Dehaene, 1992, 1993, 1997; Dehaene, Dehaene-Lambertz, & Cohen, 1998 for reviews).

A different view is that external representations need not be re-represented as internal representations in order to be involved in numerical tasks: they can directly activate perceptual processes and directly provide perceptual information that, in conjunction with the memorial information and cognitive processes provided by internal representations, determine the behavior in numerical tasks (e.g., Zhang & Norman, 1994, 1995; Zhang, 1997). Thus, in numerical tasks that involve external representations, the behavior is simply the integrative processing of the information perceived from external representations and that retrieved from internal representations through the interplay of perceptual and cognitive processes.

This article uses number comparison as an example to demonstrate the important roles of external representations in numerical tasks. We only make a few minor modifications on the two-digit number comparison task that has been previously published (Dehaene, Dupoux, & Mehler, 1990; Hinrichs, Yurko, & Hu, 1981). The results, however, are quite different. Their implications will be discussed under a framework of distributed number representations.

Number Comparison

Arabic numerals have two dimensions: a base dimension represented by the shapes of the ten digits (0, 1, ..., 9) and a power dimension represented by the positions of the digits (for details, see Zhang & Norman, 1995). The base dimension is an internal representation because the numerical value of each digit (arbitrary shape) has to be memorized, whereas the power dimension is an external representation because the position of a digit can be perceptually inspected. The information needed to compare the magnitudes of two one-digit Arabic numerals is solely in internal representations because the comparison is only on the base dimension that is internal. In contrast, the information needed to compare two multidigit Arabic numerals is not only in internal representations but also in external representations: the values of individual digits on the base dimension are stored in memory but the positions of individual digits on the power dimension are perceptually available in external representations. Thus, due to the different information sources, there might be different comparison mechanisms for one-digit and multidigit Arabic numerals.

For one-digit Arabic numerals, the time to make magnitude comparisons (larger or smaller) decreases linearly with the logarithm of the numerical distance between them (Moyer & Landauer, 1967; for reviews, see Banks, 1977; Moyer & Dumais, 1978). For example, it is faster to compare 1 and 9 than 8 and 9. This *distance effect* remains even when the comparison is based on English written number words (Foltz, Poltrock, & Potts, 1984) and Japanese kanji and kana numerals (Takahashi & Green, 1983). In addition, the distance effect resembles that found for physical stimuli such as line lengths and dot patterns (e.g., Buckley & Gillman, 1974). These results suggest that different types of numerals may have a common internal representation, which is similar to a physical continuum such as a format-independent *line-like* analog representation (e.g., Dehaene, Dupoux, and Mehler, 1990; Dehaene, 1992, 1997; Gallistel & Gelman, 1992; Restle, 1970).

For multidigit Arabic numerals, there are at least three different models of comparison (Hinrichs, Yurko, Hu, 1981). Let us use the reaction times of comparing a series of target numerals (e.g., 11 to 54 and 56 to 99) with a fixed standard numeral (e.g., 55) to describe the three models. The first model is sequential comparison: the two multidigit Arabic numerals are compared digit by digit sequentially. In this case, only the highest digits should affect the comparison unless they are not sufficient for making decisions. For example, the larger numeral of 29 and 55 can be decided by comparing their decade digits (2 and 5) alone. Thus, the reaction time (RT) for comparing 29 with 55 should be identical to the RT for comparing 21 with 55. Equation 1 is one likely sequential model for two-digit numerals, which is shown graphically in Figure 1A. D_s and D_t are the decade digits and U_s and U_t are the unit digits of the standard numeral and the target numeral, respectively; and *a*, *b*, and *d* are the parameters. When the decade digits of the standard and the target are different ($D_s \neq D_t$), the RTs of the targets within a decade are identical to each other and the RTs for different decades decrease with the logarithm of the numeric distance between the decade digits of the standard and the target. When the decade digits of the standard and the target are identical ($D_s = D_t$), that is, the target is in the same decade as the standard, the RTs decrease with the logarithm of the numerical distance of the unit digits of the standard and the target. For both cases, the comparison is in fact based on a sequential self-terminated one-digit comparison process.

$$\text{Sequential Model}: \begin{cases} RT = a - b \cdot \ln|D_s - D_t| & \text{if } D_s \neq D_t \\ RT = d - b \cdot \ln|U_s - U_t| & \text{if } U_s \neq U_t \end{cases} \qquad \text{Eq. 1}$$

The second model is parallel comparison: the lower digits may facilitate or interfere with the comparison of higher digits (i.e., a Stroop-like effect). For example, the unit digit 9 in 29 may increase the RT for comparing 29 with 55 and the unit digit 1 in 31 may decrease the RT for comparing 31 with 55. Equation 2 is one likely parallel model for two-digit numerals, which is shown graphically in Figure 1B. This model is identical to the sequential model shown in Equation 1 except that there is an extra term for the Stroop-like effect for targets whose decade digits are different from that of the standard. This Stroop-like term is a simple linear function of the numeric distance between the unit digits of the standard and the target. For targets smaller than the standard, this term has a facilitation effect for targets whose unit digits are smaller than that of the standard and an interference effect for targets whose unit digits are larger than that of the standard. For targets larger than the standard, the effects are reversed, as specified by the sign function in the equation.

$$\text{Parallel Model: } \begin{cases} RT = a - b \cdot \ln|D_s - D_t| + c \cdot (U_t - U_s) \cdot sign(D_s - D_t) & \text{if } D_s \neq D_t \\ RT = d - b \cdot \ln|U_s - U_t| & \text{if } U_s \neq U_t \end{cases} \quad \text{Eq. 2}$$

The third model is holistic comparison: multidigit numerals are first encoded as integrated representations and then their whole numerical values are compared. In this case, only the absolute numerical values should matter. Therefore, the RTs for comparing target numerals 11, 32, .., and 54 with the standard 55 should be a smooth decreasing function of the numerical distances between the target numerals and the standard. Equation 3 is one likely holistic model for two-digit numerals, which is shown graphically in Figure 1C. It shows that the unit and decade digits of the target are first integrated as a whole numerical value, which is then compared with the whole numerical value integrated from the decade and unit digits of the standard.

$$\text{Holistic Model: } RT = a - b \cdot \ln|(10 \cdot D_s + U_s) - (10 \cdot D_t + U_t)| \quad \text{Eq. 3}$$

Insert Figure 1 about here

Empirical studies have revealed a discrepancy between two-digit comparison and higher multidigit (3 or more) comparison: two-digit comparison is often holistic (Dehaene, Dupoux, & Mehler, 1990; Hinrichs, Yurko, & Hu, 1981) whereas higher multidigit comparison is sequential (Hinrichs, Berie, & Mosell, 1982; Poltrock & Schwartz, 1984). The holistic comparison for two-digit Arabic numerals is counter-intuitive because the decades of two two-digit Arabic numerals, if they are different, are sufficient to decide which numeral is larger or smaller. This finding has often been cited as evidence for a holistic analog internal representation of Arabic numerals. More recently, however, Nuerk and colleagues presented evidence that challenged this holistic representation idea (Nuerk, Weger, & Willmes, 2001). By asking participants to compare the magnitude of two vertically aligned two-digit Arabic numerals, they found a unit-decade-compatibility effect: the comparison was faster when the unit and decade digits agreed with each other (e.g., 6>5 and 7>2 when comparing 67 with 52) than when they did not. They suggested that there existed decade breaks in the mental number line and that in two-digit number comparison both the decade and unit digits might affect the two-digit number comparison performance.

In this article, we report three experiments we conducted to examine in more detail whether the comparison of two-digit Arabic numerals is sequential, parallel, or holistic. Through several minor modifications on the previously reported two-digit number comparison tasks, our results illustrate the important role external representations play in multi-digit numerical tasks.

Overview of Experiments

In the experiments reported by Dehaene, Dupoux, and Mehler (1990) and Hinrichs, Yurko, and Hu (1981), a standard (e.g., 55) was always held in memory and only the target numerals (e.g., 11 to 99 except 55) were presented on an external display and judged whether smaller or larger than the standard. In this case, the comparison was between a preprocessed internal representation of the standard and an external representation of the target. In the experiment by Nuerk et al. (2001), the two to-be-compared two-digit numerals were presented simultaneously above each other. However, there was not a standard: the numerals were chosen from a set of 240 pairs from 21 to 98. In the first and second experiments of our current study, we made three changes. First, following Dehaene et al. (1990) and Hinrichs et al. (1981), we adopted the target-standard paradigm. However, instead of one standard, we used two standards (55 and 65) as a within-subject factor to prevent participants from preprocessing a specific standard and transforming it into an internal representation. Second, similar to Nuerk et al. (2001), we presented both numerals simultaneously on the computer display such that the comparison was between two external representations. However, instead of vertically aligning the two numerals, we presented them horizontally (side-by-side). This manipulation prevented a possible confounding column-wise comparison in Nuerk et al. experiment. Third, instead of reporting whether a target was smaller or larger than the standard, participants only decided which of the two numerals was larger (Experiment 1) or smaller (Experiment 2). Such procedural changes were important for testing our hypothesis: when both numerals are presented simultaneously, the comparison is not just based on internal but also on external information, and the concurrent processing of which might generate a different pattern of behavior from that found with an internal and an external representation. To further test our hypothesis, we also conducted a control experiment (Experiment 3) in which the comparison was between an external representation and an internal representation.

To distinguish among the three comparison models we have focused on the following four effects. The first is the *unit effect*: the target numerals within a decade (e.g., 21-29) have different RTs from each other. The second is the *decade effect*: the target numerals in a decade (e.g., 21-29) have a different average RT than that in a different decade (e.g., 31-39). The third is the *discontinuity effect*: there is a sharp change in RTs for target numerals across a decade boundary (e.g., the RT difference between 29 and 30 is bigger than that between 28 and 29). The fourth is the *Stroop-like effect*: the unit digits may interfere or facilitate the comparison of the two-digit numerals. One difficulty with the observation of a Stroop-like effect is that the values of unit digits and the absolute distances from the target numerals to the standard are always confounded. For example, a faster RT to compare 21 with 55 than 29 with 55 can be either due to the facilitation of the 1 in 21 coupled with the interference of the 9 in 29, or the longer distance between 21 and 55 than between 29 and 55, or a combination of both. However, if we can observe a *reverse distance effect* across decade boundaries, then we can positively identify a Stroop-like effect. For example, the RTs for 26-29 might be slower than those for 31-34 even if 26-29 are farther away from 55 than 31-34 (see Figure 1B). Therefore, rather than testing a general Stroop-like effect, we test the reverse distance effect in the experiments, which is a stronger Stroop-like effect.

Based on the characteristics of these four effects, we can evaluate the three hypothetical models. The presence (+) or absence (-) of a specific effect for each model is shown in Table 1. First, if the comparison is sequential, then the decade effect and the discontinuity effect should be present but the unit effect and the reverse distance effect (Stroop-like effect) should not be present. Thus, if either the unit effect or the reverse distance effect is present, then sequential comparison can be rejected. Second, if the comparison is parallel, then the decade effect and the unit effect should be present and the discontinuity effect may or may not be present. The Stroop-like effect should also be present, but the stronger reverse distance effect may or may not be present. Because the Stroop-like effect is the defining feature of parallel comparison, its presence is sufficient for accepting the parallel model. Thus, if the reverse distance effect is present, then parallel comparison can be accepted. Third, if the comparison is holistic, then the unit effect and the decade effect should be present but the discontinuity effect and the reverse distance effect should not be present. Thus, if either the discontinuity effect or the reverse distance is present, then holistic comparison can be rejected.

Insert Table 1 about here

EXPERIMENT 1

In this experiment, the comparison is between two two-digit Arabic numerals that are presented side-by-side. Specifically, target numerals 11 to 54 and 56 to 99 were compared with standard 55, and target numerals 31 to 64 and 66 to 99 were compared with standard 65. The task was to decide which of two numerals was larger. We predict that unlike the comparison between an internal and an external two-digit Arabic numeral, the comparison under the current condition is no longer holistic.

Method

Participants

The participants were 32 undergraduate students in introductory psychology courses at The Ohio State University, who participated in the experiment to earn course credit. They were all native English speakers.

Design and Procedure

The participants were seated in about 40 cm from a Macintosh computer. They were told that two Arabic numerals would appear on the screen simultaneously, one on the left and one on the right side of a fixation point. They were asked to press the left key ('z') or the right key ('m') as quickly and accurately as possible depending on whether the numeral on the left or the one on the right side is larger. Each pair of numerals were presented for 2 s, preceded by a fixation point (a '+' sign) of 500 ms and followed by a blank screen of 2 s. The two-digit Arabic numerals were in 24-point bold New York font (approximately 1.0 by 0.65 cm for each digit) and with an equal distance of 0.95 cm from the fixation point.

Arabic numerals 11 to 99 (target numerals) except 55 were compared with 55 and 31 to 99 (target numerals) except 65 were compared with 65. Thus, there were 156 target-standard pairs. Each target-standard pair was presented twice, with one trial having the standard on the left side and the other having the standard on the right side. Therefore, there were a total of 312 trials. These 312 trials were randomly into four blocks with 78 trials in each block. Each participant was presented 10 randomly generated pairs of two-digit numerals for practice, followed by the four blocks of experimental trials with one minute rests between blocks.

Results

For all analyses that follow, target-standard pairs with the standard on the left side were pooled with the same pairs with the standard on the right side. The average error rate was 2.7% for standard 65 and 2.0% for standard 55. Trials with errors and trials with RTs that deviated from the mean for each target by more than three standard deviations were excluded from the analysis.

The average RTs for both standards are shown in Figure 2. For standard 65, an overall 2 (range: targets smaller vs larger than the standard) x 34 (magnitude distance) repeated-measure MANOVA showed a significant distance effect ($F(33, 693) = 16.99$, $p < 0.01$), a significant range effect ($F(1, 21) = 8.40$, $p < 0.01$), and a significant interaction ($F(33, 693) = 4.86$, $p < 0.01$). For standard 55, a similar 2 (range) x 44 (distance) MANOVA showed a significant distance effect ($F(43, 688) = 15.23$, $p < 0.01$), a marginally significant range effect ($F(1, 16) = 3.84$, $p < 0.08$), and a significant interaction ($F(43,688) = 2.25$, $p < 0.01$). While the distance effect was expected, the range effect was surprising, suggesting that participants might adopt different strategies in comparison for smaller and larger targets. Consequently, further analyses were carried out separately for targets below and above the standard.

Insert Figure 2 about here

Decade and unit effects. Several two-way (decades x units) MANOVAs on mean RTs were performed to test the decade and unit effects. Significant decade effects were found for both standards, with the targets either below the standards ($F(2, 50) = 60.75$, $p < 0.01$; and $F(3, 69) = 39.02$, $p < 0.01$, for standard 65 and 55 respectively) or above the standards ($F(2, 46) = 20.18$, $p < 0.01$; and $F(3, 60) = 74.77$, $p < 0.01$, for standard 65 and 55 respectively). However, significant unit effects were found for both standards only when the targets were below the standards ($F(8, 200) = 16.15$, $p < 0.01$; and $F(8, 184) = 8.19$, $p < 0.01$, for standard 65 and 55 respectively). When the targets were above the standards, we found significant unit effects for standard 55 ($F(8, 160) = 2.42$, $p < 0.05$) but not for standard 65. Overall, these results suggest that both decade and unit digits might affect the comparison performance. However, since the decade effect and the unit effect were often confounded with the distance effect (especially when the targets were below the standard), they should be interpreted with caution.

Discontinuity Effect. If there is a discontinuity at the boundary of two decades, there should be a sharp change in RT. Adopting a calculation similar to that used by Dehaene, Dupoux, and Mehler (1990), the change in RT across a decade boundary (e.g., RT69 - RT70) was compared with the averaged change in RT between adjacent numbers within each of the two adjacent decades (e.g., [(RT68 - RT69) + (RT70 - RT71)]/2). For standard 65, a series of ANOVAs showed significant discontinuity effects between all decades (smallest $F(1, 30) = 4.76$, $p < 0.05$) except of between 40s and 50s. For standard 55, we found a significant discontinuity effect at the boundary between 60s and 70s ($F(1, 30) = 6.75$, $p < 0.01$). Combined with above findings of significant decade effects, the occurrences of RT discontinuity at several decade boundaries suggest that decade digits did play a unique role in comparing two-digit numerals.

Reverse Distance Effect. For standard 65 with smaller targets, there was a reverse distance effect across the boundary between 30s and 40s: RTs for 36-39 were significantly slower than those for 41-44 ($F(1, 9) = 20.80$, $p < 0.01$). For standard 65 with larger targets, no reverse distance effect was found. For standard 55, no reverse distance effect was found for either larger or smaller targets. Since the reverse distance effect was a strong test of the possible interactions between decade and unit digits, the finding of the effect between 30s and 40s for standard 65 was a clear indication of such interactions. The no-shows of the effect at other boundaries, however, might be due to less adequate statistical power.

Model Fitting. The average RTs shown in Figure 2 were fitted to each of the three models (Equations 1, 2, and 3) by nonlinear regression, with separate fittings for targets below and above each standard. R^2, an index of how much of the variance in the data can be accounted for by the model, was obtained for each model, as shown in Table 2 in the parentheses. Because the three models have different numbers of parameters (2, 3, and 4 for holistic, sequential, and parallel models, respectively), we cannot directly use R^2 to decide which model has the best fit. To take the number of parameters into account, we used the Akaike Information Criterion (AIC[1]), which is an index of fit that penalizes more heavily models with more parameters as opposed to those with fewer parameters (Akaike, 1973, 1983; see also Myung & Pitt, 1997). The smaller the AIC of a model, the better the fit. For standard 65, the AIC values indicate that the parallel model is the best fit for targets below the standard and the sequential model is the best fit for targets above the standard. For standard 55, the AIC values indicate that the parallel model is the best fit for targets both below and above 55, although for targets above 55 the AIC value for the holistic model is only slightly larger than that for the parallel model.

Insert Table 2 about here

Summary and Discussion

For standard 65 with smaller targets, we found significant unit effects, decade effects, discontinuity effects, and reverse distance effects. The significance of the reverse distance effect indicates that the parallel model should be accepted (see Table 1). The significance of both the reverse distance effect and the unit effect indicates that the sequential models should be rejected. And the significance of both the reverse distance effect and the discontinuity effect indicates that the holistic model should be rejected. The AIC values also indicate that the parallel model has the best fit of the data. For standard 65 with larger targets, we found significant decade effects and discontinuity effects but not unit effects and reverse distance effects. The significance of the discontinuity effect indicates that the holistic model should be rejected (see Table 1). The parallel and the sequential models could neither be rejected nor accepted. However, the AIC values indicate that the sequential model has the best fit of data.

For standard 55 (with targets both below and above), we found significant unit effects, decade effects, and discontinuity effects but not reverse distance effects. The significance of the unit effect indicates that the sequential model should be rejected. The significance of the discontinuity effect indicates that the holistic model should be rejected. The parallel model can neither be rejected nor accepted. However, the AIC values indicate that the parallel model has the best fit of the data for targets both below and above the standard.

In sum, this experiment showed the following results. First, for both standards, the holistic model was rejected for targets both below and above the standards. Second, for both standards, the sequential model was rejected and the parallel model was accepted for targets below the standards. Third, for both standards, parallel and sequential comparisons were neither rejected nor accepted for targets above the standards, although sequential comparison was more likely than parallel comparison for standard 65 and parallel comparison was more likely than sequential comparison for standard 55.

These results are at odd with those found by Dehaene et al. (1990) and Hinrichs et al. (1981), which showed a holistic comparison for two-digit numerals. Although our results are largely consistent with Nuerk et al. (2001), in which they found a unit-decade-compatibility effect which supported a parallel comparison, we also found that a sequential comparison was more plausible when standard 65 was to be compared with a larger target. This finding, together with the inconsistent results regarding the reverse distance effect and the discontinuity effect for different standards, motivated us to conduct another experiment to replicate the findings of Experiment 1 and verify our results. The replication experiment was identical to Experiment 1 except that 65 and 45 were used as the two standards. To save space, we only give the summary of the results. The replication experiment generated nearly identical results. First, the holistic model was rejected for targets below and above both standards. Second, the sequential model was rejected but the parallel model was accepted for targets below both standards. Third, for targets above both standards, the sequential and the parallel models were neither rejected nor accepted, although the AIC values indicate that the sequential model has a slightly better fitting than the parallel model for both standards.

EXPERIMENT 2

In Experiment 1, the task was to decide which of the two externally presented numerals was larger. The result is that the comparison was somehow asymmetrical with regard to parallel and sequential comparisons. For targets below the standards, parallel comparison was the more plausible model. For targets above the standards, however, sequential and parallel comparisons were both likely. This asymmetry is quite puzzling. One explanation is that it could be due to the asymmetry of the task. What if the task is to decide which numeral is smaller? Will the results be reversed? Yet another explanation is that the asymmetry was merely an illusory phenomenon: the comparison is always parallel regardless of the whether the targets are below or above the standard and regardless of whether the task is to decide which numeral is larger or to decide which numeral is smaller. The probable sequential comparison above the standard in Experiment 1 might be merely a degraded parallel comparison. This is because the reverse distance effect we tested is a strong version of the Stroop-like effect. It could be that the Stroop-like effect was significant but the experiment lacked adequate power to detect it.

Experiment 2 examines these explanations by asking participants to decide which numeral is smaller as opposed to decide which numeral is larger in Experiment 1. If the results are reversed (i.e., the comparison is either sequential or parallel below the standard but parallel above the standard), then the hypothesis of task asymmetry is supported. If the comparison is parallel both below and above the standard, then the hypothesis of illusory asymmetry is supported.

Method

The design and procedure were exactly the same as in Experiment 1 except that instead of reporting which numeral was larger, the participants reported which numeral was smaller. There were 32 participants from the same participant pool as in Experiment 1.

Results

The average error rate was 3.5% for standard 65 and 3.1% for standard 55. Same procedures as those in Experiment 1 were used to pre-process the data before further analyses.

The average RTs for both standards are shown in Figure 3. For standard 65, an overall 2 (range: targets smaller vs larger than the standard) x 34 (magnitude distance) repeated-measure MANOVA showed a significant distance effect ($F(33, 693) = 12.28$, $p < 0.01$), a significant range effect ($F(1, 21) = 5.82$, $p < 0.05$), and a significant interaction (F(33,

693) = 6.43, $p < 0.01$). For standard 55, a similar 2 (range) x 44 (distance) MANOVA showed a significant distance effect ($F(43, 1032) = 12.48$, $p < 0.01$), a significant range effect ($F(1, 24) = 46.50$, $p < 0.01$), and a significant interaction ($F(43, 1032) = 2.03$, $p < 0.01$). These general patterns were consistent with those in Experiment 1.

Insert Figure 3 about here

Decade and unit effects. Similar two-way (decades x units) MANOVAs on mean RTs used in Experiment 1 were carried out to test the decade and unit effects. Significant decade effects were found for both standards when the targets were below the standards ($F(2, 44) = 118$, $p < 0.01$; and $F(3, 87) = 29.40$, $p < 0.01$, for standard 65 and 55 respectively). When the targets were above the standards, significant decade effects were found for standard 55 ($F(3, 84) = 19.60$, $p < 0.01$) but not for standard 65. However, significant unit effects were found for both standards, with the targets either above the standards ($F(8, 200) = 3.51$, $p < 0.01$; and $F(8, 224) = 5.69$, $p < 0.01$, for standard 65 and 55 respectively) or below the standards ($F(8, 176) = 11.58$, $p < 0.01$; and $F(8, 232) = 5.13$, $p < 0.01$, for standard 65 and 55 respectively). These patterns were largely consistent with those in Experiment 1, except of a more persistent unit effect no matter where the targets located relative to the standards.

Discontinuity Effect. Using the same discontinuity test as in Experiment 1, it was shown that for standard 65 there were significant discontinuity effects at decade boundaries between 50s and 60s ($F(1, 25) = 8.85$, $p < 0.01$) and between 60s and 70s ($F(1, 29) = 9.37$, $p < 01$). For standard 55, it was shown that was no significant discontinuity effect at all decade boundaries.

Reverse Distance Effect. For standard 65 with smaller targets, there was a reverse distance effect across the boundaries between 30s and 40s and between 50s and 60s: RTs for 36-39 were marginally slower than those for 41-44 ($F(1, 28) = 3.31$, $p < 0.10$) and

RTs for 56-59 were marginally slower than those for 61-64 ($F(1, 26) = 3.98$, $p < 0.10$). For standard 65 with larger targets, there was a reverse distance effect across the boundary between 70s and 80s: RTs for 76-79 were significantly faster than those for 81-84 ($F(1, 28) = 6.82$, $p < 0.05$). Some reverse distance effects were also found for standard 55. For those smaller targets, there was a reverse distance effect between 20s and 30s: RTs for 26-29 were significantly slower than those for 31-34 ($F(1, 29) = 3.48$, $p < 0.05$). For those larger targets, there was also a reverse distance effect between 70s and 80s: the RTs for 76-79 were marginally slower than those for 81-84 ($F(1, 26) = 6.77$, $p < 0.10$).

Model Fitting. The same nonlinear regression analyses in Experiment 1 were carried out for the present experiment. The average RTs shown in Figure 3 were fitted to each of the three models (Equations 1, 2, and 3) by nonlinear regression, with separate fittings for targets below and above each standard. The results are presented in Table 2. For standard 65, the AIC values indicate that the parallel model is the best fit for targets below the standard. For targets above the standard, both the sequential and parallel models have better fit than the holistic model but the sequential model has a slightly better fit than the parallel model. For standard 55, the parallel model has a best fit for targets both below and above the standard.

Summary and Discussion

In sum, Experiment 2 replicated the major findings of Experiment 1. For standard 65 with smaller targets, we found some significant unit effects, decade effects, discontinuity effects, and reverse distance effects. The significance of the reverse distance effect indicates that the parallel model should be accepted. The significance of both the reverse distance effect and the unit effect indicates that the sequential models should be rejected. And the significance of both the reverse distance effect and the discontinuity effect indicates that the holistic model should be rejected. The AIC also indicates that the parallel model has the best fit of the data. For standard 65 with larger targets, the unit effect, the discontinuity effect, and the reverse distance effect were significant but the decade effect was not significant. The significance of the reverse distance indicates that the parallel model should be accepted. The significance of both the reverse distance effect and the unit effect indicates that the sequential model should be rejected. And the significance of both the reverse distance effect and the discontinuity effect indicates that the holistic model should be rejected. The AIC indicates that both the sequential and parallel models have a better fit than the holistic model. Although the sequential model has a slightly smaller AIC value than the parallel model, the presence of reverse distance effects strongly supports the parallel model.

For standard 55 (with targets both below and above), we found significant unit effects, decade effects, the reverse distance effects but not discontinuity effects. The significance of the reverse distance effect indicates that the parallel model should be accepted. The significance of both the reverse distance effect and the unit effect indicates that the sequential models should be rejected. And the significance of the reverse distance effect indicates that the holistic model should be rejected. The AIC also indicates that the parallel model has the best of the data.

Overall, Experiment 2 helped clarify the issues related to the discrepancy for targets above and below the standards. It showed that for both standards, a parallel model is more plausible than a holistic or sequential one, for targets both below and above the standard. This result is not consistent with the task asymmetry hypothesis, which suggested that comparison asymmetry found in Experiment 1 was due to the fact that the task was to choose the larger one among the numerals. However, it is consistent with the illusory asymmetry hypothesis that predicted parallel comparisons for targets both below and above the standards. This suggests that the seemingly sequential comparison for targets above 65 might actually be a degraded parallel comparison because the Stroop-like effect was not strong enough to be observed as a reverse distance effect.

EXPERIMENT 3

Experiments 1 and 2 suggest that when two-digit number comparison is between two external representations, the comparison is likely parallel but not holistic. This is different from the result of the previous studies (Dehaene et al., 1990; Hinrichs et al, 1981) in which the comparison was found to be holistic when the comparison was between an internal and an external representation. To more directly test our hypothesis that the different comparison processes are due to different representations, we conducted Experiment 3 as a control experiment. In this experiment, the task was identical to the task in Experiments 1 of the present study. However, the comparison was between an internal and an external representation. If our hypothesis is correct, then we should find holistic comparison in this control experiment.

Method

The design and procedure were the same as in Experiment 1, except of the following changes. In this new design, a standard (55 or 65) was presented first for one second either on the left side or the right side of the fixation point, followed by a three-second blank interval for memory retention, and then followed by a target numeral on the other side of the fixation point. The three-second blank interval between the standard and the target was to give participants time to encode the standard as an internal representation. Participants were told to make their responses as soon as the target appeared. There were 31 participants from the same participant pool as in Experiment 1.

Results

The average error rate was 4.1% for standard 65 and 3.0% for standard 55. Same procedures as those in Experiment 1 were used to pre-process the data before further analyses.

The average RTs for both standards are shown in Figure 4. For standard 65, an overall 2 (range: targets smaller vs larger than the standard) x 34 (magnitude distance) repeated-measure MANOVA showed a significant distance effect (F(33, 627) = 9.38, $p < 0.01$), a significant range effect (F(1, 19) = 17.23, $p < 0.01$), and a significant interaction (F(33, 627) = 1.66, $p < 0.05$). For standard 55, a similar 2 (range) x 44 (distance) MANOVA showed a significant distance effect (F(43, 774) = 9.51, $p < 0.01$), a significant range effect (F(1, 18) = 17.37, $p < 0.01$), but an insignificant interaction.

Insert Figure 4 about here

Decade and unit effects. Two-way (decades x units) MANOVAs on mean RTs were carried out to test the decade and unit effects. Significant decade effects were found for both standards with the targets either below the standards (F(2, 46) = 27.40, $p < 0.01$; and F(3, 69) = 6.03, $p < 0.01$, for standard 65 and 55 respectively) or above the standards (F(2, 48) = 6.17, $p < 0.01$; and F(3, 69) = 2.98, $p < 0.01$, for standard 65 and 55 respectively). Significant unit effects were found for standard 65 only when the targets were below the standard (F(8, 184) = 2.11, $p < 0.05$) and for standard 55 only when the target were above the standard (F(8, 184) = 2.23, $p < 0.05$). There were no significant unit effects for all other conditions, indicating a much weaker influence on performance from the unit digits.

Discontinuity Effect. No significant discontinuity effect at any decade boundaries was found for either standard.

Reverse Distance Effect. No significant reverse distance effect was found for either standard.

Model Fitting. The nonlinear regression results are again presented in Table 2. The AIC values indicate that the holistic and parallel models in general have a better fit than the sequential model for both standards, both in the above and below ranges.

Summary and Discussion

The comparison in this experiment was between an internal and an external representation. We predicted that the comparison would be holistic. Although the results did not converge strongly to support a holistic comparison model, they showed signs of holistic comparison. For both standards, sequential comparison was rejected. The model fitting results indicate that the holistic model is the best fit for half of the data and the parallel model is the best for the other half of the data. However, one of the major findings of this experiment is the complete no-show of both the discontinuity effect and the reverse distance effect at any decade boundaries. Taking this into the consideration, it seems that a holistic comparison was more strongly supported. We suspect that the parallel comparison for the other half of the data might be due to the incomplete internalization of the standards, which randomly alternated between 65 and 55 and were processed for internalization for a short period of three seconds. In the previous studies that showed holistic comparison, there was only one standard that was always in memory throughout the experiment.

GENERAL DISCUSSION

The Effect of Representations

The experiments of the present study were minor modifications of the previously published experiments by Dehaene et al. (1990) and Hinrichs et al. (1981). However, the results were quite different. Our Experiments 1 and 2 showed that the number comparison was not holistic but parallel when both the to-be-compared two-digit numerals were externally presented. Our Experiment 3 showed that when one of the numerals was presented before the other one a holistic comparison became more likely.

We argue that the difference in results is a reflection of the difference in representations. In the studies by Dehaene et al., (1990) and Hinrichs et al. (1981), the standard was kept in memory as an internal representation but the target was on the screen as an external representation. Because numerals can be compared only when they are in the same form of representation (e.g., Noel & Seron, 1993; Dehaene & Akhavein, 1995), the target, which was an external Arabic numeral, and the standard, which was an internal numerical representation, must be converted into a common representation. This common representation could be a phonological representation of number words or, as proposed by those authors, a format-independent line-like analog representation. The result was a holistic comparison.

In our present study, both the target and the standard were already in the same form of representation: they were both external Arabic numerals. They need not be converted into another form of representation before they could be compared. The dimensional structures of Arabic numerals could directly activate a set of perceptual and cognitive processes. The power dimension could activate perceptual processes such as perceptual identification of the positions of individual digits and perceptual search of specific digits, and the base dimension could activate cognitive processes such as memorial retrieval of the values of individual digits and mental comparison of individual digits. Both the perceptual and the cognitive processes were the basic processes and worked together to execute the comparison task. Neither was a peripheral process for the other. The result was a parallel comparison, caused by the interaction of the perceptual and cognitive processes in the form of a Stroop-like effect.

Our results are consistent with the work by Nuerk et al. (2001). They found a strong decade and unit interaction when two externally presented (vertically aligned) two-digit Arabic numerals were compared. Although the authors did not explicitly argue for a parallel comparison model, their results supported the notion that all component digits in multi-digit numerals played an important role in multi-digit number comparison when the to-be-compared numerals are simultaneously presented.

The effects of external representations on number comparison revealed in this article have important implications on human mathematical problem solving in general. The existence of holistic, continuous, and quantity-based internal representations of numbers may not be the only plausible explanation for interesting phenomena such as the distance effect (the performance of comparing two numbers improves as the numerical distance between them increases) and the size effect (for equal numerical distance, performance decreases as their numerical magnitude increases) (see Dehaene, 1997 for a review). The increase of distance and/or size typically involves changes in external representations of numerals. These changes, together with the semantic representations of numbers and their relations, might play an important role in human mathematical thinking. The simultaneous manifestation of the distance effect and the Stroop-like effect, shown both in Nuerk et al. (2001) and our current study, suggests that it is possible that both parallel comparison and holistic comparison may work together to produce the behavior. This possibility is especially likely when both the to-be-compared numerals are simultaneously and externally available.

It is interesting to note that this view fits well with a large body of recent neuroimaging evidence that suggests that human mathematical ability emerges from the interplay of multiple brain areas, including those responsible for linguistic competence as well as those responsible for visuo-spatial processing (see Dehaene et al., 1999; Pesenti et al., 2000; Naccache & Dehaene, 2001b; Pinel, et al., 2001; Zago, et al., 2001; Dehaene, 2003). Future work is clearly needed to further clarify the nature of the interactions.

Theories of Number Representations

One of the central issues in numerical cognition is how different types of numerals are represented in the mind and support various numerical abilities. Different theoretical opinions exist. On the one hand, it has been argued that there is a single, abstract, and format-independent internal representation of numerical quantity that all different types of numerals have to be encoded into in order to be useful. The view was explicitly formulated by McCloskey and his colleagues (McCloskey, 1992; McCloskey, Caramazza, & Basili, 1985; McCloskey, Sokol, & Goodman, 1986). According to their model, to perform a numerical task, a comprehension operation has to be carried out first to transform external numerals to abstract internal representations. Various calculations and arithmetical retrievals then operate on these abstract representations. And finally a production operation is necessary to transform back the results to various output formats.

On the other hand, it has been argued that the representational effect in numerical tasks is not caused by various peripheral comprehension and production processes but by the fundamentally different internal representations that are specific to different external formats. For example, Gonzalez & Kolers (1982; 1987) found that when processing Arabic and Roman numerals people did not transform different external representations into a common abstract internal representation. Rather, they operated upon different internal representations that reflect the physical characteristics of different external representations. Deloche and Seron (1987) also suggested a transcoding mechanism that does not require a central abstract representation. In this view, one type of number representations (e.g., Arabic numerals) can be directly transcoded into another type of number representations (e.g., English number words) without the mediation of any central representations. As a representative for the format-specific view, the encoding-complex model developed by Campbell and Clark (1988, 1992; Clark & Campbell, 1991; Campbell, 1994) denies any central abstract representations. In this model, different forms of external number representations activate different forms of internal representations, which are functionally integrated in an encoding complex. Within the complex, one form of internal representations can potentially activate other forms of internal representations, and any form of internal representations can be involved in the comprehension, calculation, and production of numbers.

The triple-code model, proposed by Dehaene and colleagues (Dehaene, 1992; Dehaene & Akhavein, 1995), is an attempt to accommodate the above two extreme views. According to the triple-code model, in addition to an abstract and format-specific numerical representation, people can also represent numbers in multiple asemantic ways, including Abrabic, verbal, and phonological. Through various transcoding procedures, these different internal representations directly interact with each other. More importantly, they support different cognitive operations: while the verbal code is convenient for verbal counting and the Arabic code is often used for multi-digit calculation, it is an abstract and format-independent magnitude representation that underlies number comparison and quantity judgment.

Though the triple-code model recognizes the possibility of multiple types of internal representations of numbers, it emphasizes that these asemantic representations eventually have to converge to a common analogical representation of quantity (the mental number line) for number comparison. Dehaene and colleagues have argued that this type of abstract and format-independent quantity representation is the foundation of a “number sense” and is shared by animals and humans (Dehaene, 1997; Dehaene, et al., 1998). The existence of such an abstract quantity representation for numbers has also been supported by studies using the priming method, which demonstrated strong format-independent semantic priming in various number naming and number comparison tasks (Koechlin et al., 1999; Noccache & Dehaene, 2001a; Reynvoet et al., 2002). Noel and colleagues (Noel & Seron, 1992; Noel et al., 1997) directly examined the various format-dependent findings by Gonzalez and Kolers (1987) and Campbell and Clark (1992; Campbell, 1994). They found that the format-dependent effects might be due to a format induced encoding process but not due to different internal representations.

Distributed Representations

While the debate will certainly continue, the current study provides a different angle to understand numerical representations and emphasizes the importance of various external representations in human numerical cognition. Both the format-independent and the format-specific views described above are mostly concerned with internal representations of numbers: how people perform numerical tasks in their heads, how numbers and arithmetic facts are represented in memory, and what mental processes and procedures are involved in the comprehension, calculation, and production of numbers. In addition, both views are developed mainly from the studies of simple numerical tasks that can be performed entirely in internal representations, such as single or double-digit number comparison and simple arithmetic. Internal number representations are certainly important. However, the theories and models developed for them may not account for numerical tasks that involve external representations. As an example of numerical tasks that involve external representations but are possible to perform without them, the line-like analog representation assumed for two-digit Arabic numerals can only account for the holistic comparison when the comparison is performed on internal representations but cannot account for the parallel comparison when the comparison is performed on external representations. As an example of numerical tasks that are impossible to perform without external representations, consider the task of multiplying 735 by 278. Without special training and expertise, nobody can perform this task by encoding the two three-digit numerals in internal representations and carrying out the calculation entirely in internal representations.

Therefore, to account for the full spectrum of numerical tasks, including those involving only internal representations and those involving both internal and external representations, we should consider number representations in general as distributed representations that have internal and external representations as two indispensable components. We should also consider numerical tasks as distributed cognitive tasks that require the interactive processing of information distributed across internal and external representations (see Zhang & Norman, 1994, 1995; Zhang, 1996, 1997, 2000). There are not yet enough empirical data for us to propose a detailed process model with explicitly specified processing mechanisms for external representation based numerical tasks. Nevertheless, we can specify a set of properties of such tasks that should be considered by any potential process models. First, external representation based numerical tasks involve interactive, integrative, and dynamic processing of information perceived from external representations and that retrieved from internal representations. Second, the processes in such tasks are activated and determined by representations, not vice versa. Thus, even if different types of numerals (Arabic, Roman, etc.) might have a single internal representation, they have different distributed representations because their external representations are different. Therefore, it is different distributed representations that are responsible for the representational effect in numerical tasks. For number comparison, this means that there is no single comparison process for all types of number representations. Fourth, external representations need not be re-represented as internal representations in order to be involved in numerical tasks: they can directly activate perceptual processes and directly provide perceptual information that, in conjunction with the memorial information and cognitive processes provided by internal representations, determine the behavior in numerical tasks. Perceptual processes are basic processes of numerical tasks, just like cognitive processes. They directly operate upon external representations to participate in numerical tasks.

CONCLUSION

The present study explored the effect of external representations on number comparison tasks. A few minor modifications on the previously reported two-digit number comparison task produced different results. The difference in results between previous and current studies is a reflection of representational forms: the comparison was based on internal representations for previous studies but on external representations for our present study. This effect on number comparison caused by external representations supports the framework of distributed number presentations. In complex numerical tasks that involve external representations, number representations should be considered as distributed representations and the behavior in these tasks should be considered as the interactive processing of internal and external information through the interplay of perceptual and cognitive processes. We suggest that any general theory of number representations and any process models of numerical tasks should consider external representations as an essential component.

REFERENCES

Akaike, H. (1973). Information theory and an extension of the maximum likelihood principle. In B. N. Petrox & F. Caski (Eds.), *Second International Symposium on Information Theory (pp. 267)*. Budapest: Akademiai Kiado.

Akaike, H. (1983). Information measures and model selection. *Bulletin of the International Statistical Institute, 50*, 277-290.

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

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

Campbell, J. I. D. (1994). Architectures for numerical cognition. *Cognition, 53,* 1-44.

Campbell, J. I. D., & Clark, J. M. (1988). An encoding-complex view of cognitive number processing: Comment on McCloskey, Sokol, and Goodman (1986). *Journal of Experimental Psychology: General, 117* (2), 204-214.

Campbell, J. I. D., & Clark, J. M. (1992). Cognitive number processing: An encoding-complex perspective. In J. I. D. Campbell (Ed.). *The nature and origin of mathematical skills.* Amsterdam: Elsevier.

Clark, J. M., & Campbell, J. I. D. (1991). Integrated versus modular theories of number skills and acalculia. *Brain and Cognition, 17*(2), 204-239.

Dehaene, S. (1992). Varieties of numerical abilities. *Cognition, 44,* 1-42.

Dehaene, S. (1993). *Numerical cognition.* Cambridge, MA: Blackwell Publishers.

Dehaene, S. (1997). *The number sense: How the mind creates mathematics*. New York: Oxford University Press.

Dehaene, S. (2003). The neural basis of the Weber-Fechner law: A logarithmic mental number line. *Trends in Cognitive Sciences, 7,* 145-147.

Dehaene, S., & Akhavein, R. (1995). Attention, automaticity, and levels of representation in number processing. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 21* (2), 314-326.

Dehaene, S., Dehaene-Lambertz, G., & Cohen, Laurent (1998). Abstract representations of numbers in the animal and human brain. *Trends in Neuroscience, 21,* 355-361.

Dehaene, S., Dupoux, E. & Mehler, J. (1990). Is numerical comparison digital? Analogical and symbolic effects in two-digit number comparison. *Journal of Experimental Psychology: Human Perception and Performance, 16*, 626-641.

Dehaene, S., Spelke, E., Pinel, P., Stanescu, R., & Tsivkin, S. (1999). Sources of mathematical thinking: Behavioral and brain-imaging evidence. *Science, 284*, 970-974.

Deloche, G., & Seron, X. (1987). Numerical transcoding: A general production Model. In G. Deloche & X. Seron (Ed.), *Mathematical disabilities: A cognitive neuropsychological perspective.* Hillsdale, NJ: Lawrence Erlbaum Associates.

Foltz, G. S., Poltrock, S. E., & Potts, G. R. (1984). Mental comparison of size and magnitude: Size congruity effects. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 10*, 442-453.

Gallistel, C. R., & Gelman, R. (1992). Preverbal and verbal counting and computation. *Cognition, 44*, 43-74.

Gonzalez, E. G., & Kolers, P. A. (1982). Mental manipulation of arithmetic symbols. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 8* (4), 308-319.

Gonzalez, E. G., & Kolers, P. A. (1987). Notational constraints on mental operations. In G. Deloche & X. Seron (Ed.), *Mathematical disabilities: A cognitive neuropsychological perspective.* Hillsdale, NJ: Lawrence Erlbaum Associates.

Hinrichs, J. V., Berie, J. L., & Mosell, M. K. (1982). Place information in multidigit number comparison. *Memory and 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.

Koechlin, E., Naccache, L., Block, E., & Dehaene, S. (1999). Primed numbers: Exploring the modularity of numerical representations with masked and unmasked priming. *Journal of Experimental Psychology: Human Perception and Performance, 25,* 1882-1905.

McCloskey, M. (1992). Cognitive mechanisms in numerical processing: Evidence from acquired dyscalculia. *Cognition, 44*, 107-157.

McCloskey, M., Caramazza, A., & Basili, A. G. (1985). Cognitive mechanisms in number processing and calculation: Evidence from dyscalculia. *Brain and Cognition, 4*, 171-196.

McCloskey, M., Sokol, S. M., & Goodman, R. A. (1986). Cognitive processes in verbal-number production: Inferences from the performance of brain-damaged subjects. *Journal of Experimental Psychology: General, 115*, 307-330.

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

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

Myung, I. J., & Pitt, M. A. (1997). Applying Occam's razor in modeling cognition: A Bayesian approach. *Psychonomic Bulletin & Review, 4*, 79-95.

Naccache, L., & Dehaene, S. (2001a). Unconscious semantic priming extends to novel unseen stimuli. *Cognition, 80,* 223-237.

Naccache, L., & Dehaene, S. (2001b). The priming method: Imaging unconscious repetition priming reveals an abstract representation of number in the parietal lobes. *Cerebral Cortex, 11*, 966-974.

Noel, M. P., Fias, W., & Brysbaert, M. (1997). About the influence of the presentation format on arithmetical-fact retrieval processes. *Cognition, 63,* 335-374.

Noel, M. P., & Seron, X. (1992). Notational constraint and number processing: A reappraisal of the Gonzales & Kolers Study. *Quarterly Journal of Experimental Psychology – A, 45,* 451-478.

Noel, M. P., & Seron, X. (1993). Arabic number reading deficit: A single case study. *Cognitive Neuropsychology, 10,* 317-339.

Nuerk, H-C., Weger, U., & Willmes, K. (2001). Decade breaks in the mental number line? Putting the tens and units back in different bins. *Cognition, 82,* B25-B33.

Pesenti, M., Thious, M., Seron, X., & De Volder, A. (2000). Neuroanatomical substrates of abrabic number processing, numerical comparison, and simple addition: A PET study. *Journal of Cognitive Neuroscience, 12,* 461-479.

Pinel, P., Dehaene, S., Riviere, D., & LeBihan, D. (2001). Modulation of parietal activation by semantic distance in a number comparision task. *NeuroImage, 14,* 1013-1026.

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.

Reynvoet, B., Brysbaert, M., & Fias, W. (2002). Semantic priming in number naming. *Quarterly Journal of Experimental Psychology – A, 55,* 1127-1139.

Takahashi, A., & Green, D. (1983). Numerical judgments with kanji and kana. *Neuropsychologia, 21*, 259-263.

Zago, L., Pesenti, M., Mullet, E., Crivello, F., Mazoyer, B., & Tzourio-Mazoyer, N. (2001). Neural correlates of simple and complex mental calculation. *Neuroimaging, 13*, 314-327.

Zhang, J., & Norman, D. A. (1994). Representations in distributed cognitive tasks. *Cognitive Science, 18*, 87-122.

Zhang, J., & Norman, D. A. (1995). A representational analysis of numeration systems. *Cognition*, *57*, 271-295.

Zhang, J. (1996). A representational analysis of relational information displays. *International Journal of Human-Computer Studies*, *45*, 59-74.

Zhang, J. (1997). The nature of external representations in problem solving. *Cognitive Science*, *21*, 179-217.

Zhang, J. (2000). External representations in complex information processing tasks**.** In A. Kent (Ed.), *Encyclopedia of Library and Information Science.* New York: Marcel Dekker, Inc.

AUTHOR NOTES

This research was in part supported by Grants N00014-95-1-0241 and N00014-96-1-0472 from the Office of Naval Research, Cognitive and Neural Sciences Technology Division. We would like to thank Gwen Hall for his assistance in the experiments. Correspondence and request for reprints should be sent to Jiajie Zhang, School of Health Information Sciences, University of Texas Health Science Center at Houston, 7000 Fannin Street, Suite 600, Houston, Texas 77030, USA. Email: Jiajie.Zhang@uth.tmc.edu.

FOOTNOTES

[1] $AIC_i = -2\ln(ML_i) + 2n_i$, where ML_i is the maximum likelihood for Model i and n_i is the number of free parameters in the model. The criterion prescribes that the model that minimizes the AIC should be chosen. AIC can be rewritten as a function of SSE (sum squared error): $AIC_i = N\ln(SSE_i) + 2n_i + (-N\ln N + \frac{1}{N} + N\ln 2\pi)$, where SSE_i is the Sum Squared Error for Model i, n_i is the number of free parameters in the model, and N is the number of observations (Myung, personal communication). The latter equation was used in the present study.

Table 1. The Presence or Absence of a Specific Effect for Each Model

		Effects			
		Unit	Decade	Discontinuity	Stroop-like
Models	Sequential	-	+	+	-
	Parallel	+	+	+/-	+
	Holistic	+	+	-	-

Table 2: AIC values of Model Fitting (R^2 in parentheses)

	Models	Standard 65		Standard 55	
		Smaller than Standard	Larger than Standard	Smaller than Standard	Larger than Standard
Experiment 1	Sequential	342 (0.49)	280 (0.88)	401 (0.67)	361 (0.90)
	Parallel	311 (0.81)	291 (0.84)	368 (0.85)	356 (0.92)
	Holistic	322 (0.70)	309 (0.70)	375 (0.81)	358 (0.91)
Experiment 2	Sequential	338 (0.63)	304 (0.66)	394 (0.72)	383 (0.61)
	Parallel	317 (0.81)	306 (0.66)	370 (0.84)	376 (0.66)
	Holistic	332 (0.68)	325 (0.34)	372 (0.83)	377 (0.65)
Experiment 3	Sequential	287 (0.73)	263 (0.84)	356 (0.80)	376 (0.67)
	Parallel	284 (0.76)	255 (0.88)	347 (0.85)	369 (0.74)
	Holistic	283 (0.74)	280 (0.72)	358 (0.78)	367 (0.72)

FIGURE CAPTIONS

Figure 1. Three models of two-digit number comparison. Each graph represents the reaction times of comparing target numerals 11-54 and 56-99 with the standard 55. See Equations 1, 2, and 3 and the text for detailed explanations.

Figure 2. Reaction times for targets compared with 65 (A) and 55 (B) in Experiment 1.

Figure 3. Reaction times for targets compared with 65 (A) and 55 (B) in Experiment 2.

Figure 4. Reaction times for targets compared with 65 (A) and 55 (B) in Experiment 3.

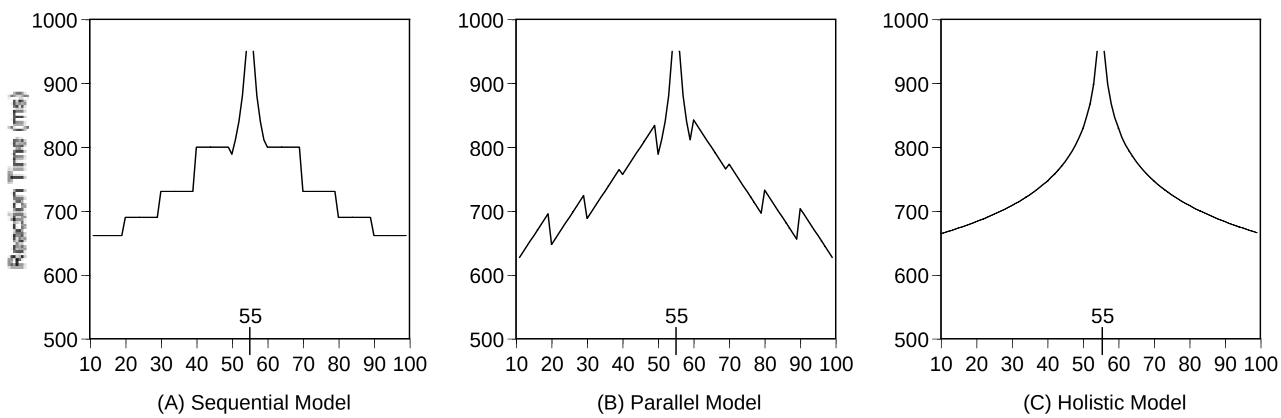
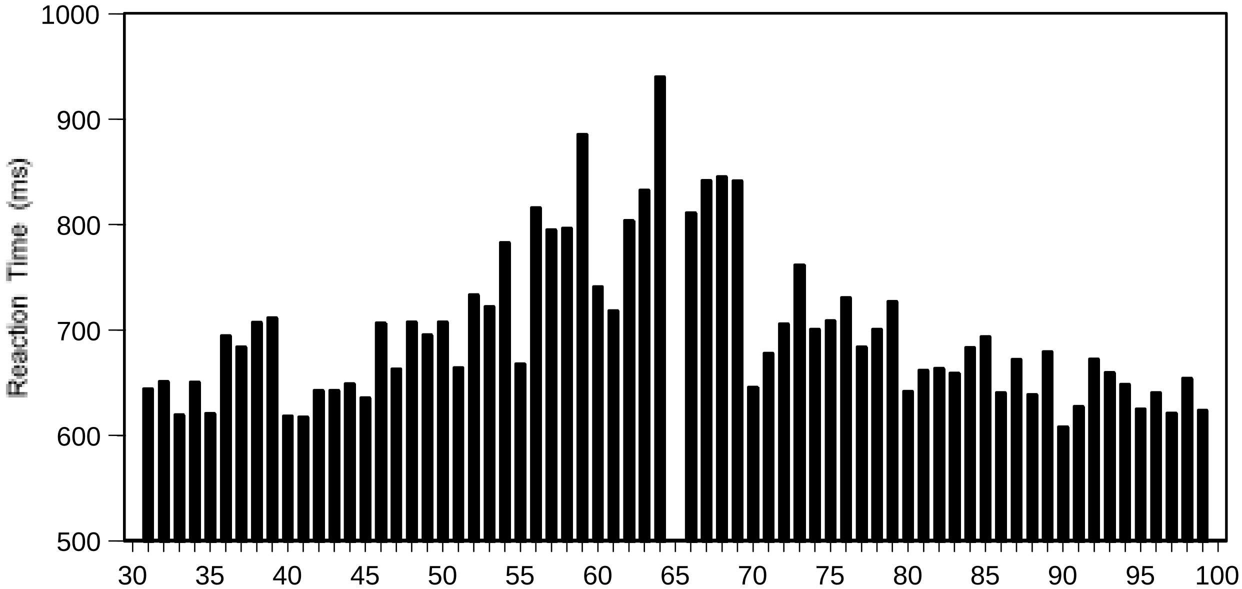
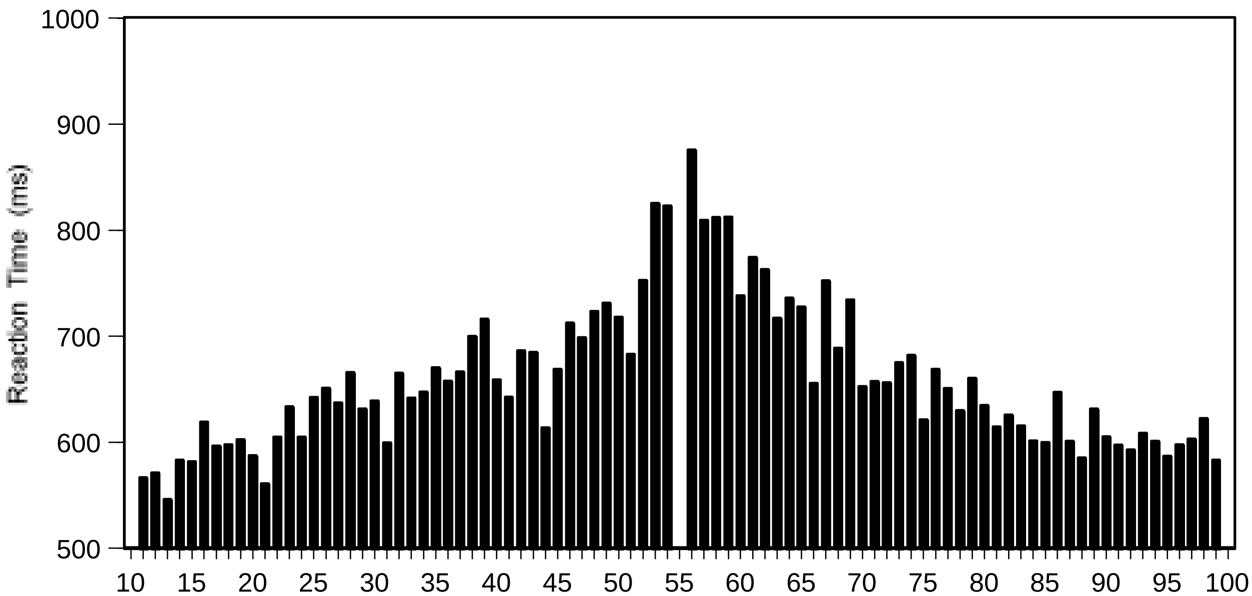

Reaction Time (ms)
1000
900
800
700
600
500
55
10 20 30 40 50 60 70 80 90 100
(A) Sequential Model
(B) Parallel Model
(C) Holistic Model

Figure 1

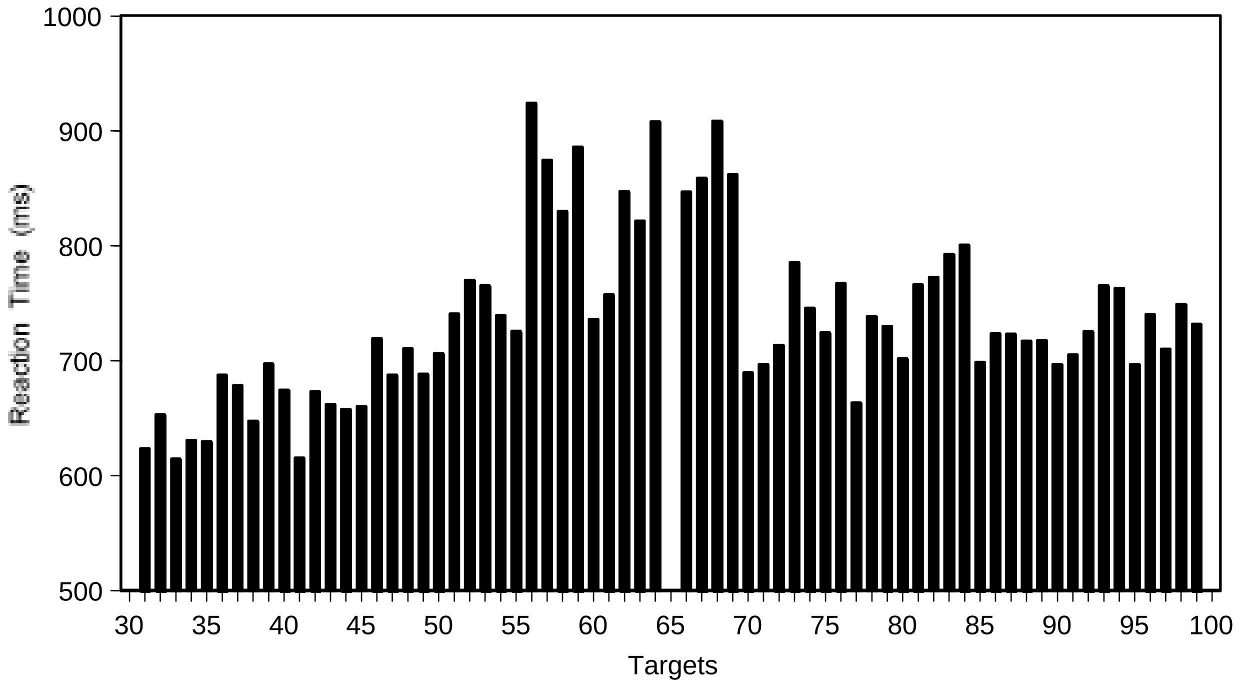


A. Standard 65

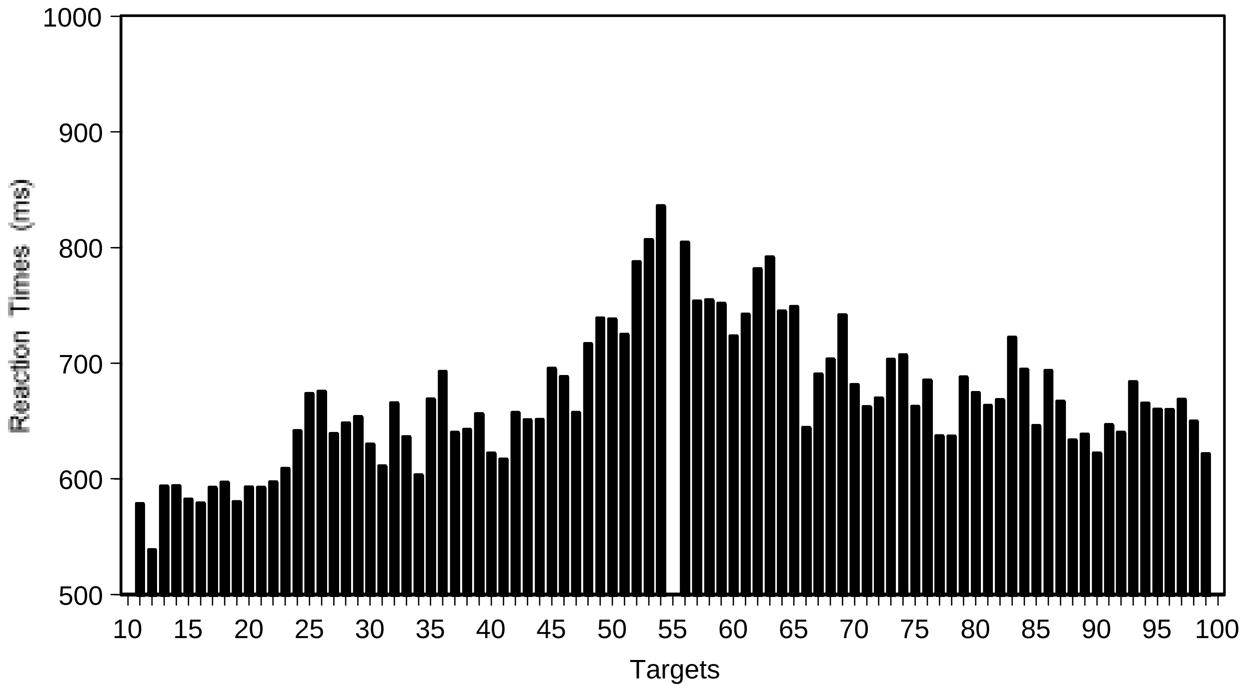


B. Standard 55

Figure 2

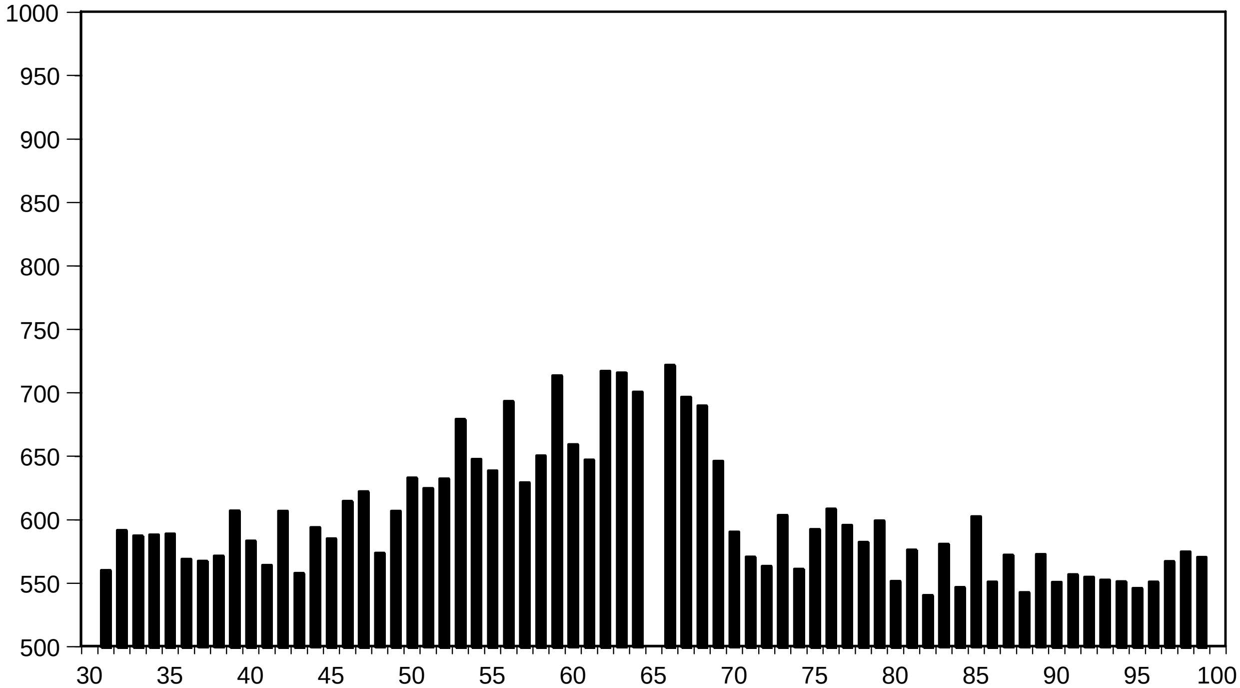


A. Standard 65

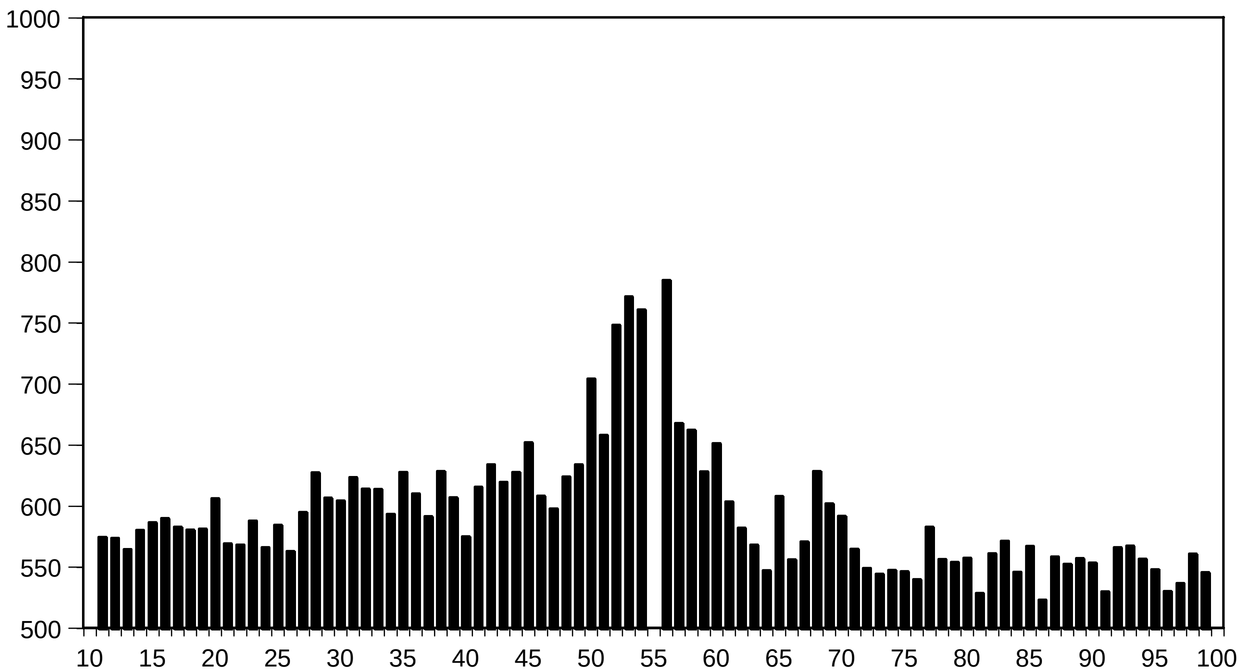


B. Standard 55

Figure 3



A. Standard 65



B. Standard 55

Figure 4