Title
The Integration of internal and External Information in Numerical Tasks

Journal
Proceedings of the Annual Meeting of the Cognitive Science Society, 17(0)

Authors
Zhang, Jiajie
Wang, Hongbin

Publication Date
1995

Peer reviewed

The Integration of internal and External Information in Numerical Tasks

Jiajie Zhang & Hongbin Wang
Department of Psychology & Center for Cognitive Science
The Ohio State University
1827 Neil Avenue
Columbus, OH 43210
zhang.52@osu.edu wang.190@osu.edu

Abstract

Numerical tasks with Arabic numerals involve the integration of internal and external information and the interaction between perception and cognition. 2-digit number comparison task was selected to study these integration and interaction processes. To compare the magnitudes of two 2digit Arabic numerals, we can (1) compare them digitbydigit sequentially, (2) compare corresponding digits in parallel, or (3) encode them as an integrated representation and compare the whole numerical values. Previous studies showed that 2-digit comparison was holistic when target numerals were compared with a standard held in memory. In our experiment target numerals and standards were presented on the same external display at the same time. Instead of a holistic comparison, we found that 2-digit comparison was a combination of sequential and parallel comparisons. The implications of this discrepancy were discussed in terms of the interplay between perception and cognition.

Although different types of numerals (e.g., Arabic, Roman, Greek, etc.) all represent the same abstract quantities—numbers, they can produce dramatically different cognitive behaviors in numerical tasks (e.g., Nickerson, 1988; Zhang & Norman, in press). This representational effect, caused by different representations of a common structure, can be easily observed by comparing the difficulties and processes of two multiplication tasks: 735 × 278 (Arabic numerals) and DCCXXXV × CCLXXVIII (Roman numerals, equivalent to 735 × 278). In addition, to perform numerical tasks, people usually need to process information distributed across the internal mind and the external environment in an interactive manner (Zhang & Norman, 1994, in press). For example, to do 735 × 278 with 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.).

In this paper, we use a simple numerical task, number comparison, to examine how internal and external information is processed and integrated in numerical tasks and what effects such interactions and integrations have on behavior.

Number Comparison

The time to compare the magnitudes (larger or smaller) of two 1-digit Arabic numerals decreases with the numerical distance between them (Moyer & Landauer, 1967). For example, it is faster to compare 1 and 9 than 8 and 9. This distance effect resembles that found for physical stimuli such as dot patterns and line lengths, implying that Arabic digits might have an internal representation analogous to a physical continuum.

Multidigit 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 positions of the digits (Zhang & Norman, in press). The base dimension is an internal representation because the numerical value of each digit has to be memorized, whereas the power dimension is an external representation because the position of a digit can be perceptually inspected. Unlike 1-digit comparison, which is solely internal, multidigit comparison is based on both internal information (the values of individual digits) and external information (positions of digits). Thus, multidigit comparison is a good task for the study of the integration of internal and external information and the interaction between perceptual and cognitive processes.

To compare the magnitudes of multidigit Arabic numerals, we can (1) compare them digit by digit sequentially, (2) compare corresponding digits in parallel, or (3) encode them as an integrated representation and compare the whole numerical values. In the first case (*sequential comparison*), only the highest digits should affect the comparison unless they are not sufficient for making decisions. In the second case (*parallel comparison*), lower digits may facilitate or interfere with the comparison of higher digits. In the third case (*holistic comparison*), only the absolute numerical distances should matter. Empirical studies have revealed a discrepancy between 2-digit comparison and higher multidigit (3 or more) comparison: 2-digit comparison is 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 of 2-digit Arabic numerals was often cited as evidence of a holistic analog internal representation of Arabic numerals.

This paper addresses two specific issues: (1) whether 2-

digit comparison is indeed holistic and (2) how internal and external information is processed and integrated in 2-digit comparison tasks.

Experiment

In the experiments reported by Dehaene et al. (1990) and Hinrichs et al. (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, one of the two numerals to be compared (the standard) was already preprocessed. In our experiment we made three procedural changes. First, the standard and targets were presented simultaneously on an external display such that both numerals had to be processed at the same time. Second, instead of one standard, we used two standards (55 and 65) as a within-subject factor to prevent subjects from preprocessing a specific standard. Third, instead of reporting whether a target is smaller or larger than the standard, subjects only decide which of the two presented numerals is larger. Such procedural changes are important for the testing of our hypothesis: when both numerals are presented externally, the comparison is based on not just internal but also external information, the concurrent processing of which can generate a different pattern of behavior.

In order to identify which of the three comparison models can best account for 2-digit comparison under our experimental conditions, we need three observations. First, if there are reaction time (RT) differences across decades but not within a decade and there are discontinuities across decade boundaries, then sequential comparison is supported. Second, if the whole numerical value resulting from the integration of the decade and unit digits is the only determinant of RT, then holistic comparison is most likely. Third, if unit digits show a Stroop-like congruity effect on RT, then parallel comparison is evident. One difficulty with the observation of a congruity effect is that the values of unit digits and the absolute distances from the standard are always confounded. For example, a shorter RT to compare 31 and 65 than 39 and 65 can be either due to the facilitation of the 1 in 31 and the interference of the 9 in 39 or the longer distance between 31 and 65 than between 39 and 65 or both. However, if we can observe a reverse distance effect across different decades, then we can still single out a congruity effect. For example, the RTs for 36-39 might be longer than those for 41-44 even if 36-39 are farther away from 65 than 41-44 are.

Method

Subjects. The subjects were 32 undergraduate students in introductory psychology courses at The Ohio State University, who participated in the experiment to earn course credit.

Design and procedure. The subjects were seated in about 40 cm from a Macintosh computer in a dark room. 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. The Macintosh computers (Quadra 700 and Centris 610) used to control the experiment could measure reaction times with a resolution of ± 1 ms. 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 2-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. The standards were presented on the left side for half of the trials and on the right side for the other half, producing a total of 312 trials. These 312 trials were randomized for each experimental session and divided into four blocks with 78 trials for each block. Each subject was presented 10 randomly generated pairs of 2-digit numerals for practice followed by the four blocks of trials with one minute rest between blocks.

Results

For all analyses that follow, trials with a standard on the left side were pooled with the corresponding trials with the same standard on the right side. Trials with errors were excluded from the analysis of reaction times (RTs). For standard 65, the average error rate was 2.7%, ranging from 1.5% in the 30s and 90s to 5.6% in the 60s. For standard 55, the average error rate was 2.0%, ranging from 0.64% in the 10s and 90s to 4.2% in the 50s. RTs deviated from the mean for each target by more than three standard deviations were excluded from analyses. Separate analyses were conducted for 65 and 55.

Standard 65. The average RTs to compare target numerals with 65 are shown in Figure 1. An analysis of the effect of the numerical distances between targets and 65 and the ranges of the targets (smaller or larger than 65) showed a significant distance effect ($F(33, 693) = 16.99, p < 0.001$), a significant range effect ($F(1, 21) = 8.40, p < 0.009$), and a significant interaction ($F(33, 693) = 4.86$, $p < 0.001$). Due to the asymmetrical range effect and the significant interaction between distance and range, we conducted separate analyses on RTs for targets smaller than 65 and those larger than 65.

Decade and unit effects. For targets 31-59, the decade effect between 40s and 50s was significant ($F(1, 25) = 77.93$, $p < 0.001$) but that between 30s and 40s was not significant ($F(1, 27) = 0.94$, $p = 0.34$). For each of the decades of 30s, 40s, and 50s, the unit effect was significant (smallest $F(8, 216) = 3.91$, $p < 0.001$). For targets 71-99, the decade effect between 70s and 80s and that between 80s and 90s were both significant (smallest $F(1, 25) = 14.02$, $p < 0.001$) None of the decades of 70s, 80s, and 90s had a significant unit effect (largest $F(8, 232) = 1.29$, $p = 0.25$).

The unit effect was further analyzed by linear regression. The RT for each target in a decade was subtracted by the mean RT of the corresponding decade, then averaged across

30s, 40s, & 50s and across 70s, 80s, & 90s (see Figure 3A). The unit effect was asymmetrical. For targets smaller than 65, the units had a strong effect with a slope of 13.7, which was significantly different than zero ($r^2 = 0.71$, $p < 0.005$). However, for targets larger than 65, the units had no significant effect: the slope (-0.66) was not significantly different than zero ($r^2 = 0.024$, $p > 0.69$). Separate regression analysis for each decade showed significant unit effect for decades 30s, 40s, and 50s (largest $p < 0.01$) but not for decades 70s, 80s, and 90s (smallest $p = 0.25$).

Discontinuity. If there is a discontinuity at the boundary of two decades, there should be a sharp change in RT. 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). An analysis of variance showed significant discontinuity effects between 30s and 40s, 50s and 60s, 60s and 70s, 70s and 80s, and 80s and 90s (smallest $F(1, 30) = 4.76$, $p < 0.05$). The effect between 40s and 50s was not significant ($F(1, 30) = 0.32$, $p = 0.57$).

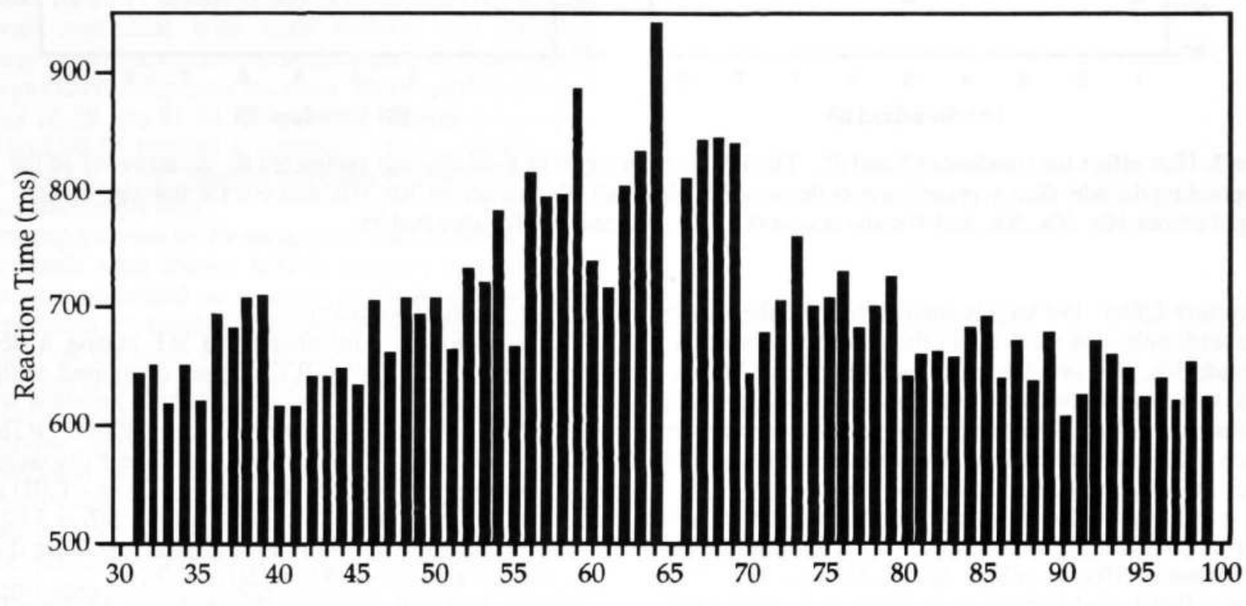

Figure 1. Reaction times for targets compared with 65.

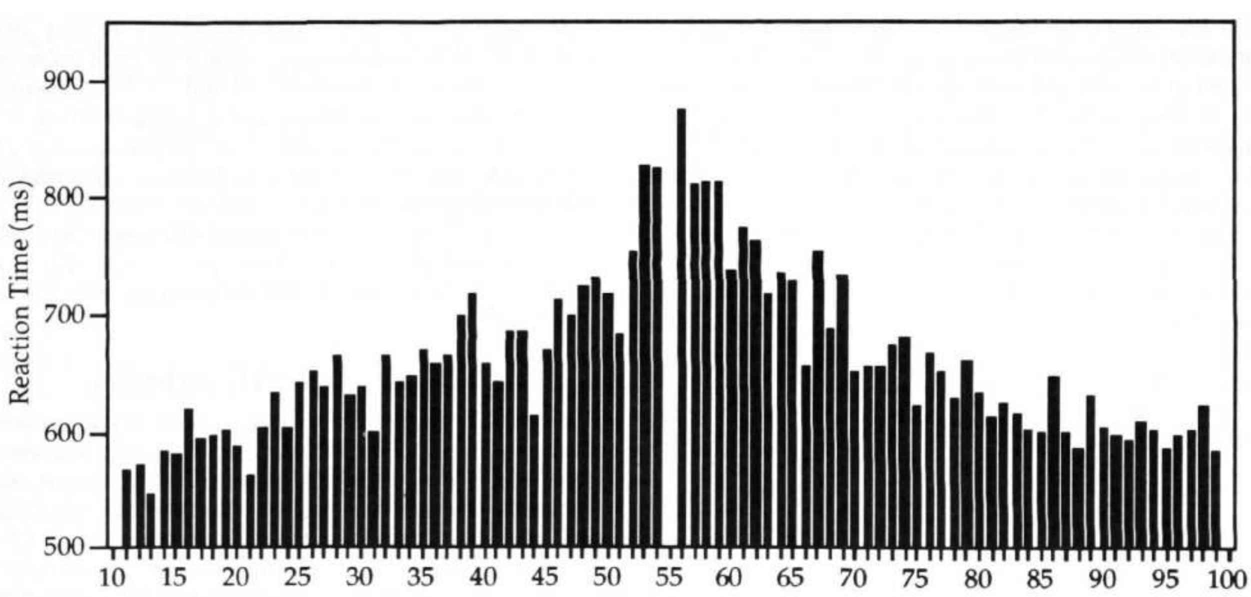

Figure 2. Reaction times for targets compared with 55.

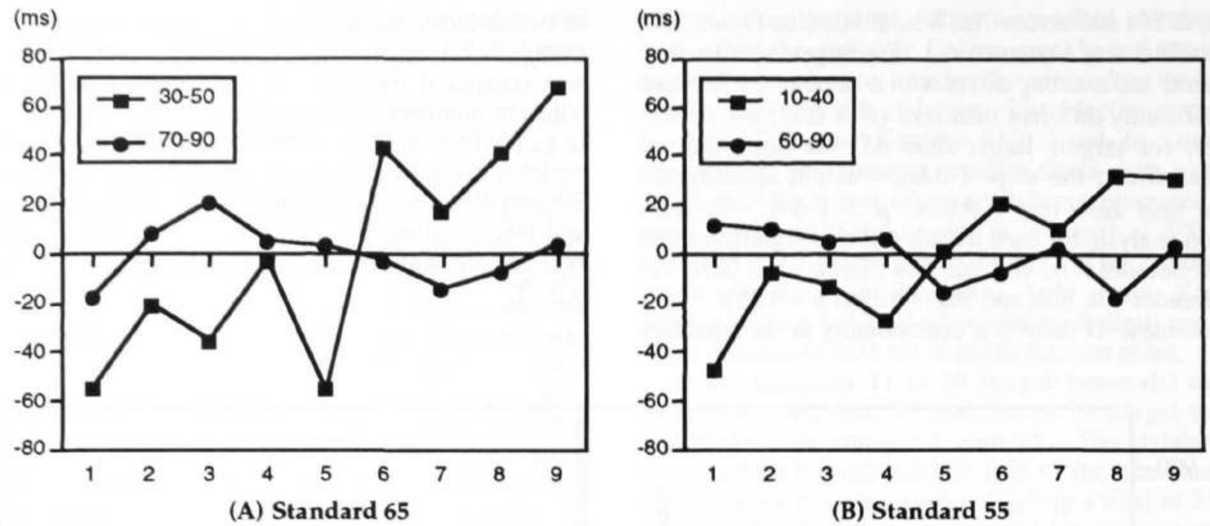

Figure 3. Unit effect for standards 65 and 55. The RT for each target in a decade was subtracted by the mean RT of the corresponding decade, then averaged across decades 30s, 40s, and 50s and across 70s, 80s, and 90s for standard 65 and averaged across 10s, 20s, 30s, and 40s and across 60s, 70s, 80s, and 90s for standard 55.

Congruity Effect. For targets smaller than 65, the RTs for targets with units 1-4 were faster than those with units 6-9 for decade 30s, 40s, and 50s (smallest $F(1, 29) = 23.82$, $p < 0.001$), which was consistent with both a congruity effect and a distance effect. Most importantly, there was a reverse distance effect across the boundary between 30s and 40s: RTs for 36-39 were slower than those for 41-44 ($F(1, 9) = 20.80$, $p < 0.001$)—clear evidence for congruity effect. For targets larger than 65, there was neither a congruity effect nor a distance effect within a decade (largest $F(1, 29) = 0.74$, $p = 0.40$), and there was no reverse distance effect across decade boundaries.

Standard 55. The average RTs to compare target numbers with 55 are shown in Figure 2. There was a marginal asymmetrical range (smaller or larger than 55) effect ($F(1, 16) = 3.84$, $p = 0.07$) and a significant distance effect ($F(43, 688) = 15.23$, $p < 0.001$). The interaction between range and distance was also significant ($F(43, 688) = 2.25$, $p < 0.001$). Separate analyses were conducted for targets smaller than 55 and those larger than 55.

Decade and Unit Effects. For targets 11-49, the decade effect was significant between 10s and 20s and between 20s and 30s (smallest $F(1, 25) = 22.34$, $p < 0.001$), but not between 30s and 40s ($F(1, 26) = 2.38$, $p = 0.13$). For decades 10s, 20s, 30s, and 40s, the unit effect was all significant (smallest $F(8, 224) = 2.27$, $p < 0.02$). For targets 61-99, the decade effect was significant between 60s and 70s and between 70s and 80s (smallest $F(1, 22) = 27.66$, $p < 0.001$), but not between 80s and 90s ($F(1, 24) = 0.50$, $p = 0.49$). The unit effect was significant for 60s ($F(8, 232) = 2.32$, $p = 0.02$) but not for 70s, 80s, and 90s (largest $F(8, 208) = 1.41$, $p = 0.19$).

Linear regression analysis showed an asymmetrical unit effect similar to that found for standard 65 (see Figure 3B): the slope for each decade below 55 was significantly different than zero (largest $p < 0.02$) whereas that above 55 was not (smallest $p = 0.19$).

Discontinuity. 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). There was a significant discontinuity effect between 60s and 70s ($F(1, 30) = 6.75$, $p < 0.01$) and a marginal effect between 50s and 60s ($F(1, 30) = 3.80$, $p = 0.06$). There was no discontinuity effect at other decade boundaries (largest $F(1, 30) = 2.91$, $p = 0.10$).

Congruity Effect. For each decade below 55, the RTs for targets with units 1-4 were larger than those with units 6-9 (smallest $F(1, 28) = 8.93$, $p < 0.006$), which was consistent with both a congruity effect and a distance effect. Although there was no reverse distance effect across decade boundaries, there was no significant difference between 16-19 and 21-24 and between 26-29 and 31-34, implying that there was neither a congruity nor a distance effect, or that there was a congruity effect countering a distance effect. For decades above 55, there was neither a congruity nor a distance effect within 70s, 80s, and 90s (smallest $F(1, 26) = 0.85$, $p = 0.37$) and only a marginal difference between 61-64 and 66-69 ($F(1, 29) = 4.68$, $p < 0.05$). No reverse distance effect across decade boundaries was found for decades above 55.

Discussion and Conclusion

The experimental results showed a complex pattern that can not be easily fitted into a single model. Both standards 65 and 55 showed an asymmetrical range effect: targets smaller and those larger than the standards were compared in different ways. For standard 65, targets larger than 65 were compared sequentially, supported by an insignificant unit effect within each decade, a zero slope of linear regression for each decade, a decade effect, and discontinuities across decade boundaries. Targets smaller than 65 were compared in parallel, supported by a unit effect within each decade and a Stroop-like congruity effect (reverse distance effect). The holistic comparison model can be clearly rejected for the case of standard 65.

For standard 55, though the results were not as clearly cut as those for standard 65, the trend was similar. Targets larger than 55 seemed to be compared sequentially, supported by an insignificant unit effect within each decade (except 60s, which can be accounted for by an abnormal low RT for 66), a zero slope for each decade, a decade effect between 60s and 70s and between 70s and 80s, and discontinuities at the boundaries between 50s and 60s and between 60s and 70s. For targets smaller than 55, a strong unit effect for every decade ruled out sequential comparison but was consistent with both holistic and parallel comparison. Although no reverse distance effect was found, an insignificant difference between 16-19 and 21-24 and between 26-29 and 31-34 but a strong difference between 21-24 and 26-29 implied a congruity effect countering a distance effect. Thus, parallel comparison was more likely than holistic comparison.

In the experiments by Dehaene et al. and Hinrichs et al., the standards were always held in memory and only the targets were presented on an external display, whereas in our experiment the standards and targets were presented and compared on the same external display at the same time. Although the abstract task was the same for their studies and our studies (i.e., comparing 2-digit numerals), the comparison processes were different. This is a demonstration of the representational effect in numerical tasks—different representations of a common abstract structure can cause dramatically different behaviors.

Our experimental results are clearly evidence against the claim that there is a common internal representation for all types of number representations. Different representations are not encoded and transformed into an abstract internal representation. Rather, they activate representation-specific processes. It is these representation-specific processes, not the abstract processes for the abstract representation, that are the actual mechanisms in numerical tasks.

Acknowledgements

This research was supported by a Seed Grant from The Ohio State University. We would like to thank Gwen Hall for his assistance in the experiments.

References

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.

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.

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

Nickerson, R. (1988). Counting, computing, and the representation of numbers. *Human Factors, 30*, 181-199.

Poltrock, S. E. & Schwartz, D. R. (1984). Comparative judgments of multidigit numbers. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 10*, 32-45.

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

Zhang, J. & Norman, D. A. (in press). A representational analysis of numeration systems. To appear in *Cognition*.