Decomposing the problem-size effect: A comparison of response time distributions across cultures

MARCIE PENNER-WILGER, CRAIG LETH-STEENSEN, and JO-ANNE LeFEVRE
Carleton University, Ottawa, Ontario, Canada

Is the locus of the problem-size effect in mental arithmetic different across cultures? In a novel approach to this question, the ex-Gaussian distributional model was applied to response times for large (e.g., 8×9) and small (e.g., 2×3) problems obtained from Chinese and Canadian graduate students in a multiplication production task (LeFevre & Liu, 1997). The problem-size effect for the Chinese group occurred in μ (the mean of the normal component), whereas the problem-size effect for the Canadian group occurred in both μ and τ (the mean of the exponential component). The results support the position that the problem-size effect for the Chinese group is purely a memory-retrieval effect, whereas for the Canadian group, it is an effect of both retrieval and the use of nonretrieval solution procedures.

Understanding the *problem-size effect* (i.e., that large problems, such as 8×9 or $9 + 7$, take longer to solve than smaller problems, such as 3×4 or $2 + 6$) is an integral part of understanding how arithmetic knowledge is represented in the mind. Research on the problem-size effect has shown differences across cultures (Campbell & Xue, 2001; Geary, 1996; LeFevre & Liu, 1997), individuals (Hecht, 1999; LeFevre, Sadesky, & Bisanz, 1996), and problem types (e.g., *ties and fives* effects; LeFevre, Sadesky, & Bisanz, 1996). These complex patterns of results have fuelled considerable controversy about the cause of the problem-size effect (for a review, see LeFevre, Sadesky, & Bisanz, 1996). In the present paper, we undertake a novel analysis of the problem-size effect by examining the shapes of response time distributions for solutions to large and small multiplication problems. In this analysis, we made use of the *ex-Gaussian* distributional model in order to obtain quantitative measures of distributional shape (Heathcote, 1996). The goal of the present research was to obtain a more detailed empirical account of the problem-size effect in order to further evaluate existing theoretical accounts.

The Problem-Size Effect

The problem-size effect, first reported by Groen and Parkman (1972), is the most robust phenomenon in mathematical cognition (Ashcraft, 1992, 1995). As such, any cognitive model of mental arithmetic must provide an ac-

count of this effect. Proponents of the view that adults always retrieve arithmetic facts from memory have suggested two explanations for the problem-size effect. According to learning-based models, the effect is due to differences in the associative strength of small versus large problems, arising from differences in the amount of exposure to problems as they are learned (e.g., Ashcraft, 1992; Hamann & Ashcraft, 1986; Siegler & Shrager, 1984). According to structural accounts, characteristics of the numerical relations among problems, such as magnitude or relative position, are reflected in their mental organization and subsequent retrieval (e.g., Campbell, 1995; Miller, Perlmutter, & Keating, 1984). Thus, larger problems will be retrieved more slowly than smaller problems, because their representations are less differentiated in memory. In both the learning-based and structural accounts, the problem-size effect is due to decreased efficiency of retrieval as problem size increases. In contrast, LeFevre, Sadesky, and Bisanz (1996) posited that different procedure choices for small and large problems (i.e., less retrieval for large problems) could account for the problem-size effect, because many nonretrieval procedures take longer to perform than does direct retrieval. For example, one's solving 6×7 by repeatedly adding 6 is slow and error prone. Similarly, the decomposition of 9×8 into $10 \times 8 - 8$ is likely to be slower than the retrieval of 72 from memory. Thus, the effect might be due to the use of a greater variety of procedures on large than on small problems. In the literature, the source or sources of the problem-size effect remain a matter of debate.

The effect of problem size varies with culture, in that adults educated in North America show a larger problem-size effect than do adults educated in China. This finding is specific to recent generations; problem-size effects were not different across cultures for older generations (Geary et al., 1997). This cross-generation pattern suggests that

experiential differences and not language differences are the locus of the exaggerated effect. North American adults report using a variety of nonretrieval procedures, as well as direct retrieval to solve single-digit problems (Campbell & Timm, 2000; Campbell & Xue, 2001; Geary, Frensch, & Wiley, 1993; Geary & Wiley, 1991; Hecht, 1999; LeFevre, Bisanz, et al., 1996; LeFevre & Morris, 1999; LeFevre, Sadesky, & Bisanz, 1996; cf. Kirk & Ashcraft, 2001). The types of nonretrieval procedures reported for multiplication include visualization of the multiplication table, series (e.g., $4 \times 5 = 5, 10, 15, 20$), repeated addition (e.g., $3 \times 4 = 4 + 4 + 4$), and the use of derived facts (e.g., $3 \times 6 = 3 \times 5 + 3$). In contrast, Chinese adults report exclusive use of retrieval for solving basic multiplication problems (Campbell & Xue, 2001; Geary, 1996).

We hypothesized that the source of the differences in the problem-size effect across Chinese and North American adults is the latter group's use of multiple solution procedures to solve simple multiplication problems. Campbell and Xue (2001) recently showed that the problem-size effect for their Canadian participants could be attributed both to less efficient retrieval and to the use of slower solution procedures on large versus small multiplication problems. For their Chinese participants, Campbell and Xue showed that the problem-size effect could be fully attributed to less efficient retrieval for large versus small problems. These findings were based on patterns of procedure use across cultural groups that were derived from participants' self-reports describing their solutions.

Findings based on participants' self-reports have been challenged, however. Kirk and Ashcraft (2001) found that participants' reports of their arithmetic solution procedures were susceptible to biasing instructions. To avoid potential biasing effects of self-report instructions, we adopted a novel approach to test the hypothesis that group differences in the problem-size effect are, at least in part, related to differential use of solution procedures. We examined the shapes of the distributions of response times obtained from Chinese and Canadian adults. If the patterns of procedure use differ across these two groups, the shapes of these distributions should differ as well.

Ex-Gaussian Distribution

Response time distributions are usually positively skewed. For the purpose of analyses, however, most researchers ignore the underlying shape of the distribution and use either mean or median response times (e.g., LeFevre, Bisanz, et al., 1996; LeFevre, Sadesky, & Bisanz, 1996). Alternatively, researchers sometimes either trim extreme response times (e.g., Campbell & Timm, 2000) or perform scale transformations (e.g., Jaffe-Katz, Budescu, & Wallsten, 1989) on them to induce normality. Data trimming, for example, is typically performed under the assumption that extreme scores can be regarded as random, unrepresentative outlying responses. However, it is also important to consider that the true distribution itself might be skewed. Moreover, if the skew were representative of the underlying process that is being studied, trimming or transform-

ing the data would reduce the validity of the findings. Practically, however, methods to extract robust quantitative measures for nonnormal distributions are unavailable in traditional statistical packages.

The ex-Gaussian distribution (Hockley, 1984; Hohle, 1965; Ratcliff & Murdock, 1976) provides a good fit to most response time data and allows three quantitative summary measures of distributional shape to be obtained. The ex-Gaussian distributional model consists of a normally distributed (Gaussian) leading edge and an exponentially distributed tail. Measures derived from this distribution include: μ, the mean of the normal component, σ, the standard deviation of the normal component; and τ, the mean of the exponential component. These three measures can be used to describe the full profile of a set of response times and, hence, provide much more information than can be obtained from mean response time measures alone (however, note that the mean will always equal $\mu + \tau$). Moreover, Heathcote (1996) has developed a statistical package, RTSYS, that allows researchers to easily obtain values for μ, σ, and τ. Ex-Gaussian analyses have provided insight into cognitive processing and have allowed for more stringent evaluations of existing theories of response time phenomena in a variety of areas (e.g., Balota & Spieler, 1999; Heathcote, Popiel, & Mewhort, 1991; Leth-Steensen, King Elbaz, & Douglas, 2000; Wixted & Rohrer, 1993).

The leading edge of a response time distribution (the general location of which is reflected in the value of μ) is composed of the relatively faster set of response times in the distribution, and the tail (the size of which is reflected in the value of τ) is composed of the relatively slower set of response times. Because the use of direct retrieval of arithmetic facts produces relatively fast responses, whereas the use of nonretrieval procedures produces relatively slow responses (Campbell & Timm, 2000; LeFevre, Bisanz, et al., 1996; LeFevre, Sadesky, & Bisanz, 1996), the value of μ should be reflective of retrieval use, and the value of τ should be reflective of procedure use. Thus, if the size of τ reflects the use of procedures other than retrieval, individuals who are more likely to use such procedures should have larger values of τ than individuals who rely exclusively on retrieval.

The Present Analyses

To test our hypotheses about the origin of the problem-size effect, we reanalyzed data originally published by LeFevre and Liu (1997). LeFevre and Liu collected response times for single-digit multiplication problems from two groups of adults: One group was educated in Canada, and another group was educated in China. In accord with research suggesting that Chinese-educated individuals rely on retrieval, whereas North American-educated individuals use a mixture of procedures, LeFevre and Liu found that the adults educated in China showed smaller effects of problem size on response latencies than did the adults educated in Canada. The results of interviews of the participants suggested that the Chinese participants relied

almost exclusively on retrieval, whereas the Canadian participants used more varied solution procedures. With respect to response time distributions, if the Canadians were using different solution processes than the Chinese adults (i.e., a mix of procedures and retrieval vs. exclusive retrieval), different distributional patterns should emerge for the two groups. More specifically, we hypothesized that the Canadian group would have larger values of τ than would the Chinese group. Second, we hypothesized that, if the exaggerated problem-size effect found for North American relative to Chinese adults is a result of procedure use, the Canadian group would show a problem-size effect in both μ and τ, whereas the Chinese group would show a problem-size effect in only μ.

METHOD

The data analyzed in the present paper were originally collected and reported by LeFevre and Liu (1997).

Participants

Forty graduate students participated in this experiment and received a $10 honorarium. Twenty participants (10 male, 10 female; mean age = 31.1 years) had completed their primary education in the United States ($n = 2$) or in Canada ($n = 18$) and their secondary and postsecondary education in Canada. Twenty participants (10 male, 10 female; mean age = 29.4 years) had completed their primary and secondary education in the People's Republic of China.

Materials

The participants completed three blocks of the 100 multiplication problems from 0×0 to 9×9. Three randomized lists of the 100 problems were created, and the order of presentation of the lists was randomly selected for each participant. In the present analyses, we omitted the 0-operand and 1-operand problems, since these problems can be solved by applying a rule (i.e., anything times 0 is 0, a number times one is itself; Campbell & Xue, 2001). Therefore, three blocks of 64 problems were available for analysis for each participant.

Procedure

For each trial, an asterisk was presented in the center of a computer screen. To start the trial, the participants pressed a microswitch. After a 600-msec blank interval, the problem was presented one line below the location of the asterisk. The problem remained on the screen until the participant responded. Responses were detected by a voice-operated relay connected to the computer. A millisecond-accurate timer on a Digitek I/O board was responsible for timing. A 5-sec time limit was imposed for each trial. After each trial, the experimenter entered the participant's response into the computer. The computer then gave auditory feedback on the correctness of each trial. The problems were presented in Arabic numerals; the Canadian students responded in English, the Chinese students in Mandarin. The same bilingual experimenter tested all participants.

RESULTS

Dependent measures included mean response time, standard deviation, percent error, and the three ex-Gaussian parameters obtained by fitting the ex-Gaussian distribution to the data of individual participants, separately for small and large problems. Small problems were defined as those with products less than or equal to 25 ($n = 32$), and large problems were defined as those with products greater than 25 ($n = 32$). A 2 (problem size: large, small) $\times$ 2 (group: Canadian, Chinese) mixed-factors analysis of variance (ANOVA) was performed on each dependent measure, with group as the between-subjects factor and problem size as the within-subjects factor. Means for each measure are shown in Figure 1 (traditional analyses) and Figure 2 (ex-Gaussian analyses). Of the 7,680 total trials, 70 (0.9%) were invalid, owing to either a premature firing of the voice-activated relay or to a failure to respond within the 5-sec time limit, and 222 (2.9%) of the responses were incorrect, leaving 7,388 (96.2%) correct trials available for analysis.

Traditional Analyses

Mean response time. The participants solved large problems more slowly than small problems (1,584 vs. 1,276 msec) [$F(1,38) = 152.3$, $MS_e = 12,395.4$, $p < .001$]. The Canadian participants solved the problems more slowly than did the Chinese participants (1,558 vs. 1,303 msec) [$F(1,38) = 14.2$, $MS_e = 91,484.3$, $p < .01$]. There was also a problem size $\times$ group interaction
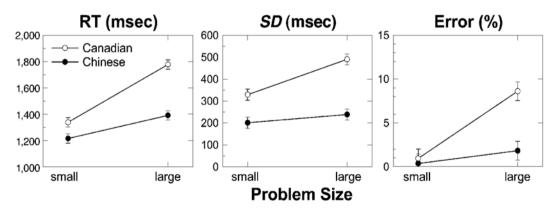

Figure 1. Traditional measures as a function of problem size for Canadian and Chinese participants. Error bars indicate the 95% confidence intervals based on the MS_e values for the illustrated two-way interactions, calculated by using the techniques described by Loftus and Masson (1994).
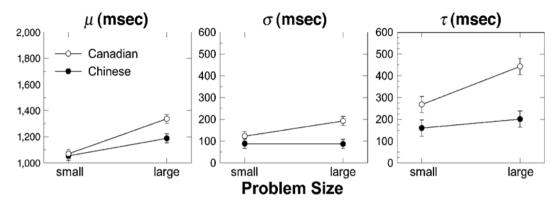
Figure 2. Ex-Gaussian measures as a function of problem size for Canadian and Chinese participants. Error bars indicate the 95% confidence intervals based on the MS_e values for the illustrated two-way interactions, calculated by using the techniques described by Loftus and Masson (1994).

$[F(1,38) = 28.3, MS_e = 12,395.3, p < .001]$. As is shown in Figure 1, the Canadian participants showed a larger problem-size effect than did the Chinese participants.

Standard deviation. Response times were more variable on the large than on the small problems (364 vs. 265 msec) $[F(1,38) = 31.4, MS_e = 6,263.9, p < .001]$. Response times of the Canadian participants were more variable than those of Chinese participants (409 vs. 219 msec) $[F(1,38) = 34.0, MS_e = 21,100.7, p < .001]$. There was also a problem size × group interaction $[F(1,38) = 12.3, MS_e = 6,263.9, p < .01]$. As is shown in Figure 1, the Canadian participants showed a problem-size effect, whereas the Chinese participants did not.

Percent error. The percentage of errors made was fairly low, perhaps owing to an emphasis on accuracy in the instructions. The participants made more errors when solving large problems than small problems (5.2 vs. 0.6%) $[F(1,38) = 36.7, MS_e = 11.3, p < .001]$. The Canadian participants made more errors than did the Chinese participants (4.7% vs. 1.1%) $[F(1,38) = 21.1, MS_e = 12.8, p < .001]$. There was also a problem size × group interaction $[F(1,38) = 17.0, MS_e = 11.3, p < .001]$. As is shown in Figure 1, the Canadian participants showed a problem-size effect, whereas the Chinese participants did not.

Ex-Gaussian Analyses

Model fits. Goodness-of-fit measures for the ex-Gaussian distributional model in terms of a chi-square statistic were provided by RTSYS, with significant chi-square values indicating poor fit. Chi-square tests were significant ($p < .05$) for only 6 the 40 fits for the Canadian participants and for only 4 the 40 fits for the Chinese participants. Ex-Gaussian parameter values for cases in which the ex-Gaussian model provided a poor fit to the distributional data were included in the following analysis, because these values still provide meaningful quantitative information about the general shapes of the distributions.

μ. The mean of the ex-Gaussian normal component was greater for the large problems than for the small problems (1,263 vs. 1,062 msec) $[F(1,38) = 71.4, MS_e =$

11,326.2, $p < .001$]. In contrast to the mean response times, however, μ did not differ significantly across groups $[F(1,38) = 3.3, MS_e = 41,089.4, p > .05]$. There was a problem size × group interaction $[F(1,38) = 7.9, MS_e = 11,326.2, p < .01]$. As is shown in Figure 2, the Canadian participants showed a larger problem-size effect than did the Chinese participants.

σ. The variability of the ex-Gaussian normal component was greater for the large problems than for the small problems (139 vs. 105 msec) $[F(1,38) = 5.3, MS_e = 4,324.9, p < .05]$. σ was greater for the Canadian participants than for the Chinese participants (156 vs. 87 msec) $[F(1,38) = 21.2, MS_e = 4,599.4, p < .001]$. There was also a problem size × group interaction $[F(1,38) = 5.6, MS_e = 4,324.9, p < .05]$. As is shown in Figure 2, the Canadian participants showed a problem-size effect, whereas the Chinese participants did not.

τ. The mean of the ex-Gaussian exponential component was greater for large problems than for small problems (322 vs. 214 msec) $[F(1,38) = 16.8, MS_e = 13,774.3, p < .001]$. τ was greater for the Canadian than for the Chinese participants (356 vs. 181 msec) $[F(1,38) = 18.7, MS_e = 32,547.0, p < .001]$. There was also a problem size × group interaction $[F(1,38) = 6.5, MS_e = 13,774.3, p < .05]$. As is shown in Figure 2, the Canadian participants showed a problem-size effect, whereas the Chinese participants did not.

Vincentized group response time distributions. Group response time distributions were derived for the Canadian and Chinese groups for both small and large problems by using the Vincent averaging technique recommended by Ratcliff (1979; see also Heathcote, 1996). These distributions are displayed in Figure 3 as Vincent histograms. Within such histograms, distributions of observations are represented as a series of rectangles of equal area. The solid curves in Figure 3 represent the ex-Gaussian fits to each of those distributions.

The group differences in response time distributional patterns that are evident in Figure 3 correspond to the reported group differences in the three ex-Gaussian para-
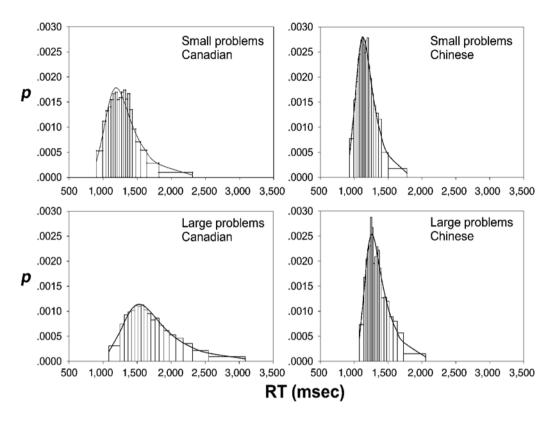
Figure 3. Group probability density histograms for distributions of response times to small and large problems for Canadian and Chinese participants. Solid lines represent the fitted ex-Gaussian distribution functions.

meters. The Canadian distributions (left-hand panel) are less peaked, with a substantial amount of positive skew. In contrast, the Chinese distributions (right-hand panel) are highly peaked, with little positive skew. As problem size increases (from the top to the bottom plots), the Canadian distributions show both a shift in the leading edge and a large increase in the size of the tail. In contrast, the Chinese distributions show only a shift in the leading edge, without any notable change in distributional shape.

DISCUSSION

The goal of the present paper was to explore differences in the source of the problem-size effect across cultures. As hypothesized, the shapes of the response time distributions for solution performances to single-digit multiplication problems, as reflected in the obtained values of the three ex-Gaussian parameters, differed across the Chinese and Canadian participants. Namely, the size of the tails of the distributions (i.e., τ) was much larger for the Canadian than for the Chinese participants. This finding is consistent with the conclusion derived from self-report data (e.g., Campbell & Xue, 2001) that these two groups are using different combinations of solution procedures. Moreover, because Chinese participants typically report using retrieval exclusively and the Canadian participants typically report using a mix of retrieval and nonretrieval

procedures, our findings suggest that large values of the ex-Gaussian τ parameter could be used to provide an alternative index (i.e., to self-report) of nonretrieval procedure use.

Also as hypothesized, the manner in which the problem-size effect was manifested in the shapes of the response time distributions differed for the Chinese and Canadian participants. Namely, we found that the problem-size effect for the Chinese participants occurred in μ alone, whereas for the Canadian participants, it occurred in both μ and τ. If the Chinese participants relied solely on retrieval, any problem-size effect for them should have been due entirely to differences in retrieval efficiency for large and small problems (Campbell & Xue, 2001; Geary, 1996). Hence, our findings suggest that shifts in μ over problem size are indicative of a decreased efficiency of retrieval processes for large versus small problems. The finding that the Canadian participants showed an even larger problem-size effect in μ than did the Chinese participants implies that retrieval processes on large problems are even less efficient for the Canadian participants than for the Chinese participants. However, this enhanced μ effect for the Canadian participants was also accompanied by a problem-size effect in the ex-Gaussian parameter σ (there was no corresponding effect in σ for the Chinese participants). As noted by one reviewer, the location of the leading edge of a response time distribution can be more precisely defined

by the value $[\mu - (1.96 \times \sigma)]$, because it approximates the 2.5 percentile point. Hence, the locations of the leading edges for the large problems do not actually differ greatly between the Canadian and the Chinese participants (see Figure 3). Finally, the increase in τ across problem size for the Canadian participants is entirely consistent with the notion that a substantial part of the problem-size effect for these participants is due to an increased use of less efficient procedures when they solve large problems (Campbell & Xue, 2001; Geary, 1996).

Why do Canadian-educated adults show differential patterns of performance on simple multiplication problems compared with Chinese-educated adults? Researchers have pointed to differences in educational practices, such as greater amounts of practice of basic facts (Geary, 1996), increased emphasis on larger problems that, in grade 3 students, results in a reversed problem-size effect (Geary, 1996), and a reduced reliance on calculators in Chinese relative to North American schools (Campbell & Xue, 2001; LeFevre & Liu, 1997). Also, the finding that Chinese adults still show a problem-size effect despite their reporting exclusive use of retrieval suggests that procedure choices do not fully capture the problem-size effect. Thus a role for structural factors (Geary, 1996) or magnitude effects (Campbell, 1995) in the problem-size effect is maintained. In accord with this view, LeFevre, Sadesky, and Bisanz (1996; LeFevre, Bisanz, et al., 1996) showed that the problem-size effect was smaller, but still statistically significant, when it was based only on problems on which participants reported using retrieval, as compared with when latencies were averaged across procedures (cf. Siegler, 1987). Thus, the present research is consistent with the view that the problem-size effect arises from a variety of sources that depend on culturally linked factors, such as practice and selection of solution procedures, as well as on structural factors, such as magnitude, that are inherent to the mental representation of numbers.

It is also important to emphasize that the ex-Gaussian distributional model has been used to fit these response time data, simply because it provides useful quantitative measures of distributional shape. Its use here is not meant to imply that it represents a theoretical model of the response processes involved in solving single-digit multiplication problems. We believe that a more appropriate theoretical model for response times would, in fact, involve a probabilistic mixture of separate (relatively faster) retrieval and (relatively slower) nonretrieval distributions. A formal characterization of such a model for the four factorial conditions of our study is presented in the Appendix.

Finally, note that these retrieval and nonretrieval distributions each represent mixtures of solution times for the individual problems subsumed within the small- and large-problem categories. Consequently, increases in τ over problem size could be observed if solution retrieval times for the individual problems increased in an expansive nonlinear fashion with the increase in problem size (especially given that there is a greater range of product sizes within the large-problem category than in the small-problem category). If there is a nonlinear relation between retrieval latency and problem size that is greater for Canadian adults than for Chinese adults, it could also account for the observed differences in τ across the groups. Furthermore, if the retrieval of solutions from memory could be assumed to take place within a diffusion process (i.e., a continuous random walk) of the type proposed by Ratcliff (1978), decreases in evidence-accumulation rates (e.g., for large problems or Canadian individuals) could lead to corresponding increases in both distributional location and skew. Relatedly, within the context of an explicit arithmetic fact retrieval model, such as Campbell's (1995) network interference model, it is conceivable that increases in distributional skew could result from specific cases in which the activated node for the correct answer takes an extraordinarily long time to resolve the interference from other answer nodes. Hence, the results of the present analyses do not specifically rule out the possibility that differences in the efficiency of retrieval processes between the two groups alone could have produced the observed pattern of differential problem-size effects in τ. However, these alternative interpretations are mitigated by the finding that, for the Chinese adults, τ did not significantly change across problem-size conditions. In summary, a demonstration of convergence among alternative measures of procedure use would strengthen the conclusion that the size of τ provides an index of nonretrieval procedure usage.

CONCLUSION

The present findings lend support to the position that the problem-size effect is due to different factors across cultures (Campbell & Xue, 2001) and is due in part to procedure use for Canadian adults (Campbell & Xue, 2001; Hecht, 1999; LeFevre, Bisanz, et al., 1996; LeFevre, Sadesky, & Bisanz, 1996). Data on the use of procedures by North American adults has generally been collected with self-report techniques (Campbell & Timm, 2000; Hecht, 1999; LeFevre, Bisanz, et al., 1996; LeFevre, Sadesky, & Bisanz, 1996). Recent research has pointed to potential shortcomings in the self-report technique (Kirk & Ashcraft, 2001). Kirk and Ashcraft suggested that participants are reactive to the demands involved in providing self-reports and therefore data from self-reports are limited in their ability to illuminate cognitive processes. Although complementary analyses of response times have provided some evidence that self-reports are indeed veridical (LeFevre, Bisanz, et al., 1996; LeFevre, Smith-Chant, Hiscock, Daley, & Morris, in press), Smith-Chant and LeFevre (in press) found evidence for reactivity that varied with skill and instructional demands. Thus, it is important to develop alternative techniques for assessing variability in the use of solution procedures. The ex-Gaussian τ parameter shows promise as an objective indicator of procedure use; one that avoids the potential bias engendered by asking participants to describe their solution procedures.

REFERENCES

ASHCRAFT, M. H. (1992). Cognitive arithmetic: A review of data and theory. *Cognition*, **44**, 75-106.

ASHCRAFT, M. H. (1995). Cognitive psychology and simple arithmetic: A review and summary of new directions. *Mathematical Cognition*, **1**, 3-34.

BALOTA, D. A., & SPIELER, D. H. (1999). Word frequency, repetition, and lexicality effects in word recognition tasks: Beyond measures of central tendency. *Journal of Experimental Psychology: General*, **128**, 32-55.

CAMPBELL, J. I. D. (1995). Mechanisms of simple addition and multiplication: A modified network-interference theory and simulation. *Mathematical Cognition*, **1**, 121-164.

CAMPBELL, J. I. D., & TIMM, J. C. (2000). Adults' strategy choices for simple addition: Effects of retrieval interference. *Psychonomic Bulletin & Review*, **7**, 692-699.

CAMPBELL, J. I. D., & XUE, Q. (2001). Cognitive arithmetic across cultures. *Journal of Experimental Psychology: General*, **130**, 299-315.

GEARY, D. C. (1996). The problem-size effect in mental addition: Developmental and cross-national trends. *Mathematical Cognition*, **2**, 63-93.

GEARY, D. C., FRENSCH, P. A., & WILEY, J. G. (1993). Simple and complex mental subtraction: Strategy choice and speed-of-processing differences in young and elderly adults. *Psychology & Aging*, **8**, 242-256.

GEARY, D. C., HAMSON, C. O., CHEN, G.-P., LIU, F., HOARD, M. K., & SALTHOUSE, T. A. (1997). Computational and reasoning abilities in arithmetic: Cross-generational change in China and the United States. *Psychonomic Bulletin & Review*, **4**, 425-430.

GEARY, D. C., & WILEY, J. G. (1991). Cognitive addition: Strategy choice and speed-of-processing differences in young and elderly adults. *Psychology & Aging*, **6**, 474-483.

GROEN, G. J., & PARKMAN, J. M. (1972). A chronometric analysis of simple addition. *Psychological Review*, **79**, 329-343.

HAMANN, M. S., & ASHCRAFT, M. H. (1986). Textbook presentations of the basic addition facts. *Cognition & Instruction*, **3**, 173-192.

HEATHCOTE, A. (1996). RTSYS: A DOS application for the analysis of reaction time data. *Behavior Research Methods, Instruments, & Computers*, **28**, 427-445.

HEATHCOTE, A., POPIEL, S. J., & MEWHORT, D. J. K. (1991). Analysis of response time distributions: An example using the Stroop task. *Psychological Bulletin*, **109**, 340-347.

HECHT, S. A. (1999). Individual solution processes while solving addition and multiplication math facts in adults. *Memory & Cognition*, **27**, 1097-1107.

HOCKLEY, W. E. (1984). Analysis of response time distributions in the study of cognitive processes. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, **6**, 598-615.

HOHLE, R. H. (1965). Inferred components of reaction time as a function of foreperiod duration. *Journal of Experimental Psychology*, **69**, 382-386.

JAFFE-KATZ, A., BUDESCU, D. V., & WALLSTEN, T. S. (1989). Timed magnitude comparisons of numerical and nonnumerical expressions of uncertainty. *Memory & Cognition*, **17**, 249-264.

KIRK, E. P., & ASHCRAFT, M. H. (2001). Telling stories: The perils and promise of using verbal reports to study math strategies. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, **27**, 157-175.

LEFEVRE, J.-A., BISANZ, J., DALEY, K. E., BUFFONE, L., GREENHAM, S. L., & SADESKY, G. S. (1996). Multiple routes to solution of single-digit multiplication problems. *Journal of Experimental Psychology: General*, **125**, 284-306.

LEFEVRE, J.-A., & LIU, J. (1997). The role of experience in numerical skill: Multiplication performance in adults from Canada and China. *Mathematical Cognition*, **3**, 31-62.

LEFEVRE, J.-A., & MORRIS, J. (1999). More on the relation between division and multiplication in simple arithmetic: Evidence for mediation of division solutions via multiplication. *Memory & Cognition*, **27**, 803-812.

LEFEVRE, J.-A., SADESKY, G. S., & BISANZ, J. (1996). Selection of procedures in mental addition: Reassessing the problem size effect in adults. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, **22**, 216-230.

LEFEVRE, J.-A., SMITH-CHANT, B. L., HISCOCK, K., DALEY, K. E., & MORRIS, J. (in press). Young adults' strategic choices in simple arithmetic: Implications for the development of mathematical representations. In A. J. Baroody & A. Dowker (Eds.), *The development of arithmetic concepts and skills: Recent research and theory*. Hillsdale, NJ: Erlbaum.

LETH-STEENSEN, C., KING ELBAZ, Z., & DOUGLAS, V. I. (2000). Mean response times, variability, and skew in the responding of ADHD children: A response time distributional approach. *Acta Psychologica*, **104**, 167-190.

LOFTUS, G. R., & MASSON, M. E. J. (1994). Using confidence intervals in within-subjects designs. *Psychonomic Bulletin & Review*, **1**, 476-490.

MILLER, K., PERLMUTTER, M., & KEATING, D. (1984). Cognitive arithmetic: Comparison of operations. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, **10**, 46-60.

RATCLIFF, R. (1978). A theory of memory retrieval. *Psychological Review*, **85**, 59-108.

RATCLIFF, R. (1979). Group reaction time distributions and an analysis of distribution statistics. *Psychological Bulletin*, **86**, 446-461.

RATCLIFF, R., & MURDOCK, B. B. (1976). Retrieval processes in recognition memory. *Psychological Review*, **83**, 190-214.

SIEGLER, R. S. (1987). The perils of averaging over strategies: An example from children's addition. *Journal of Experimental Psychology: General*, **116**, 250-264.

SIEGLER, R. S., & SHRAGER, J. (1984). Strategy choices in addition and subtraction: How do children know what to do? In C. Sophian (Ed.), *Origins of cognitive skills* (pp. 229-293). Hilllsdale, NJ: Erlbaum.

SILVERMAN, B. W. (1986). *Density estimation for statistics and data analysis*. London: Chapman & Hall.

SMITH-CHANT, B. L., & LEFEVRE, J.-A. (in press). Doing as they are told and telling it like it is: Self-reports in mental arithmetic. *Memory & Cognition*.

VAN ZANDT, T. (2000). How to fit a response time distribution. *Psychonomic Bulletin & Review*, **7**, 424-465.

WIXTED, J. T., & ROHRER, D. (1993). Proactive interference and the dynamics of free recall. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, **19**, 1024-1039.

YANTIS, S., MEYER, D. E., & SMITH, J. E. K. (1991). Analyses of multinomial mixture distributions: New tests for stochastic models of cognition and action. *Psychological Bulletin*, **110**, 350-374.

APPENDIX

We suggest that the response time distributions for the four experimental conditions in the present study can be characterized more formally in terms of probabilistic mixtures according to the following set of equations:

$$RT_{Ca, Sm} = p_{R, Ca, Sm} RT_{R, Ca, Sm} + p_{NR, Ca, Sm} RT_{NR, Ca, Sm}, \tag{A1}$$

$$RT_{Ca, Lg} = p_{R, Ca, Lg} RT_{R, Ca, Lg} + p_{NR, Ca, Lg} RT_{NR, Ca, Lg}, \tag{A2}$$

$$RT_{Ch, Sm} = RT_{R, Ch, Sm}, \tag{A3}$$

$$RT_{Ch, Lg} = RT_{R, Ch, Lg}. \tag{A4}$$

Where, within each equation, the mixture probabilities (i.e., the p_R's for retrieval and the p_{NR}'s for nonretrieval) have the property that $p_R = 1 - p_{NR}$ with $p_R > p_{NR}$, and where Equations A3 and A4 represent the fact that for the Chinese adults, p_{NR} could be assumed to be equal to (or, at least, very close to) 0 for both problem sizes.

Given a set of binary mixture distributions, methods are available for estimating the values of the various mixture probabilities (Yantis, Meyer, & Smith, 1991). However, these methods are applicable to experimental situations in which the underlying distributions that are being mixed are the same across the conditions of an experiment, and it is only the mixture probabilities that are varying. For these data, this assumption seems tenable. For example, the differences in μ across both culture and problem size suggest that the locations of underlying retrieval distributions (i.e., the RT_Rs) are not the same within the four conditions (with the possible exception that, perhaps, $RT_{R, Ca, Sm} = RT_{R, Ch, Sm}$). Similarly, given that the nature and efficiency of the nonretrieval strategies used by the Canadian adults should vary markedly across small and large problems, it seems rather unlikely that $RT_{NR, Ca, Sm} = RT_{NR, Ca, Lg}$.

Although it does not necessarily have to be the case, one signature of a binary mixture distribution is the presence of bimodality (Yantis et al., 1991). Unfortunately, the use of the ex-Gaussian distributional model here does not allow for the detection of bimodality, because all of the fits to the data are constrained to take the functional form of the ex-Gaussian (a shortcoming shared by all other parametric distributional models, except the mixture model). An alternative method would involve the use of nonparametric density estimation techniques (Van Zandt, 2000). Such techniques can provide accurate descriptive characterizations of the true underlying shape of a distribution. They do not, however, provide quantitative measures of distributional shape that will allow for the kinds of quantitative comparisons that are made here with the use of the three ex-Gaussian parameters (although there are some quantitative methods available; see, e.g., Silverman, 1986).