Complex Mental Arithmetic: The Contribution of the Number Sense

Julie Nys and Alain Content
Université Libre de Bruxelles

Young adults were asked to solve two-digit addition problems and to say aloud the result of each calculation step to allow the identification of computation strategies. We manipulated the position of the largest addend (e.g., 25 + 48 vs. 48 + 25) to assess whether strategies are modulated by magnitude characteristics. With some strategies, participants demonstrated a clear preference to take the largest addend as the starting point for the calculation. Hence, rather than applying strategies in an inflexible manner, participants evaluated and compared the operands before proceeding to calculation. Further, mathematically skilled participants tended to use those magnitude-based strategies more often than less skilled ones. The findings demonstrate that magnitude information plays a role in complex arithmetic by guiding the process of strategy selection, and possibly more so for mathematically skilled participants.

Keywords: approximate number system, mathematical ability, mental arithmetic, number sense, strategy

There is now a large body of evidence showing that humans as well as several other animal species possess a specific representational system for quantity, which provides an approximate code for abstract numerical magnitude. This approximate number system is considered as the foundation of the “number sense” (Dehaene, 1997, 2001), a cover term accounting for the core nonverbal abilities allowing to quickly apprehend, estimate, and roughly manipulate numerosities. How the approximate number sense relates to arithmetical abilities is however an open question. The present research aimed at examining the role of the number sense in complex arithmetic by investigating whether the relative magnitude of operands influences the way strategies are selected and applied. Evidence for an influence of the magnitude would be indicative of a contribution of the number sense, because it encompasses magnitude apprehension and comparison.

How do we solve problems such as “48 + 25”? Previous studies have shown that many strategies can be used, and that different people choose amongst different strategies for different problems (e.g., Blöte, Klein, & Beisiuizen, 2000; Fuson et al., 1997; Green, Lemaire, & Dufau, 2007; Heidsfield, 2000; Lemaire & Arnaud, 2008; Lucangeli, Tressoldi, Bendotti, Bonanomi, & Siegel, 2003). Amongst problem features, carry, problem size and parity have been shown to influence arithmetical processing (e.g., Fürst & Hitch, 2000; Imbo, Vandierendonck, & de Rammelaere, 2007; LeFevre, Sadesky, & Bisanz, 1996; Lemaire & Reder, 1999).

Surprisingly, other features such as the position of the largest addend, or more generally the magnitude relations across operands, have hardly been examined despite indications that they might influence the selection and application of calculation strategies. With single-digit addition problems, young children generally prefer to add the value of the smallest addend to the other digit (Groen & Parkman, 1972). This technique, known as the Min strategy, implies that the magnitude of the addends was apprehended precociously during processing and determined the execution of the incrementation. Interestingly, Butterworth, Zorzi, Girelli, and Jonckheere (2001) produced evidence for an addend-comparison stage in adults’ arithmetical facts retrieval (but see Robert & Campbell, 2008).

Because magnitude characteristics have an impact on the strategies used to solve simple arithmetic problems in children, and possibly on the organisation of the arithmetical facts network in adults as suggested by Butterworth et al. (2001), they might also intervene in complex arithmetic. Indeed, Trbovich and LeFevre (2003) manipulated operand order and found that, with horizontal presentation, participants were faster to solve two-digit + one-digit problems (e.g., 52 + 3) than the reverse (3 + 52). Although the study did not examine calculation strategies, this finding suggests that the relative magnitude of operands may influence processing. Here, we wondered whether strategies are influenced by magnitude in adults. In a pilot study with complex additions, participants reported preferring to start the calculation from the largest of the two operands rather than systematically from the left (or right) one. Therefore, in the present experiment, we varied the position of the largest addend in a controlled way by presenting the problems twice, as $a + b$ and $b + a$, to test whether adults evaluate magnitude and use the largest operand as calculation anchor, whatever its position.

Strategies were identified by requiring participants to verbalise online the result of each calculation step. For instance, when solving the “48 + 25” problem, one participant would successively produce 68, then 73. Another might say 50, 75, 73, or 60, 13, 73. We assumed that the Intermediate Results Verbalization (hereafter, IRV) technique would explicitly reveal mental calculation
steps without imposing a strong cognitive overload that would modify the course of processing (Kirk & Ashcraft, 2001; Smith-Chant & LeFevre, 2003), while at the same time avoiding the risk of post hoc reevaluation (Ericsson & Simon, 1980). Latencies to the first intermediate result, as an index of the time required for strategy choice and initialization, and IRV sequences were recorded. Participants also received a standard mathematical test to evaluate the relation between mathematical ability and strategic preferences.

Method

Participants

Fifty undergraduate French-speaking students at the Université libre de Bruxelles (mean age = 20.96 years, 25 men) took part. All had normal or corrected vision. Most of them had received their education in the French Community of Belgium or in France.

Stimuli

Forty-two pairs of addition problems were used. All contained two-digit operands and required a carry. Addends were sampled across all tens (excluding the tens themselves) to avoid biases in the response range. In each pair, the position of the largest addend was manipulated by reversing the order of the two addends, so that the largest was located to the right or left side of the sign (see Table 1). Unit-decade compatibility effects have been reported in Arabic number comparison studies (e.g., Nuerk, Weger, & Willmes, 2001), such that interference was observed when the comparison of the tens and of the units did not match (for instance, 64 vs. 37 slower than 67 vs. 34). Here, we only used compatible addends, so that the unit of the largest addend was always larger than the unit of the smallest addend. Moreover, we modulated the salience of the position manipulation, by varying the numerical distance between the addends. For small distance problems, the difference between addends was smaller than 40; for large distance problems, it was larger than 40. Distance was partly confounded with problem size, which ranged between 30 and 140 and from 70 to 150, respectively (r = .58).

The set of problems was divided into three blocks, in such a way that each block comprised equal numbers of the four kinds of problems. Paired problems appeared in different blocks.

Procedure

Testing occurred individually in a quiet room and required about 40 min. The word-problems subtest of the WAIS (“Arithmetic subtest,” Wechsler, 1989) was administered first, as a measure of mathematical ability. Then, participants were asked to mentally solve the experimental addition problems. They were asked to say the result of each calculation step as soon as possible. One training block (14 trials) was then passed, followed by three blocks of 28 trials presented in a fixed pseudorandom order. Block order was varied across participants and short breaks were allowed between blocks.

Presentation of the stimuli and timing were controlled by PsyScope running on an Apple Macintosh computer. Stimuli were displayed at the centre of the screen using Arial 120 and participants were seated at a comfortable viewing distance (~50 cm). Each trial started with a central fixation point (a “X” sign) during 600 ms. Then, the problem appeared in a horizontal configuration, and remained on the screen until the end of the trial. The experimenter recorded IRV sequences as well as final responses. Response latencies from the appearance of the stimulus to the first intermediate result produced were collected through a microphone and voice key. When ready, the experimenter pressed the space key to initiate the next trial after an 800 ms blank interval.

Results

An α level of .05 was used for all statistical tests. Overall, participants made very few errors when solving the problems (M = 7.1%, SD = 7.4). Error rates were entered in a within-subject ANOVA with Position (left vs. right largest addend) and Distance (small vs. large). The analysis revealed no significant effect (Fs < 1).1

Regarding calculation processes, we first analysed the latencies of first verbalizations. All latencies above 500 ms, except voice-key malfunctions, were taken into account (8% discarded), whatever the final response. The ANOVA revealed a significant effect of Position, F(1, 49) = 5.90, MSE = 317.873, p = .019, η² = .11. Indeed, latencies were shorter when the largest addend was located on the left side of the operation sign (2,871 vs. 3,065 ms). No other significant effect was found (Fs < 1).

Based on previous studies (Blöte et al., 2000; Heirdsfield, 2000), we classified the IRV sequences into four categories (Aggregation, Rounding, Tens & Units, and Pen & Paper). Virtually all responses (97.7%) could be categorized according to that scheme. The four categories generally entailed distinct IRV sequences (see Table 2), so there was little ambiguity in the classification.2 Tens & Units was the most used strategy in terms of overall frequency (34.9%, 24.9%, 23.7%, and 14.3%, respectively, for Tens & Units, Aggregation, Rounding, and Pen & Paper), as well as in terms of number of participants (28, 19, 20, and 11 participants used the respective strategies at least once). About 60% of the participants used only one strategy (11, 6, 6, and 6, respectively), 30% used two, with virtually all possible combinations, and few showed evidence of using more than two strategies.

To assess whether problem characteristics influence strategy selection, we conducted ANOVAs on the percentage of use of each strategy, with Position and Distance as within-subject factors. The only significant effect was that Rounding was employed more frequently when the distance was large (26.0%) than when it was small (21.3%), F(1, 49) = 11.52, MSE = 95.9, p = .001, η² = .19.

---

1 This analysis and the following ones were also run with Gender as an additional factor. No significant effect of Gender was observed, and the inclusion of Gender did not change any of the conclusions.

2 Interrater agreement on a sample of 1,680 items (20 participants) was 96.8%.
To assess whether problem characteristics influence strategy application, Aggregation and Rounding were further distinguished as different IRV sequences occur depending on the anchor addend, the operand to which the strategy is applied. In Left-anchored Aggregation, the left addend is used without decomposition (e.g., 48 + 25 would produce the sequence “68, 73”); whereas the opposite happens in Right-anchored Aggregation (i.e., starting from 25: “65, 73”). Similarly, Rounding was split into two variants, which correspond, respectively, to the rounding up of the left (“50, 75, 73”) or right addend (“30, 78, 73”). Overall, left-anchored strategies were used more frequently than right-anchored strategies, suggestive of a default left-to-right preference (35.3 vs. 14.6% for Left- and Right-anchored Aggregation; 25.3 vs. 22.0% for Left- and Right-anchored Rounding).

However, over and above this default preference, participants who used these two strategies tended to start quite systematically from the largest operand. As shown in Figure 1, Left-anchored Aggregation was used more when the largest addend was on the left (23.5%) than when it was on the right (11.8%), and the opposite trend was observed for Right-anchored Aggregation (1.3 vs. 13.3%). Similarly for Rounding (see Figure 2), the left-anchored strategy prevailed when the largest addend was on the left (24.1%) relative to the right (1.2%) and the mirror pattern came out for Right-anchored Rounding (0.7 vs. 21.3%, respectively). As a result, 74% of Aggregation IRVs and 96% of Rounding IRVs were based on the largest addend, and both rates were significantly higher than the value expected if participants started from either addend at random, t(18) = 2.45; p = .025; t(19) = 45.60; p < .001 for Aggregation and Rounding, respectively. The magnitude-based preferences did not vary significantly according to distance, both Fs ~1.

Because unit values were systematically larger in the largest addend, one might wonder whether the magnitude-based preferences described above rely on the comparison of the addends values themselves or of their respective units magnitudes. To disentangle the two factors, we ran multiple regression analyses on the percentage of use of Aggregation and Rounding across problems, using the Addends Difference and the Units Difference as predictors. The Addends Difference (left addend value minus right addend value) accounts for the distance between operands and for the position of the largest addend (positive difference when the left operand is the largest, negative otherwise), and similarly for the Units Difference. Hence, if the trend to start from the largest operand occurs because of a comparison of the unit values rather than of the addends values, Units Difference should come out rather than Addends Difference. Furthermore, as the distance manipulation was partially confounded with problem size, the Addends Sum was entered as an additional predictor to capture problem magnitude. Addends Difference, Units Difference, and Addends Sum were entered simultaneously in the analyses.

Each model explained about 75% of the variance, with independent contributions of both Units Difference and Addends Difference (see Table 3). The β coefficients were positive for Left-
anchored strategies, indicating that their use augmented with both larger left addends and larger left units. Conversely, the $\beta$ were negative for Right-anchored strategies, suggesting that Right-anchored Aggregation and Rounding were increasingly used with larger right addend and unit values. In sum, the regressions confirm that Aggregation and Rounding are magnitude-based strategies in that they are sensitive to the relative magnitude of both the addends and the units.

Finally, to assess whether strategy selection was related to mathematical skill, we examined the correlations between the percentage of use of each strategy and the raw score obtained in the WAIS subtest (Raw score: $M = 13.4; SD = 4.7$; Standardised score: $M = 9.7; SD = 2.6$). No correlation was found for the Tens & Units strategy ($r = .009, p = .95$). The Pen & Paper strategy was negatively correlated with the WAIS subtest ($r = -.45, p = .001$), and there was a nonsignificant trend toward a positive correlation for Aggregation ($r = .26, p = .07$), as well as for Rounding ($r = .11, p = .45$). When the two magnitude-based strategies were considered together, a positive correlation ($r = .31, p = .03$) was observed, suggesting that mathematically skilled participants were inclined to use magnitude-based strategies more often.

**Discussion**

Does the number sense contribute to complex mental arithmetic? Our findings lead to a positive answer by demonstrating that participants use magnitude information to guide and adapt the application of their calculation strategies. The position of the largest addend was one critical determinant of the way Aggregation and Rounding were executed, so that participants choose to start calculation from the largest addend most of the time. Interestingly, this finding is in line with recent observations on complex subtraction problems. Torbeyns, Ghesquière, and Verschaffel (2009) manipulated the numerical features of subtraction problems and reported that when the subtrahend was large (e.g., 71–59), participants used a strategy of indirect addition ($59 + 10 = 69, 69 + 2 = 71$), so the response is $10 + 2 = 12$ more often than when it was small (e.g., 71–29). Both observations suggest that calculators evoke the magnitude of the operands, compare them, and use the result of this comparison to organise the computation.

By contrast, in our study, no evidence indicative of an influence of magnitude properties was found for the Pen & Paper and Tens & Units strategies, and one might think that they are more dependent on the formal structure of the Arabic notation than on the magnitude characteristics of the addends. Interestingly, the correlations with the WAIS subtest revealed that mathematically skilled participants tend to use Pen & Paper less and either magnitude-based strategy more than less skilled ones. While the correlations could be mediated by cultural, educational, as well as instructional differences (see, e.g., Imbo & LeFevre, 2009), they nevertheless support the distinction between magnitude-based and non-magnitude-based calculation strategies.

Why would one want to evaluate and compare operands before doing the computation? Is adding 25 to 48 easier than adding 48 to 25? Every colleague with whom we discussed the present findings shared the intuition that it is. One possible explanation is that starting from the largest addend constitutes a remnant of the Min strategy (Groen & Parkman, 1972). Further, if one assumes, as for the elementary facts (Butterworth et al., 2001), that sums of tens (e.g., 40 + 20) are stored in long-term memory following a MAX + MIN organisation, the anchoring of Aggregation and Rounding on the largest addend would suit memory retrieval constraints.

An alternative hypothesis is that starting from the largest addend enables the excitation of a region of the analogical number representation system that is closer to the region of the sum (i.e., 73 is closer to 48 than to 25; for a similar hypothesis, see Restle, 1970). This might facilitate the activation of the numerosity detectors corresponding to the response. The envisaged process is similar to one mechanism operating in number priming experiments (Kochcin, Naccache, Block, & Dehaene, 1999; Reynvoet, Brysbaert, & Fias, 2002), in which a prime facilitates the naming or numerical categorization of a target, and more so when the prime and the target are numerically close to each other. Further experiments would be required to determine whether starting from the largest operand is indeed more efficient than starting from the smallest, especially when the former is close to the sum.

In conclusion, we believe that the present findings have implications for the understanding of both arithmetic processing and individual differences in numerical cognition. Regarding processing, they corroborate the notion of a tight connection between magnitude representations and exact calculation mechanisms. Even though it would seem premature to propose a processing model, the data offer several constraints that should be taken into account. Left-anchored strategies were used more frequently overall than right-anchored strategies, suggestive of a natural left-to-right preference. The default left-to-right procedure can however be modulated by the evaluation of operands and units values, possibly leading to a reorganization of the calculation. Indeed the

**Table 3**

<table>
<thead>
<tr>
<th>Predictors</th>
<th>Left-anchored Aggregation $R^2$</th>
<th>Right-anchored Aggregation $R^2$</th>
<th>Left-anchored Rounding $R^2$</th>
<th>Right-anchored Rounding $R^2$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Addends difference $\beta$</td>
<td>$.70^{**}$</td>
<td>$.74^{**}$</td>
<td>$.40^{**}$</td>
<td>$.78^{**}$</td>
</tr>
<tr>
<td>Units difference $\beta$</td>
<td>$.41^{**}$</td>
<td>$.37^{**}$</td>
<td>$.40^{**}$</td>
<td>$.44^{**}$</td>
</tr>
<tr>
<td>Addends sum $\beta$</td>
<td>$-.09$</td>
<td>$-.01$</td>
<td>$.004$</td>
<td>$.06$</td>
</tr>
</tbody>
</table>

**$p < .005.$**
longer latencies for the first intermediate result when the largest operand appeared on the right suggest such a reorganization (see also Trbovich & LeFevre, 2003, for a similar finding). Exactly how the comparison of operands and of units influence processing requires further research, but the observation that both come into play is reminiscent of the current debate on the holistic or componential nature of mental magnitude (e.g., Nuerk et al., 2001). In summary, the present findings invite to consider more systematically the influence of magnitude characteristics in behavioural as well as in neuroimaging investigations of calculation processes.

Regarding individual differences, it is worth noticing that not all participants used either Aggregation or Rounding. The trend for mathematically skilled participants to use magnitude-based strategies more often than less skilled ones is compatible with Gallistel and Gelman’s assumption that “the acquisition and performance of verbal arithmetic is mediated by the preverbal system for represented numerosity and doing arithmetic computation” (Gallistel & Gelman, 1992, p. 67). One may further wonder whether the strategic preferences are related to individual differences in the number sense, given Halberda, Mazzocco, and Feigenson’s (2008) recent data showing a link between the acuity of the approximate number system and mathematical ability. Studies bearing on the role of magnitude in calculation strategies such as the present one contribute to clarify ways in which the approximate number system could impact on learning and performing arithmetic. For instance, children prone to access and use magnitude information would select adaptive incrementation strategies that might enhance the structure of their arithmetic facts network and the strength of the stored associations. This view opens further perspectives for the analysis of the causal mechanisms through which a core number system deficit might contribute to clarify ways in which the approximate number system and mathematical ability. Studies bearing on the exponential nature of mental magnitude (e.g., Nuerk et al., 2001). In summary, the present findings invite to consider more systematically the influence of magnitude characteristics in behavioural as well as in neuroimaging investigations of calculation processes.

Résumé

Des jeunes adultes devaient résoudre des problèmes d’addition à deux chiffres et décrire à haute voix toutes les étapes de leurs calculs afin d’en identifier les stratégies. Nous avons manipulé la position de l’opérande le plus grand (par ex., 25_48 vs 48_25) afin de tester si les stratégies sont modulées par des caractéristiques de magnitude. Avec certaines stratégies, les participants ont démontré une nette préférence pour prendre le plus grand opérande comme point de départ du calcul. Ainsi, plutôt que d’appliquer les stratégies de façon inflexible, les participants ont évalué et comparé les opérandes avant de procéder au calcul. De plus, les participants habiles en mathématiques ont montré une plus grande tendance à utiliser ces stratégies basées sur la magnitude. Les résultats démontrent que l’information de magnitude joue un rôle dans l’arithmétique complexe en guidant le processus de sélection de stratégie, et ce, possiblement de façon plus marquée pour les participants habiles en mathématiques.

Mots-clés : système de nombres approximatifs, habileté mathématique, arithmétique mentale, sens du nombre, stratégie

References


Received December 29, 2009
Accepted May 17, 2010