ELSEVIER

COGNITION

Cognition 80 (2001) B21–B30

www.elsevier.com/locate/cognit

Brief article

Strategy choice for arithmetic verification: effects of numerical surface form

Jamie I.D. Campbell*, Jonathan Fugelsang

Department of Psychology, University of Saskatchewan, 9 Campus Drive, Saskatoon, S7N 5A5, Canada

Received 22 November 2000; accepted 10 January 2001

Abstract

Canadian university students ($n = 48$) solved simple addition problems in a true/false verification task with equations in digit format ($3 + 4 = 8$) or written English format (three + four = eight). Participants reported their solution strategy (e.g. retrieval or calculation) after each trial. Reported use of calculation strategies was much greater with word (41%) than digit stimuli (26%), and this difference was exaggerated for numerically larger problems. Word-format costs on reaction time (RT) were correspondingly greater for large than for small problems, but this Format × Size RT effect was bigger for true than for false equations. The results demonstrate that surface format affects central, rather than only peripheral, stages of cognitive arithmetic. © 2001 Elsevier Science B.V. All rights reserved.

Keywords: Calculation; Numerical surface form; Strategy choice

1. Introduction

How numerical format (e.g. Arabic digits, written number words, roman numerals, etc.) affects arithmetic is a central issue of numerical cognition research. Some researchers have argued that effects of format are localized to systems that encode numerical stimuli, and do not penetrate downstream to affect calculation (Dehaene & Cohen, 1995; McCloskey, 1992; Noël, Fias, & Brysbaert, 1997). Others have reported evidence that format can directly affect calculation (Blankenberger &

* Corresponding author. Tel.: +1-306-966-6664; fax: +1-306-966-6630.
E-mail address: jamie.campbell@usask.ca (J.I.D. Campbell).

PII: S0010-0277(01)00115-9

Vorberg, 1997; Campbell, 1994; Campbell, Kanz, & Xue, 1999; McNeil & Warrington, 1994; Sciama, Semenza, & Butterworth, 1999).

One previously unexplored way that format could affect calculation is by influencing strategy. Direct memory retrieval is the predominant strategy used by skilled individuals for simple arithmetic (e.g. $6 + 3 = 9$, $7 \times 9 = 63$), but even educated adults report using calculation procedures such as counting or transformation some of the time (e.g. $6 + 7 = 6 + 6 + 1$) (Geary & Wiley, 1991; LeFevre, Sadesky, & Bisanz, 1996). It is plausible that format could affect strategy choice. According to Siegler's *Adaptive Strategy Choice Model* (Siegler & Shipley, 1995), an arithmetic strategy is selected based on its relative efficiency (i.e. speed and probability of success). Consequently, manipulations that reduce the efficiency of retrieval promote a switch to calculation (cf. Campbell & Timm, in press). Problem familiarity is another factor that affects strategy choice for arithmetic. Schunn, Reder, Nhouyvanisvong, Richards, and Stroffolino (1997) showed that participants first judge the strength of familiarity produced by a problem in order to decide whether to attempt retrieval or calculate.

These considerations suggest that the presentation format of arithmetic problems (i.e. Arabic digits vs. written number words) could directly affect strategy choice: because simple arithmetic problems are rarely encountered in written word format, the visual familiarity of word problems is low relative to the familiarity of digit format problems. The results of Schunn et al. (1997) suggest that the low familiarity of word problems could result in less retrieval and more use of calculation. Moreover, relative to digit format, word format greatly increases problem difficulty; reaction times (RTs) and errors both are increased by as much as 30% (Campbell, 1994). According to the Siegler and Shipley (1995) model, the greater difficulty encountered with word stimuli could discourage retrieval and promote use of calculation.

To pursue this hypothesis, we tested university students on simple addition problems using a verification task ($3 + 4 = 8$, true or false?). Verification equations may be solved by a familiarity-based recognition strategy (Zbrodoff & Logan, 1990), by evaluating rules (Lemaire & Reder, 1999), or by generating the correct answer (using direct retrieval or calculation) and comparing it to the presented answer (Ashcraft, Fierman, & Bartolotta, 1984). After each trial, participants selected the strategy used from a list of candidate strategies (cf. Campbell & Timm, in press; Campbell & Xue, in press). As it is rare to encounter addition equations in word format (e.g. two + five = eight), we anticipated that the familiarity-based recognition strategy would be relatively ineffective for word problems, producing a shift to calculation for word equations.

2. Method

2.1. Participants

Thirty-three female and 15 male volunteers (mean age 22.1 years) participated to fulfill a requirement of their introductory psychology course at the University of Saskatchewan.

2.2. Design and stimuli

Participants received four blocks of 72 computer-displayed addition verification trials. On odd-numbered trials, the operands and answer appeared in Arabic digit format ($3 + 4 = 7$). On even trials, the equations appeared in lower-case English (three + four = seven). The problems involved pairs of addends between 2 and 9 (i.e. $2 + 2 = 4$ to $9 + 9 = 18$). There are 36 possible pairings of the numbers 2 through 9 when commuted pairs (e.g. $4 + 5$ and $5 + 4$) are counted as one problem. In each block, all 36 problems were tested in random order, once in digit and once in word format, separated by at least 18 trials. Within each set of 36 digit and 36 word trials in a block, there were 18 true and 18 false equations. Across the four blocks, each of the 36 problems was tested in each format twice in a true equation and twice with a different false answer. For each participant, four false-answer sets were generated pseudo-randomly. Within each set, each of the numbers from 4 to 18 (i.e. the range of true answers) occurred at least once and no more than four times. False answers were within ± 4 of the correct answer and never corresponded to either the difference or the product of the operands.

Equations appeared as white characters against a dark background. Each character space was approximately 3 mm wide and 5 mm high. For word problems, the two operands were separated by the operation sign (+) with a space on each side (e.g. three + eight =). For digit problems the two operands were separated from the + sign by three character spaces (e.g. 3   +   8 =). The answer-to-be-verified appeared simultaneously with the problem operands, centered 10 mm below the + sign.

2.3. Procedure

Even-numbered participants indicated "true" with the right button of the response box and odd-numbered participants responded "true" with the left button. General instructions explained the verification task and requested the participant to respond quickly but accurately. The following strategy instructions were given: "After each equation please indicate how you solved the problem by choosing from among the following possible strategies…RECOGNITION = you thought the equation was true because it seemed familiar or looked right, or false because it seemed unfamiliar or looked wrong. REMEMBER & COMPARE = you remembered the correct answer and then compared it to the presented answer. CALCULATE & COMPARE = you calculated to get the correct answer and then compared it to the presented answer. ODD/EVEN RULES = you used odd/even rules to deduce that the equation was false. OTHER = you used some other calculation strategy (e.g. subtraction) or are uncertain." It was explained that recognition or remembering an answer involved direct retrieval without any intermediate steps, inferences, or calculations. In contrast, calculating an answer involved strategies such as counting or deriving the answer based on knowledge of a related problem.

Prior to the experimental trials, participants received 12 practice trials in alternating digit and word format involving the operand 0 or 1 paired with 0 to 9. For each practice and experimental trial, a fixation dot appeared at the center of the screen. When the participant pressed a button, the fixation dot flashed for 1 s and then was replaced by an equation with the + at fixation. Timing (accurate to ± 1 ms) began when the equation appeared and was stopped by the button-press response. After the response, a green "C" for correct or a red "E" for error appeared for 300 ms at the center of the screen. The prompt "Strategy Choices" then appeared with the strategy labels Recognition, Remember & Compare, Calculate & Compare, Odd/ Even Rules, and Other aligned below. The experimenter recorded the strategy choice by pressing a button on the computer keyboard. The screen then cleared and displayed the fixation dot for the next trial.

3. Results

"Tie" problems (e.g. $2 + 2 = 4$, $3 + 3 = 6$, etc.) were excluded from analysis because of their unique encoding characteristics (cf. Noël et al., 1997). In general, the difficulty of simple arithmetic problems increases with the numerical size of the operands, and use of calculation strategies is more common for the larger, more difficult problems (Geary & Wiley, 1991; LeFevre et al., 1996). To operationalize problem size, "small" problems had addends that produced a product less than or equal to 25, else a problem was "large" (Campbell et al., 1999; Campbell & Xue, in press). This creates small and large problem sets that contain the same number of items (four ties and 14 non-ties).[1]

3.1. Performance

Table 1 presents mean correct RT and percentage of errors as a function of format (digit, word), truth (true, false), and size (small problems, large problems). ANOVA confirmed the standard effects of problem size and format. RTs were longer for large problems (1749 ms) than small problems (1424 ms) ($F(1,47) = 170.61$, $MSe = 59338.89$), and longer for words (1860 ms) than digits (1314 ms) ($F(1,47) = 259.91$, $MSe = 109993.95$). As is often found (Ashcraft & Stazyk, 1981; Campbell & Tarling, 1996), responses to true equations were faster (1492 ms) than responses to false equations (1682 ms) ($F(1,47) = 109.03$, $MSe = 31971.06$). A Truth × Size effect occurred because the problem-size effect (i.e. the mean difference between small and large problems) was larger for true (+368 ms) than for false equations (+281 ms) ($F(1,47) = 19.95$, $MSe = 9856.84$).

There was a strong Format × Size interaction: word-format costs were greater overall for large (+625 ms) than for small problems (+466 ms) ($F(1,47) = 39.66$, $MSe = 15259.11$). This Format × Size effect has been observed previously in simple arithmetic tasks requiring production of the answer (Campbell, 1994; Noël et al., 1997). The Format × Size × Truth interaction was also significant,

[1] The same pattern of results was obtained defining small problems as "both operands less than or equal to five" or "sum less than or equal to 10".

Table 1
Mean RT, and percentage of errors as a function of truth, format, and problem size[a]

Format	True			False		
	Small	Large	L − S	Small	Large	L − S
RT						
Digits	1069	1331	+ 262	1314	1542	+ 228
Words	1546	2020	+ 474	1769	2103	+ 334
W − D	+ 477	+ 689	+ 212	+ 455	+ 561	+ 106
% Errors						
Digits	2.8	6.3	+ 3.5	4.5	8.4	+ 3.9
Words	6.3	13.2	+ 6.9	3.1	7.1	+ 4.0
W − D	+ 3.5	+ 6.9	+ 3.4	−1.4	−1.3	+ 0.1

[a] True, true equations; false, false equations; digits, digit presentation format; words, word presentation format; small, small problems; large, large problems.

however ($F(1,47) = 6.87$, $MSe = 9856.84$, $P = 0.012$). For true equations, there was a strong Format × Size effect, with greater word-format costs for large true equations (+689 ms) than small true equations (+477 ms); the difference between these word-format costs (+212 ms) represents the magnitude of the Format × Size effect for true equations. For false equations, the word-format cost also was greater for large (+561 ms) than for small equations (+455 ms), but the difference (+106 ms) was about half as large compared to true equations.

The corresponding analysis of errors indicated more errors for large (8.8%) than small problems (4.2%) ($F(1,47) = 30.79$, $MSe = 66.34$), and more errors with word (7.4%) than digit stimuli (5.5%) ($F(1,47) = 6.91$, $MSe = 51.94$, $P = 0.012$). Unlike the RT analysis, however, the Format × Size ($F(1,47) = 2.89$, $MSe = 24.48$, $P = 0.10$) and the Format × Size × Truth ($F(1,47) = 2.97$, $MSe = 21.55$, $P = 0.09$) effects only approached significance. Nonetheless, the pattern was similar to the RT analysis. For true equations, the word-format cost in errors tended to be greater for large (+6.9%) than small problems (+3.6%). In contrast, there was practically no effect of format on errors for large false (−1.3%) or for small false equations (−1.4%).

3.1.1. Discussion of performance analyses

The results replicated previous research examining effects of surface form on simple arithmetic production (e.g. Campbell, 1994; Noël et al., 1997), and show that these effects are also observed using a verification procedure. Specifically, performance was substantially worse with word than digit stimuli, but this word-format cost was greater for the larger, more difficult problems. It has been a point of controversy whether the Format × Size interaction arises in arithmetic processes (Campbell, 1994, 1999) or during encoding of problem operands (e.g. McCloskey, Macaruso, & Whetstone, 1992; Noël et al., 1997). Here, however, the Format × Size effect was substantially larger for true than for false equations. Given that encoding

Table 2
Percentage reported use of strategies as a function of truth, format, and problem size[a]

Format	True			False		
	Small	Large	L − S	Small	Large	L − S
Calculate & Compare						
Digits	20.8	26.1	+ 5.3	24.3	29.3	+ 5.0
Words	34.8	46.2	+ 11.4	36.7	47.0	+ 10.3
W − D	+ 14.0	+ 20.1	+ 6.1	+ 12.4	+ 17.7	+ 5.3
Recognition						
Digits	50.2	42.2	− 8.0	48.8	43.9	− 4.9
Words	35.7	27.0	− 8.7	36.2	30.8	− 5.4
W − D	− 14.5	− 15.2	− 0.7	− 12.6	− 13.1	− 0.5
Remember & Compare						
Digits	28.6	29.8	+ 1.2	21.6	21.5	− 0.1
Words	28.2	23.6	− 4.6	21.4	16.4	− 5.0
W − D	− 0.4	− 6.2	− 5.8	− 0.2	− 5.1	− 4.9

[a] True, true equations; false, false equations; digits, digit presentation format; words, word presentation format; small, small problems; large, large problems.

conditions were practically identical for true and false equations this effect cannot be attributed to encoding. Before considering interpretation of the triple interaction, however, it is necessary first to examine the strategy reports.[2]

3.2. Strategy reports

The most common strategy reported was Recognition (39.5% of all trials), followed by Calculate & Compare (33.1%), Remember & Compare (23.9%), Odd/Even Rules (2.2%) and Other strategies (1.4%). As selection of the Odd/Even and Other categories occurred with low frequency, we focused on the other three strategy categories. Table 2 presents the mean percentage use of Calculate & Compare, Recognition, and Remember & Compare as a function of format, truth, and size. We analyzed percentage use of Calculate & Compare, which mirrors percentage use of the two retrieval strategies. Calculation was reported more for large (37.2%) than small problems (29.2%) ($F(1, 47) = 17.41$, $MSe = 353.23$). This replicates previous research examining strategy usage for simple addition using the arithmetic production task (e.g. Geary & Wiley, 1991; LeFevre et al., 1996). Critically, format had a large effect on calculation use: participants were much more likely to report calculation with word stimuli (41.2% of trials on average) than digit

[2] A general concern about strategy reports is whether collection of strategy information affects the way people perform arithmetic (Kirk & Ashcraft, in press). We tested another group of 48 participants using the identical arithmetic verification procedure, but did not collect strategy reports. The patterns of means and of significant effects in the Format × Truth × Size analyses of RT and errors were the same as in the current study. The authors thank Gabe Buettner for his contributions to the no-report control experiment.

stimuli (25.1%) ($F(1,47) = 71.26$, $MSe = 347.94$). Furthermore, format and size interacted ($F(1,47) = 11.39$, $MSe = 69.09$); specifically, the increase in calculation use with words was greater for large problems (+18.9%) than for small problems (+13.2%). There was no suggestion of a triple Format × Size × Truth interaction on calculation use ($F(1,47) < 1$, $MSe = 42.86$).

Table 2 shows that the main effect of format on calculation use reflected a decrease primarily in use of Recognition with the word format. Thus, as we anticipated, use of Recognition memory was reduced by the unfamiliar word format. In contrast, the Format × Size interaction on calculation use reflected a trade-off with use of the Memory & Compare strategy: there was a strong Format × Size interaction in usage of Memory & Compare ($F(1,47) = 10.26$, $MSe = 66.96$, $P = 0.002$) but not in use of Recognition ($F(1,47) < 1$, $MSe = 64.48$). This suggests that the Format × Size interaction on calculation usage occurred specifically because participants were less likely to retrieve the correct answer to large problems given word stimuli.

3.2.1. Strategy performance

Few participants used all three strategies in all eight cells of the design. Consequently, to examine performance differences across strategies we collapsed the data over equation type (i.e. true and false) and analyzed errors ($n = 32$) and RT ($n = 25$) in Format × Size × Strategy repeated measures ANOVAs. In the error analysis, there were no significant effects involving the strategy factor ($F(2,62) < 1$, $MSe = 303.03$ for the main effect of strategy). The RT analysis demonstrated that Calculate & Compare RTs were slower on average (2007 ms) than Recognition (1528 ms) or Memory & Compare RTs (1537 ms) ($F(2,48) = 62.21$, $MSe = 120853.22$). Moreover, the problem-size effect on RT was larger for Calculate & Compare trials (+346 ms) than for Recognition (+252 ms) or Memory & Compare trials (+214 ms) ($F(2,48) = 3.15$, $MSe = 35700.87$, $P = 0.05$). This replicates previous research showing that calculation strategies for simple addition are slow relative to retrieval, but this difference is greater for large problems (LeFevre et al., 1996).

4. Discussion

The results demonstrate that cognitive arithmetic can be strongly affected by format, and are incompatible with the assumption that surface form affects only problem encoding processes (Dehaene & Cohen, 1995; McCloskey, 1992). The unfamiliar word format greatly increased use of calculation strategies. This is consistent with the proposal of Schunn et al. (1997) that familiarity of arithmetic stimuli influences strategy use; specifically, the less familiar a problem is, the less likely one is to attempt direct retrieval and the more likely one is to use a calculation strategy. Furthermore, increased use of calculation with word-format equations was greater for larger, more difficult problems. This suggests that the word format was particularly disruptive of retrieval processes for large problems, thereby increasing the relative utility of calculation strategies, especially for the large word problems

(Siegler & Shipley, 1995). Both neuropsychological and brain imaging studies support the distinction between a direct-retrieval route for simple arithmetic and an indirect, semantically-mediated route that supports calculation strategies (Chochon, Cohen, van de Moortele, & Dehaene, 1999; Dehaene & Cohen, 1997; Hittmair-Delazer, Semenza, & Denes, 1994). The present results suggest that surface form can influence which number processing routes or modules are recruited for arithmetic.

A second important implication of our results is that a Format × Size interaction in arithmetic performance arises, in part, because format can affect the probability that people retrieve as opposed to calculate answers. Specifically, relative to the digit format, the word format promoted more use of relatively slow calculation strategies, but especially so for large problems. As calculation performance was about 500 ms slower on average compared to retrieval, the disproportionate use of calculation for large, word-format problems would contribute to the observed Format × Size interaction on RTs.

This cannot be the whole story of the Format × Size RT interaction, however. The Format × Size effect (i.e. greater word-format RT costs for large than for small problems) was almost twice as big for true as for false equations. Error rates presented a similar pattern. As encoding and response requirements were the same for true and false equations, the triple interaction cannot be attributed to effects arising at these stages. Instead, surface format apparently affected the efficiency of retrieval or calculation processes differently for true and false equations. Why would this occur?

In general, performance on true verification trials is a more sensitive measure of the efficiency of arithmetic processes than is performance on false trials. Indeed, the problem-size effect was substantially larger for true (+368 ms) than for false equations (+281 ms). This probably occurs because arithmetic processing is less likely to need to run to completion for false equations. For example, relatively large magnitude discrepancies (e.g. more than ±2 from correct) may be detected while retrieval or calculation is in progress (Ashcraft & Stazyk, 1981).[3] In contrast, the difference in difficulty between small and large problems is more fully expressed on true trials. Consequently, the problem-size effect on true trials is more sensitive to manipulations that affect the efficiency of retrieval or calculation. Thus, the fact that the Format × Size effect was larger for true than for false equations suggests that the word format reduced the efficiency of arithmetic processes. Greater experience performing arithmetic operations on digits than on words may give digits a stronger capacity to directly activate number-fact representations (Campbell, 1994; Campbell et al., 1999).

[3] Consistent with this, reported usage of the Remember & Compare strategy (Table 2) was lower for false (20.2%) than for true (27.6%) equations ($F(1,47) = 5.92$, $MSe = 872.42$, $P = 0.019$). This would be expected if answer retrieval was less likely to run to completion for false than for true equations.

Acknowledgements

This work was funded by the Natural Sciences and Engineering Research Council of Canada.

References

Ashcraft, M. H., Fierman, B. A., & Bartolotta, R. (1984). The production and verification tasks in mental addition: an empirical comparison. *Developmental Review*, *4*, 157–170.

Ashcraft, M. H., & Stazyk, E. H. (1981). Mental addition: a test of three verification models. *Memory and Cognition*, *9*, 185–196.

Blankenberger, S., & Vorberg, D. (1997). The single-format assumption in arithmetic fact retrieval. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *23*, 721–738.

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

Campbell, J. I. D. (1999). The surface form × problem-size interaction in cognitive arithmetic: evidence against an encoding locus. *Cognition*, *70*, 25–33.

Campbell, J. I. D., Kanz, C. L., & Xue, Q. (1999). Number processing in Chinese-English bilinguals. *Mathematical Cognition*, *5*, 1–39.

Campbell, J. I. D., & Tarling, D. (1996). Retrieval processes in arithmetic production and verification. *Memory and Cognition*, *24*, 156–172.

Campbell, J. I. D., & Timm, J. (in press). Adults' strategy choices for simple addition: effects of retrieval interference. *Psychonomic Bulletin and Review*.

Campbell, J. I. D., & Xue, Q. (in press). Cognitive arithmetic across cultures. *Journal of Experimental Psychology: General*.

Chochon, F., Cohen, L., van de Moortele, P. F., & Dehaene, S. (1999). Differential contributions of the left and right parietal lobules to number processing. *Journal of Cognitive Neuroscience*, *11*, 617–630.

Dehaene, S., & Cohen, L. (1995). Toward an anatomical and functional model of number processing. *Mathematical Cognition*, *1*, 83–120.

Dehaene, S., & Cohen, L. (1997). Cerebral pathways for calculation: double dissociation between rote verbal and quantitative knowledge of arithmetic. *Cortex*, *33*, 219–250.

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

Hittmair-Delazer, M., Semenza, C., & Denes, G. (1994). Concepts and facts in calculation. *Brain*, *117*, 715–728.

Kirk, E. P., & Ashcraft, M. H. (in press). Do adults over-report strategy use in single digit arithmetic? An investigation of instructional bias in verbal reports. *Journal of Experimental Psychology: Learning, Memory, and Cognition*.

LeFevre, J., 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, and Cognition*, *22*, 216–230.

Lemaire, P., & Reder, L. (1999). What affects strategy selection in arithmetic? The example of parity and five effects on product verification. *Memory and Cognition*, *27*, 364–382.

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

McCloskey, M., Macaruso, P., & Whetstone, T. (1992). The functional architecture of numerical processing mechanisms: defending the modular model. In J. I. D. Campbell (Ed.), *The nature and origins of mathematical skills* (pp. 493–537). Amsterdam: Elsevier.

McNeil, J., & Warrington, E. K. (1994). A dissociation between addition and subtraction with written calculation. *Neuropsychologia*, *32*, 717–728.

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

Schunn, C. D., Reder, L. M., Nhouyvanisvong, A., Richards, D. R., & Stroffolino, P. J. (1997). To calculate or not to calculate: a source activation confusion model of problem familiarity's role in strategy selection. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *23*, 3–29.

Sciama, S. C., Semenza, C., & Butterworth, B. (1999). Repetition priming in simple addition depends on surface form and typicality. *Memory and Cognition*, *27*, 116–127.

Siegler, R. S., & Shipley, C. (1995). Variation, selection, and cognitive change. In G. Halford, & T. Simon (Eds.), *Developing cognitive competence: new approaches to process modelling* (pp. 31–76). Hillsdale, NJ: Erlbaum.

Zbrodoff, N. J., & Logan, G. D. (1990). On the relation between production and verification tasks in the psychology of simple arithmetic. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *16*, 83–97.