Mental Arithmetic Processes: Testing the Independence of Encoding and Calculation

Adam R. Frampton and Thomas J. Faulkenberry

Tarleton State University

Author Note

Adam R. Frampton and Thomas J. Faulkenberry, Department of Psychological Sciences, Tarleton State University.

This research was supported in part by funding from the Tarleton State University Office of Student Research and Creative Activities.

Correspondence concerning this article should be addressed to Thomas J. Faulkenberry, Department of Psychological Sciences, Tarleton State University, Box T-0820, Stephenville, TX 76402. Phone: +1 (254) 968-9816. Email: faulkenberry@tarleton.edu

Abstract

Previous work has shown that the cognitive processes involved in mental arithmetic can be decomposed into three stages: encoding, calculation, and production. Models of mental arithmetic hypothesize varying degrees of independence between these processes of encoding and calculation. In the present study, we tested whether encoding and calculation are independent by having participants complete an addition verification task. We manipulated problem size (small, large) as well as problem format, having participants verify equations presented either as Arabic digits (e.g., "3 + 7 = 10") or using words (e.g., "three + seven = ten"). In addition, we collected trial-by-trial strategy reports. Though we found main effects of both problem size and format on response times, we found no interaction between the two factors, supporting the hypothesis that encoding and calculation function independently. However, strategy reports indicated that manipulating format caused a shift from retrieval based strategies to procedural strategies, particularly on large problems. We discuss these results in light of two competing models of mental arithmetic.

Keywords: mental arithmetic, encoding, calculation, strategy reports

Mental Arithmetic Processes: Testing the Independence of Encoding and Calculation

Mental arithmetic is a daily skill that involves several cognitive processes. For example, consider the period of time at the end of a meal when it is time to calculate a tip. Calculating a gratuity (e.g., 20%) requires first looking at the lunch total and encoding the total into a mental representation. The gratuity is then calculated through some manipulation of previously learned mathematical facts (possibly including practiced procedures or direct memory retrieval), culminating in the production of an amount that is then written on the receipt. Though most researchers agree on framing mental arithmetic in terms of encoding, calculation, and retrieval (Campbell & Clark, 1988; Dehaene, 1992; McCloskey, 1992), there is considerable debate regarding the independence of these processes. The purpose of the present study was to investigate the interaction between encoding and calculation in mental arithmetic.

To study encoding and calculation, we relied upon two specific empirical effects found throughout the literature. First, a common signature of the calculation process is the problem size effect, which is the finding that response times and errors increase as problem operands grow in magnitude (Ashcraft, 1992; Campbell, Parker, & Doetzel, 2004; Groen & Parkman, 1972). For example, the simple addition problem 1+1 is typically solved more quickly and accurately than 8+9. One early explanation for the problem size effect was Groen and Parkman's (1972) "min" (minimum addend) model, in which the larger operand of a simple arithmetic problem (9) is fixed in the mind and the smaller operand (8) is added in increments of one until the total (17) is reached. Thus, the increase in RT from the problem 1+1 is attributed to an increase the number of increments calculated by the thinker. Such "min" counting is a prevalent operation employed by children, but adults are known to revert to this incremental technique only when direct retrieval fails (Groen & Parkman, 1972). Later accounts of the problem size effect included the network retrieval model (Ashcraft, 1987) and the network interference model (Campbell, 1987), both of which explained the problem size effect as due to the structure of arithmetic facts in long term memory. Further explanations of the problem size effect have included roles for number fact acquisition and strategy selection. For example, smaller problems are encountered more frequently and may be less sensitive to interference during retrieval from long term memory (Campbell & Alberts, 2009). In addition, larger numbers are more likely to evoke procedural strategies (e.g., incremental counting) and are more error prone due the failing of retrieval activation (Campbell & Alberts, 2009; LeFevre, Sadesky, & Bisanz, 1997).

Though there are several competing explanations for the problem size effect, the increase in response times for larger problems is robust, and hence, the problem size effect is a reliable marker of mental arithmetic. Similarly, the encoding process in mental arithmetic is reflected by the format effect (Campbell, 1994; Dehaene & Cohen, 1995; Noël, Fias, & Brysbaert, 1997), where response times and errors increase when problems are presented in word problem format (Nine + three), as opposed to Arabic digits (9 + 3). The penalty associated with changes in surface format may be attributed to lack of familiarity of word problems (Schunn et al., 1997). However, it is not clear whether these format effects are localized to the encoding process alone, or whether changes in format directly affect the calculation process.

As such, the current research reveals a debate over how the processes of encoding and calculation interact. Models formed on the premise of no direct communication between encoding and calculation are called additive models (e.g., the Triple Code Model of Dehaene and Cohen, 1995), for they assert no interaction between encoding and calculation. That is, arithmetic codes exiting the encoding phase will not affect the manner in which the arithmetic problem is calculated. To be specific, such a model assumes that any differences in encoding (e.g., manipulating surface format, such as “4+5” versus “four + five”) would not directly affect calculation processes (Dehaene & Cohen, 1995).

One example of an additive model is Dehaene and Cohen’s (1995) Triple Code Model. This model accounts for numerical processing through three separate code systems: the auditory verbal word frame, analog magnitude representation, and visual Arabic number form. Each code subsystem is responsible for a different number processing task. The auditory verbal word frame mediates written and spoken input and output, while the visual Arabic number form handles digital input and output as well as multi-digit operations. Finally, the analog magnitude representation is recruited in core number operations of estimation, magnitude comparison, and potentially subitizing (Campbell & Epp, 2005, p. 347). Critically, the Triple Code Model predicts that response time costs due to surface format manipulations can be attributed to the efficiency of transcoding the visual stimuli into their appropriate internal codes. As calculation would take place entirely within the visual Arabic number form, there would be no downstream effect of format on the actual calculation process.

A second class of models is characterized by an interaction between the encoding and calculation processes. That is, the processes involved in calculation directly depend upon the format in which stimuli are encoded (Campbell & Alberts, 2009; Campbell, Parker, & Doetzel, 2004). To illustrate, think back to the lunch bill example presented earlier. If the bill was illustrated in Roman numerals rather than the typical Arabic digit format, one would expect that the time to calculate the tip would increase. What is not clear is whether the source of this increase is due to the efficiency of encoding the Roman numerals into a mental representation consisting of Arabic digits (e.g., the Triple Code Model), or whether the process of encoding the Roman numerals fundamentally affects the processes involved in calculation. Campbell and Epp

(2005) present such an interactive approach with their Encoding Complex model. They argue that problem operands automatically trigger a network of associations and operations related to the problem encoded. Like the Triple Code Model, it views number processing as dependent upon representational codes (e.g., visual, verbal, magnitude). However, successful number processing depends on the strength of the relationships between representational codes, termed skilled processing (Campbell & Epp, 2005). Presentation of problems in an unfamiliar format would result in greater response times and errors because participants are unable to "maximize activation of relevant information and minimize activation of irrelevant information" (Campbell & Epp, 2005, p. 350). That is, the manipulation of format would directly affect the calculation process.

Evidence for an interactive model of mental arithmetic came from Campbell and Fugelsang (2001), who required participants to record their strategy choice after completing simple arithmetic problems. Participants were presented single-digit addition problems in a true/false verification task with equations displayed in either Arabic digit format ("6 + 3 = 8") or word format ("six + three = eight"). Campbell and Fugelsang (2001) found that the problem size effect was larger for word problems than for digit problems. This format x size interaction indicated that the response time costs associated with the word format were carried into the calculation stage, thus supporting an interactive model of mental arithmetic. Further, Campbell and Fugelsang found that the most commonly reported strategy was direct retrieval from long term memory. However, participants reported using procedural strategies more often than retrieval when solving word format problems, especially when the problem operands were large. That is, format had a direct effect on the types of strategies used in calculation, which further supports the interactive model.

The purpose of our study was to perform a replication of Campbell and Fugelsang (2001) with a different population. The participants in Campbell and Fugelsang (2001) were psychology students at a Canadian university. Though Campbell and Fugelsang did not report educational backgrounds of their participants, Campbell (personal communication) has reported that many participants in his lab are educated in China, which may lead to differences in performance on mental arithmetic (Campbell & Xue, 2001). We performed our replication on a sample of participants from Texas. Based on the past literature, we expected to replicate both the effects of problem size (small problems should be solved faster than large problems) and problem format (digit problems should be solved faster than word problems). The critical test concerns whether format (digit, word) interacts with problem size (small, large) on response times. If there is an interaction, the results would lend support for the interactive model of mental arithmetic (Campbell & Epp, 2005). If there is no such interaction, the results instead support an additive model of mental arithmetic (Dehaene & Cohen, 1995), calling the results in Campbell and Fugelsang (2001) into question.

Method

Participants

Twenty-three undergraduate students (18 female, mean age = 25.2 years, age range 19 to 60) participated in this experiment in exchange for partial course credit in their psychology courses. Within this sample, 16 self-identified their ethnicity as White, two self-identified as Black, and five as Hispanic. The experiment was reviewed and approved by the institutional review board at Tarleton State University.

Design and Stimuli

Each participant completed 288 trials consisting of four blocks of 72 single-digit addition verification problems. On even numbered trials, questions were presented as word problems in lower case English (“five + seven = twelve”). On odd numbered trials, questions were presented in Arabic digit format (“5 + 7 = 12”). Problems were composed of addends ranging from 2 to 9, resulting in a set of 36 problems ranging between 2 + 2 = 4 and 9 + 9 = 18 (note that commuted pairs such as 2 + 6 and 6 + 2 were counted as one problem). In each block, each of the 36 problems was presented once in digit format and once in word format. Problem size was defined as either small (product of operands less than or equal to 25) or large (product of operands greater than 25). Within each set of 36 problems, 18 were presented as true equations (e.g., “2 + 4 = 6”) and 18 were presented as false equations (e.g., “2 + 4 = 7”). Across all four blocks, each addition problem was tested in each format twice in a true equation and twice with a different false answer. False answers were generated pseudo-randomly to be within ± 4 of the correct answer and never corresponded to either the difference or the product of the operands. Within each set of false answers, each of the numbers 4 to 18 (i.e. the range of true answers) occurred at least once but no more than four times.

All equations appeared as white characters against a black background, displayed in 36 point Lucida Grande font. For all equations the two operands were separated by a single space on either side of the + sign (e.g. three + eight = eleven). The answer to be verified appeared simultaneously with the problem operands. Following each verification trial, participants were asked to indicate the strategy they used by selecting one of five strategy descriptions (originally presented in Campbell and Fugelsang, 2001): “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 remember 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."

Procedure

Even numbered participants selected "true" responses by pressing the right button of a response box, and odd numbered participants selected "true" responses by pressing the left button. Each participant was instructed to respond quickly but accurately.

Prior to the first block, each participant completed 12 practice trials in alternating word and digit format using the operand 0 or 1 paired with 0 to 9. At the beginning of each trial, a fixation point appeared at the center of the screen. When ready to begin, participants initiated the presentation of the equation with a button press. The fixation dot flashed for 1 second and was then replaced with an equation. The timer began with the presentation of this equation and ended with the participant's manual response (a button press indicating true or false). All response times were accurate to ± 1 ms. After each response, feedback was given; a green C for correct or red E for error flashed on the screen for 300ms. On the subsequent screen, the prompt "Strategy Choices" appeared with the cues Recognition, Remember & Compare, Calculation, Odd/Even Rules, or Other aligned vertically below. The experimenter recorded the strategy choice on each trial with a press of a keyboard button, clearing the screen and prompting the fixation point for the next trial.

Results

Participants completed a total of 6,624 experimental trials. Of these, we removed 423 trials that contained an error response (6.4%). From the remaining correct trials, we removed an additional 118 trials (1.9%) for which response time (RT) exceed 3 standard deviations from the mean RT over all trials (M = 1708 ms, SD = 1056 ms). All RT analyses were performed on the remaining 6,083 trials.

Response time analysis

We submitted correct RTs to a 2 (Problem Size: small, large) x 2 (Format: digit, word) x 2 (Truth Value: true, false) repeated measures analysis of variance. Results can be seen in Figure 1. As expected, there was a significant main effect of Problem Size, F (1, 22) = 72.3, $p < 0.001$, $\eta_p^2 = 0.77$. RTs for large problems (M = 1819 ms) were longer than for small problems (M = 1453 ms). Also, there was significant main effect of Format, F (1, 22) = 327.8, $p < 0.001$, $\eta_p^2 =$ 0.94, with word problems (M = 1880 ms) taking longer to verify than digit problems (M = 1392 ms). Finally, there was a significant main effect for Truth Value, F (1, 22) = 35.1, $p < 0.001$, η_p^2 = 0.61, with false problems (M = 1742 ms) taking longer to verify than true problems (M = 1530 ms). There was a small, but statistically significant interaction between Problem Size and Truth Value, F (1, 22) = 4.6, $p = 0.04$, $\eta_p^2 = 0.17$. As can be seen in Figure 1, the problem size effect, operationalized as the difference between RT for large problems and small problems, was slightly smaller for false problems (mean difference = 333 ms) than for true problems (mean difference = 400 ms). Critically, there was no interaction between Problem Size and Format, F (1, 22) = 0.03, $p = 0.86$, $\eta_p^2 < 0.01$, lending support for an additive model of mental arithmetic (Dehaene & Cohen, 1995) over an interactive model (Campbell & Epp, 2005). No other terms in the ANOVA model were significant (all F-values less than 0.5).

Strategy Reports

Similar to Campbell and Fugelsang (2001), we calculated the mean percentage use of three prevalent strategies, Calculate & Compare, Recognition, and Remember & Compare, which is presented in Table 1. Notice that Recognition was used on most trials, but this strategy shifted to Calculate & Compare for large word problems. Whereas the RT analysis above did not bear out a Problem Size x Format interaction, the strategy reports do seem to indicate that format and problem size interact with regard to the types of strategies used.

Discussion

The purpose of the present study was to perform a replication of Campbell and Fugelsang (2001) and test the independence of encoding and calculation. We did this by having participants perform an arithmetic verification task and give trial-by-trial strategy reports. We tested 23 participants who verified whether addition equations were true or false in multiple conditions derived from manipulating problem format (Arabic digits or word problems) and problem size (small or large).

Given that the effects of problem size and format are strong in the literature (Ashcraft, 1992; Campbell, 1994; Campbell, Parker, & Doetzel, 2004; Dehaene & Cohen, 1995; Groen & Parkman, 1972; Noël, Fias, & Brysbaert, 1997), we expected to find significant main effects of both. This expectation was confirmed. We found a large main effect of problem size; response times increased when problem operands increased in magnitude. We also found a large main effect of surface format; response times for equations in word format were larger than response times for equations in Arabic digit format.

The critical test for independence of encoding and calculation came from testing the interaction between the factors of problem size and format. Recall that an additive model (e.g., Dehaene & Cohen, 1995) is based on the premise of no direct communication between encoding and calculation processes. That is, codes acquired in the encoding phase do not affect the manner in which calculation occurs. For example, Dehaene and Cohen's (1995) Triple Code model would predict that the verbal auditory word frame subsystem would encode stimuli in word format (e.g., “six + nine = fifteen”) and transform them into the appropriate internal code for calculation, which would be the visual Arabic form. Subsequently, this visual Arabic number form would account for calculation independently of the initial form of the stimulus. Thus, the effect of problem size (a calculation effect) would be the same regardless of initial format, which would imply that there is no interaction between problem size and format.

Alternatively, an interactive model (e.g., Campbell & Epp, 2005) would predict a direct influence of encoding on calculation. For example, Campbell and Epp’s (2005) Encoding Complex Model would predict that the presented problems would activate a rich network of associations, including activations of both correct and incorrect answers which are all used in the calculation process. Presenting problems in an unfamiliar format (such as word format) would lessen the activation strength of the correct answer and potentially increase activation strength of incorrect answers, thus exacerbating the effect of problem size on RT. That is, encoding factors would directly affect the calculation process, implying that there would be an interaction between problem size and format.

Critically, we did not find a statistically significant interaction between format and problem size on RT. This finding is in opposition to that of Campbell and Fugelsang (2001). Rather, our results support an additive model of arithmetic processing. However, our participant strategy reports mirror those from Campbell and Fugelsang (2001). Like Campbell and Fugelsang’s (2001) participants, our participants utilized the recognition strategy most often to solve equations, but showed a shift from recognition to procedural strategies when encountering large word problems. This strategy shift suggests that manipulations in the presentation format of equations may alter the strategy used to complete the equations. That is, format manipulations seem to have downstream effects on the calculation process, which is the signature of an interactive model of arithmetic processing.

Thus, the results of our study are mixed. Response time patterns indicate support for an additive model of arithmetic processing, whereas strategy reports support an interactive model. However, some limitations to this study demand that the results be interpreted tentatively. With a relatively small sample size, we may not have had sufficient power to detect the interaction of problem size and format on response times. However, the expected effects of problem size and format were sufficiently robust, and the extremely small F-ratio ($F = 0.03$) on the problem size x format interaction makes reduced power an unlikely culprit. The nature of the verification task should be interpreted in context as well. Unlike a production task in which the answer is provided by the participant, a choice is made about a potential answer provided on each trial. This could affect the RT because the participant could rely on recognizing the answer provided rather than reaching the answer independently. That is, the verification task may not truly reflect calculation processes in the same way as a production task.

Given these contradictory findings, which result should we believe? Strategy reports are known to be questionable indicators of mental processes (Cooney & Ladd, 1992; Russo, Johnson, & Stephens, 1988). Response time patterns have long been the gold standard in cognition research, and as such, they should be interpreted accordingly. The absence of an interaction between problem size and format gives us a fair amount evidence in support of an additive model of mental arithmetic (e.g., Dehaene & Cohen, 1995). Future work could investigate the nature of strategy reports in mental arithmetic using free response setting rather than the forced choice setup we used.

In summary, we found support for an additive model of mental arithmetic performance, as predicted by Dehaene and Cohen (1995). Our results indicate that the stages of encoding and calculation in mental arithmetic are functionally independent. This result helps to further specify the cognitive mechanisms behind mental arithmetic and further clarify how adults go about the daily task of doing mental calculations.

References

Ashcraft, M.H. (1987). Children's knowledge of simple arithmetic: A developmental model and simulation. In J. Bisanz, C.J. Brainerd, & R. Kail (Eds.), *Formal methods in developmental psychology: Progress in cognitive development research* (pp. 302-338). New York: Springer- Verlag. doi:10.1007/978-1-4612-4694-7_9

Ashcraft, M. H. (1992). Cognitive arithmetic: A review of data and theory. *Cognition, 44*(1-2), 75-106. doi:10.1016/0010-0277(92)90051-i

Ashcraft, M., & Stazyk, E. (1981). Mental addition: A test of three verification models. *Memory & Cognition, 9*(2), 185. doi:10.3758/BF03202334

Campbell, J. I. D. (1987). Network interference and mental multiplication. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 13*(1), 109–123. doi:10.1037/0278-7393.13.1.109

Campbell, J. I. D. (1994). Architectures for numerical cognition. *Cognition, 53*(1), 1–44. doi:10.1016/0010-0277(94)90075-2

Campbell, J. D., & Alberts, N. M. (2009). Operation-specific effects of numerical surface form on arithmetic strategy. *Journal of Experimental Psychology: Learning, Memory,and Cognition*, *35*(4), 999-1011. doi:10.1037/a0015829

Campbell, J. I., & Clark, J. M. (1988). An encoding-complex view of cognitive number processing: Comment on McCloskey, Sokol, and Goodman (1986). *Journal of Experimental Psychology: General, 117*(2), 204–214. doi:10.1037/0096-3445.117.2.204

Campbell, J. D., & Epp, L. J. (2005). Architectures for arithmetic. In J. I. D. Campbell (Ed.), *Handbook of mathematical cognition* (pp. 347-360). New York, NY, US: Psychology Press.

Campbell, J. I. D., & Fugelsang, J. (2001). Strategy choice for arithmetic verification: effects of numerical surface form. *Cognition, 80*(3), B21–B30. doi:10.1016/s0010-0277(01)00115-9

Campbell, J. D., Parker, H. R., & Doetzel, N. L. (2004). Interactive effects of numerical surface form and operand parity in cognitive arithimetic. *Journal of Experimental Psychology: Learning, Memory & Cognition, 30*(1), 51-64. doi:10.1037/0278-7393.30.1.51

Campbell, J. I. D., & Xue, Q. (2001). Cognitive arithmetic across cultures. *Journal of Experimental Psychology: General, 130*(2), 299–315. doi:10.1037/0096-3445.130.2.299

Cooney, J. B., & Ladd, S. F. (1992). The influence of verbal protocol methods on children's mental computation. *Learning and Individual Differences*, *4*(3), 237-257. doi:10.1016/1041-6080(92)90004-x

Dehaene, S. (1992). Varieties of numerical abilities. *Cognition, 44*(1-2), 1–42. doi:10.1016/0010-0277(92)90049-n

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

Groen, G. J., & Parkman, J. M. (1972). A chronometric analysis of simple addition. *Psychological Review, 79*(4), 329-343. doi:10.1037/h0032950

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, and Cognition, 22*(1), 216–230. doi:10.1037/0278-7393.22.1.216

McCloskey, M. (1992). Cognitive mechanisms in numerical processing: Evidence from acquired dyscalculia. *Cognition, 44*(1-2), 107–157. doi:10.1016/0010-0277(92)90052-j

Morey, R. D. (2008). Confidence intervals from normalized data: A correction to cousineau

(2005). *Tutorial in Quantitative Methods for Psychology*, *4*(2), 61–64.

Noël, M.-P., Fias, W., & Brysbaert, M. (1997). About the influence of the presentation format on arithmetical-fact retrieval processes. *Cognition, 63*(3), 335–374. doi:10.1016/s0010-0277(97)00009-7

Russo, J. E., Johnson, E. J., & Stephens, D. L. (1989). The validity of verbal protocols. *Memory & Cognition, 17*(6), 759–769. doi:10.3758/bf03202637

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. doi:10.1037/0278-7393.23.1.3

Table 1

Percentage reported use of strategies as a function of truth, format, and problem size

		True problems		False problems	
	Format	Small	Large	Small	Large
Recognition	Digits	54.4	55.5	52.1	54.2
	Words	43.7	27.4	41.1	29.1
Remember and compare	Digits	24.4	21.4	23.0	21.8
	Words	31.3	27.2	23.7	22.9
Calculate and compare	Digits	20.4	22.1	24.2	22.7
	Words	24.2	43.5	34.0	47.5

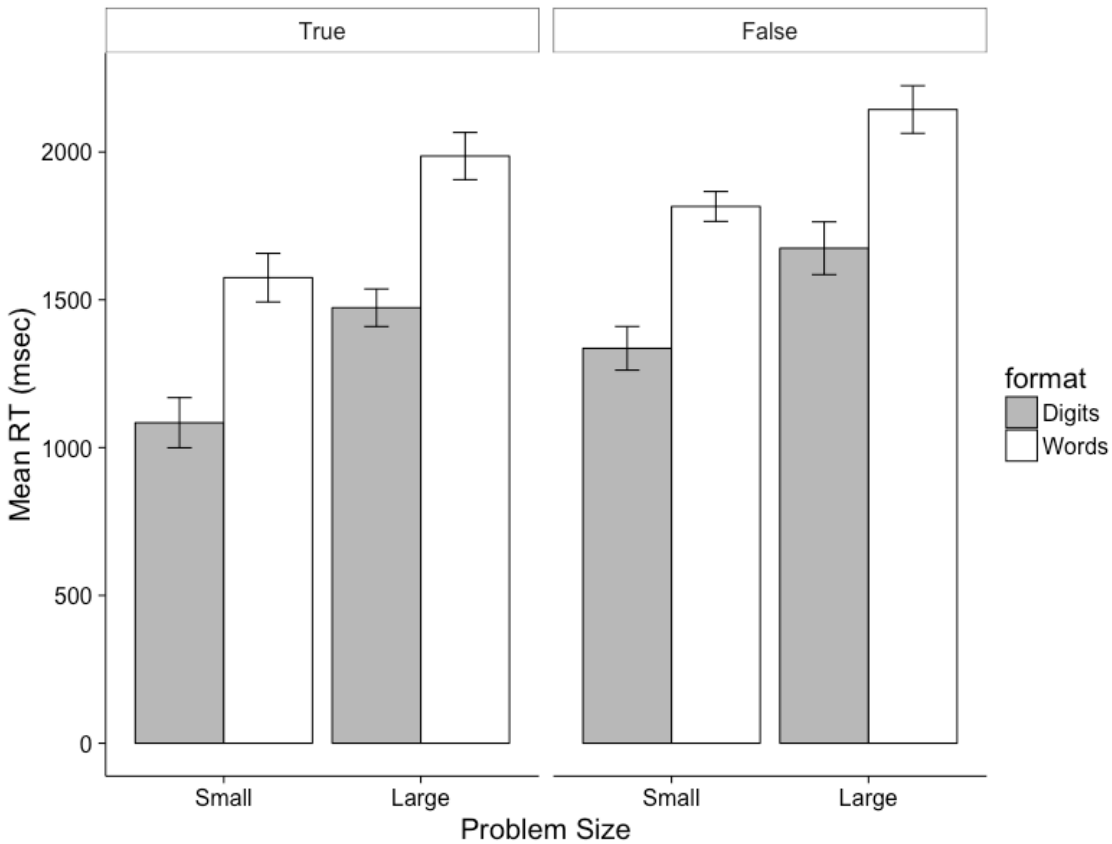

Figure 1. Mean RT as a function of problem size (small, large), format (digits, words), and truth value (true, false). Error bars represent within-subject 95% confidence intervals as recommended by Morey (2008).