

**When does the story matter? No evidence for the foregrounding hypothesis in math story problems**

Sabrina M. Di Lonardo Burr<sup>1</sup>, Jill Turner<sup>1</sup>, Jesse Nietmann<sup>2</sup>, and Jo-Anne LeFevre<sup>1,2</sup>

<sup>1</sup>Carleton University, Department of Cognitive Science

<sup>2</sup>Carleton University, Department of Psychology

Corresponding author: Sabrina Di Lonardo Burr, Department of Cognitive Science, Carleton University, 1125 Colonel By Drive, Dunton Tower, Ottawa, ON, K1S 5B6

email: [sabrina.dilonardo@carleton.ca](mailto:sabrina.dilonardo@carleton.ca)

**ORCID**

Sabrina M. Di Lonardo Burr: <https://orcid.org/0000-0001-6338-9621>

Jo-Anne LeFevre: <https://orcid.org/0000-0002-1927-7734>

### **Abstract**

Math story problems are difficult for many solvers because comprehension of mathematical and linguistic content must occur simultaneously. Across two studies, we attempted to conceptually replicate and extend findings reported by Mattarella-Micke and Beilock (2010) and Jarosz and Jaeger (2019). Mattarella-Micke and Beilock found that multiplication word problems in which an irrelevant number was associated with the protagonist of the problem (i.e., foregrounded in the text) were solved less accurately than problems in other conditions. Jarosz and Jaeger used similar materials but tested the more general inconsistent-operations hypothesis that association with the protagonist would interfere with multiplication whereas dissociation would interfere with division. They found partial support: When division problems were primed with dissociative scenarios, solvers made more errors, but they failed to replicate the associative findings for multiplication. In the present research, we conducted two studies ( $Ns = 205$  and  $359$ ), in which we similarly manipulated whether irrelevant content was associated with or dissociated from the story protagonist. In these studies, we did not find support for either the foregrounding or inconsistent-operations hypotheses. Exploratory error analyses suggested that solvers' errors were most often the result of calculation difficulties or inappropriate operation choices and were unrelated to the presence of associative or dissociative story elements. Our careful implementation of this manipulation and much greater power to detect effects suggests that the association manipulation in irrelevant text does not influence adults' performance on simple math story problems.

Keywords: word problems, math performance, numerical interference, extraneous elements

Abstract word count: 236

When does the story matter? No evidence for the foregrounding hypothesis in math story problems

People often struggle with story problems even when they are capable of doing the underlying mathematics because they have difficulty extracting relevant information from the text to determine the appropriate mathematical equation and procedure (Daroczy, Wolska, Meurers, & Nuerk, 2015; Mayer & Hegarty, 1996; Schley & Fujita, 2014; Thevenot & Barrouillet, 2015). To be successful, solvers must use both mathematical and linguistic knowledge. For example, linguistic knowledge about word order may be critical in extracting the appropriate equation from the text (Fuchs, Gilbert, Fuchs, Seethaler, & Martin, 2018): '*more than*' implies a different operation than does '*then there were more*'. In the present research, we addressed the question of whether subtle differences in the words used in story scenarios influence adults' performance on mathematically simple problems.

Mattarella-Micke and Beilock (2010) reported that word choice in irrelevant story text coupled with numerical interference influenced math performance. More specifically, their *foregrounding* hypothesis reflected predictions based both on situation model research in text comprehension (e.g., Kintsch & van Dijk, 1978) and retrieval processes in arithmetic cognition (e.g., Siegler, 1988). Within a text, the situation model represents the characters, their behaviour, and object-character interactions (Zwaan & Radvansky, 1998). Shifts in the situation model can occur if important story elements are foregrounded (Glenberg, Meyer, & Lindem, 1987). Foregrounding occurs when objects are made more or less salient to participants by either spatially associating or dissociating extraneous information with the protagonist in a narrative (Glenberg et al., 1987). For example, '*He put on his sweatshirt*' is spatially associative because the act of *putting on* the sweatshirt associates the sweatshirt with the protagonist's body. In

contrast, '*He took off his sweatshirt*' is spatially dissociative because the act of *taking off* the sweatshirt dissociates the sweatshirt from the protagonist's body. In texts where the object was spatially associated with the protagonist as opposed to dissociated, Glenberg and colleagues (1987) found that participants were faster to confirm or disconfirm the presence of the object. Based on this finding, Mattarella-Micke and Beilock hypothesized that foregrounding could affect math story problems when irrelevant interfering numbers were foregrounded and thus associated with the protagonist of the story.

Predictions consistent with the distribution of association (DOA) model of simple arithmetic (Siegler, 1988) were used by Mattarella-Micke and Beilock (2010) to refine the foregrounding hypothesis. According to the DOA model, both correct and incorrect answers to math problems are activated and compete for retrieval. Thus, if an incorrect answer is more highly activated than the correct answer, the incorrect response may be retrieved. The incorrect answer may also be retrieved if the solver activates the wrong information in response to a presented stimulus; this phenomenon is known as retrieval interference. For example, when participants are presented with single-digit addition problems (e.g.,  $2 + 3$ ) followed by single-digit multiplication problems (e.g.,  $2 \times 3$ ), they are more likely to add when they should be multiplying (e.g., respond that  $2 \times 3 = 5$ ; Campbell & Timm, 2000). For adults who have years of addition and multiplication practice, the correct response (e.g., 12 to  $3 \times 4$ ) receives more activation than the highly-interfering answer (e.g., 7) under most conditions. However, the activation of these associations is sensitive to semantic context (Bassok, Pedigo, & Oskarsson, 2008). Thus, Mattarella-Micke and Beilock proposed that if an interfering answer became semantically foregrounded (i.e., through association with the protagonist), math answer retrieval would be disrupted because these slight shifts in the story problem text would change the

accessibility of problem content. In essence, Mattarella-Micke and Beilock made the strong claim that minor word changes to text that is irrelevant to constructing an appropriate situation model can nevertheless influence problem solving of skilled adults, especially when coupled with an interfering number.

In each word problem used by Mattarella-Micke and Beilock (2010), a number that represented a quantity of objects was either associated with or dissociated from the protagonist. For example, consider this problem from Mattarella-Micke and Beilock: “*Alexis, an architect, was busy designing a new office building in town. She picked up 16 sketches she had been working on and took the elevator to a meeting where she would deliver her ideas. Alexis took the elevator 7 floors. If each floor is 9 feet tall, how many feet did she travel?*” In this problem, the 16 sketches are associated with the protagonist because they were *picked up*. In the dissociation version, the object would be dissociated with the protagonist, such as “*She set aside 16 sketches...*”. The number connected with the associated or dissociated information was either highly-interfering or less-interfering relative to the required calculation. In this example, the highly-interfering number is 16, which is the sum of 7 and 9 (i.e., the critical numbers for the mathematical problem). In contrast, the less-interfering version of this scenario contained a number that was similar in magnitude to the interfering number (e.g., 15). Mattarella-Micke and Beilock hypothesized that foregrounded information that included a highly-interfering number would disrupt mathematical problem solving. In support of their hypothesis, participants made more multiplication errors when the story included a highly-interfering number (i.e., the sum of the critical numbers) and when the object in the story was associated with the protagonist.

Jarosz and Jaeger (2019) attempted to replicate Mattarella-Micke and Beilock’s (2010) results for multiplication and extend the findings to division. They proposed the *inconsistent-*

*operations hypothesis* which includes the prediction that association will activate addition and adds the further prediction that dissociation will activate subtraction. For example, the dissociative words *put down* are often used in subtraction problems, whereas the associative words *picked up* are often used in addition problems. Thus, in the example above, the addition operation would be activated in the associative condition, potentially interfering when the problem required multiplication, whereas subtraction would be activated in the dissociative condition, potentially interfering when the problem required division. Addition and multiplication interfere with each other in arithmetic studies without story contexts (Butterworth, Zorzi, Girelli, & Jonckheere, 2001; Campbell & Metcalfe, 2007; Campbell & Timm, 2000; LeFevre et al., 1996), and subtraction interferes with division if solvers are using the procedural strategy of repeated subtraction to calculate the quotient (Kouba, 1989; Mulligan & Mitchelmore, 1997). Note that although Mattarella-Micke and Beilock included division problems in their study, they did not analyze them.

The foregrounding hypothesis (Mattarella-Micke & Beilock, 2010) and the inconsistent-operations hypothesis (Jarosz & Jaeger, 2019) make the same prediction for multiplication problems: Solvers will make more errors on multiplication problems for associative than dissociative scenarios. In contrast, the inconsistent-operations hypothesis also predicts that solvers will make more errors on division problems for dissociative than for associative scenarios because dissociative scenarios contain subtraction language. Consistent with this prediction, across all three of their studies, Jarosz and Jaeger found that participants made more division errors when the problem scenario was dissociative rather than associative. The presence or absence of an interfering number did not influence their results. Moreover, they failed to replicate Mattarella-Micke and Beilock's findings in that there were no differences in

performance for associative versus dissociative problems for multiplication. Thus, the foregrounding hypothesis was not strongly supported when all five studies across these two papers are considered. Furthermore, a search among published work that cited Mattarella-Micke and Beilock (2010) did not uncover any other studies in which associative/dissociative language was manipulated in math word problems, despite the substantial literature on interactions between linguistic and numerical factors in word problem solving (Daroczy et al., 2015).

The foregrounding and the interfering-operations hypotheses have implications for models of the interplay between text elements and numerical components of story problems (Daroczy et al., 2015). The lack of supporting literature and the inconsistent patterns across the two papers suggested to us, however, that the effects of these highly-specific manipulations on problem-solving performance might not be strong. We hypothesized that minor wording changes and the presence of interfering numbers in irrelevant text may not be sufficient to disrupt the construction of the situation model for story problems, especially among skilled adults. Accordingly, we attempted to conceptually replicate and extend the findings of Mattarella-Micke and Beilock (2010).

A conceptual replication was necessary for several reasons. First, we were unable to obtain the original stimuli used by Mattarella-Micke and Beilock (2010). Second, Mattarella-Micke and Beilock's design confounded specific stories with experimental conditions (i.e., story content was not counterbalanced across conditions) and other aspects of the problems were not controlled (i.e., problems varied in length). We redesigned the materials to be more uniform and to avoid confounds between specific stories and conditions. We also presented the word problems all at once, rather than in two parts, to maximize the ecological validity of the word problems. Third, sample sizes in Mattarella-Micke and Beilock's studies were small (i.e., 38 and

21 in Experiments 1 and 2, respectively) and effect sizes were small, indicating that larger sample sizes would be prudent for replication.

After we had designed and implemented the current research, we became aware of similar work by Jarosz and Jaeger (2019) in which they also attempted to replicate Mattarella-Micke and Beilock (2010). Thus, not all the features of the studies conducted by Jarosz and Jaeger were included in the present research. Fortunately, we could compare performance across operations and across association conditions for both operations, because Jarosz and Jaeger had also modelled their manipulations on those of Mattarella-Micke and Beilock. We did not, however, have the same numerical interference manipulations as Jarosz and Jaeger. In summary, our analyses constitute a conceptual replication of the research of Mattarella-Micke and Beilock, and a partial conceptual replication of the research of Jarosz and Jaeger.

## Study 1

### Method

**Participants.** Two-hundred and forty-seven undergraduate students accepted the invitation to participate in the study. However, after exclusion criteria were applied (see Results for details), data for 205 participants (83.0%) were analyzed (147 females;  $M_{age} = 22.4$  years,  $SD = 5.9$ ). Of these participants, all reported speaking fluent English, with 75.6% reporting English as their first language. After English, the most common first languages reported were Mandarin (6.3%), Arabic (5.9%), and French (2.9%). The remaining 9.3% of participants reported other first languages with a frequency of less than 1%. The study was approved by the Carleton University Research Ethics Board.

**Measures.**

*The Math Background and Interests Questionnaire (MBIQ).* The MBIQ (LeFevre, Smith-Chant, Hiscock, Daley, & Morris, 2003) is a 20-item measure of math attitudes ( $n = 3$ ), language and writing skills ( $n = 6$ ), and demographic information ( $n = 11$ ). Demographic information included questions with respect to gender, age, and program of study. Math attitudes, and language and writing skills were reported on a 5-point Likert-type scale, ranging from “1” (i.e., very low/almost never) to “5” (i.e., very high/always). With the exception of age and first language, these data were not used in the present research.

*Math Story Problems.* Six counterbalanced problem sets composed of 32 multiplication and 16 division problems were created based on the format used by Mattarella-Micke and Beilock (2010). Problems required single-digit multiplication with operands from 2 to 9 (e.g.,  $4 \times 5$ ), excluding ties (e.g.,  $4 \times 4$ ); division problems used the corresponding operands (e.g.,  $20 \div 4$ ). For multiplication problems, the scenarios were either associative or dissociative and contained either highly-interfering (i.e., sum of the critical problem numbers) or less-interfering numbers (i.e., numbers similar in magnitude to the interfering number). For division problems, the scenarios were either associative or dissociative and contained the same number used in the corresponding highly-interfering multiplication problem. Mattarella-Micke and Beilock had an equal number of multiplication and division problems (36 of each, total of 72). They described division as filler problems and did not provide details of their design or any analyses. Jarosz and Jaeger had 32 multiplication and 32 division problems (total of 64) in their problem set. Because our study was designed prior to the Jarosz and Jaeger (2019) publication, we did not have highly- and less-interfering conditions for division. In all studies, half of the problems in each operation were associative and half were dissociative.

The story problems consisted of two components, although the whole story problem was presented simultaneously to the participants. The first part introduced the story problem protagonist and a scenario that included a number of items (highly-interfering or less-interfering for multiplication) that were either associated with or dissociated from the protagonist (see Table 1 for examples). The second part presented a mathematical question that required either multiplication or division. The highly-interfering numbers were the sums of the two numbers presented in the multiplication problem (e.g., “5” for  $3 \times 2$ ). The less-interfering numbers were similar in magnitude to the highly-interfering numbers (e.g., “4” for  $3 \times 2$ ). For division problems, the interfering number was the same number used in the highly-interfering multiplication problem (e.g., “5” for  $6 \div 3$ ).

Table 1. *Examples of math story problems used in the present research.*

Operation	Story Text
Multiplication	Jane, a wristwatch model, is preparing for her next photoshoot. <i>She put nail polish on 5/4/some fingers/She took nail polish off 5/4/some fingers</i> and then headed to work. Jane had 2 photoshoots today and each shoot lasted 3 hours. How many hours did Jane work today?
Division	Amy, a museum tour guide, was busy preparing for the next big tour. <i>She picked up 15/some maps/She put down 15/some maps</i> and collected payment from the tour group attendees. She collects \$56 in total. If each person in the group paid \$7, how many people are in her tour group?

*Note.* Italics are used in this table to highlight the information that was manipulated. Associative/dissociative content and highly-interfering number/less-interfering number/interfering-word (multiplication) and interfering number/interfering word (division) content is separated by slashes.

Mattarella-Micke and Beilock (2010) and Jarosz and Jaeger (2019) presented the components of the story problems sequentially. However, in the present research our goal was to make the story problems similar to those encountered in typical situations. In Mattarella-Micke and Beilock (2010), participants were asked to rate the introduction sentences for clarity and similarity to other passages to ensure that they read all of the text. In contrast, in the present research, participants were asked to answer a three-item multiple-choice comprehension problem related to the protagonist (e.g., “What kind of model was Jane?” – swimsuit, wristwatch, shoe or “Where does Amy give tours?” – zoo, museum, or art gallery) on a separate screen after responding to the story problem. The comprehension question was included to ensure that participants read the whole story problem, including the associating or dissociating information and the interfering number.

**Procedure.** Participants were recruited through an online system and received course credit for their participation. Participants logged in, selected the study, and received a link to a survey created with an online survey tool (i.e., Qualtrics). Electronic informed consent was obtained prior to the start of the survey. Participants completed the MBIQ followed by the story problems. Prior to beginning the story problems, participants were asked to complete the problems as quickly and accurately as possible and informed that they could use pencil and paper if needed but were asked not to use a calculator. After participants responded to the story problem by typing a number into the box, the problem disappeared, and the comprehension question appeared. A single random order of the arithmetic problems was used, but different participants saw different initial scenarios for a particular problem. There were six different lists of problems (3 [multiplication highly-interfering; multiplication less-interfering, division] by 2 [associative, dissociative]). Participants were randomly assigned to one of the six lists.

After every eight story problems, an entertaining meme (i.e., a humorous photo combined with text) appeared on the screen. The meme signalled to participants that they could take a break for up to three minutes before continuing with the story problem task. The memes were inserted to reduce fatigue and encourage engagement so that participants would try their best and answer all of the questions. The study took approximately 60 minutes to complete. Stimuli, data, and analyses for Study 1 and Study 2 can be found here: <https://osf.io/8kvef/>.

## Results

Several criteria were used to exclude participants who did not seem to be putting forth their best effort. First, data were excluded for participants who obtained a total score of less than 50% on the mathematical part of the story problems ( $n = 40$ ). Second, participants who failed to answer at least 8 of the last 10 story problems were excluded ( $n = 2$ ), on the assumption that they had decided not to complete the study. Thus, 205 participants were included in the analyses.

Analyses were conducted using repeated-measures ANOVAs. Error bars in figures are standard errors of the means, to be consistent with Mattarella-Micke and Beilock (2010) and Jarosz and Jaeger (2019). Furthermore, in addition to reporting frequentist statistics, Bayes factors were computed to evaluate the fit of the data under the null and alternative hypotheses. The Bayes factor,  $BF_{01}$ , is “a ratio that contrasts the likelihood of the data fitting under the null hypothesis with the likelihood of fitting under the alternative hypothesis” (Jarosz & Wiley, 2014). For example, a Bayes factor of 12.2 indicates that the data are 12.2 times more likely to occur under the null hypothesis than the alternative hypothesis (Jarosz & Wiley, 2014). Taking the inverse,  $BF_{10}$ , puts the Bayes factor in terms of the alternative hypothesis (e.g.,  $BF_{01} = 12.2$ ,  $BF_{10} = 1/12.2 = 0.082$ ). The Bayes factors were calculated in JASP. All interpretations of the Bayes factors with respect to the strength of the evidence for the null or alternative hypothesis in

the present analyses are in accordance with the guidelines developed by Jeffreys (1961) and as listed in Table 4 of Jarosz and Wiley (2014).

**Story Problems.** Story problem comprehension scores ( $M = 89.6\%$ ,  $SD = 10.9\%$ ) suggested that participants were reading the story problems in their entirety. Participants made errors on 7.2% ( $SD = 11.2$ ) of multiplication problems and 10.9% ( $SD = 16.2$ ) of division problems. This level of performance was similar to the percentage of errors for multiplication ( $M_s = 5 - 7\%$ ) and division ( $M_s = 7 - 9\%$ ) reported by Jarosz and Jaeger (2019) in three studies and to the percentage of error reported by Mattarella-Micke and Beilock (2010) for multiplication ( $M_s$  approximately 6 - 7%) in two studies.

**Multiplication Problems.** To replicate the analyses performed by Mattarella-Micke and Beilock (2010), percentage error for multiplication story problems were analyzed in a 2 (association: associated, dissociated) by 2 (interfering number: highly, less) repeated measures ANOVA (see Figure 1; compare to Figures 1 and 2 of Mattarella-Micke & Beilock). The effect of association was not significant,  $F(1, 204) = 0.14$ ,  $p = 0.71$ ,  $\eta_p^2 = 0.001$ . The estimated Bayes factor,  $BF_{01} = 12.2$ , indicates strong evidence for the null hypothesis. The effect of interference was not significant,  $F(1, 204) = 0.33$ ,  $p = .57$ ,  $\eta_p^2 = .002$ . The estimated Bayes factor,  $BF_{01} = 11.2$ , indicates strong evidence for the null hypothesis. Critically, in contrast to the results of Mattarella-Micke and Beilock, the interaction between interference and association was not significant,  $F(1, 204) = 1.06$ ,  $p = .31$ ,  $\eta_p^2 = .005$ . The estimated Bayes factor,  $BF_{01} = 7.1$ , indicates substantial evidence for the null hypothesis. In summary, no evidence was found for the pattern of results reported by Mattarella-Micke and Beilock.

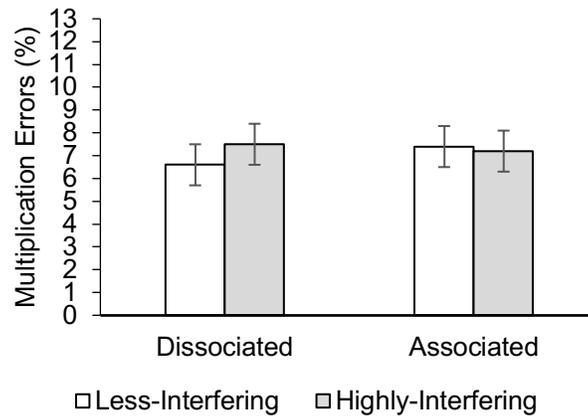
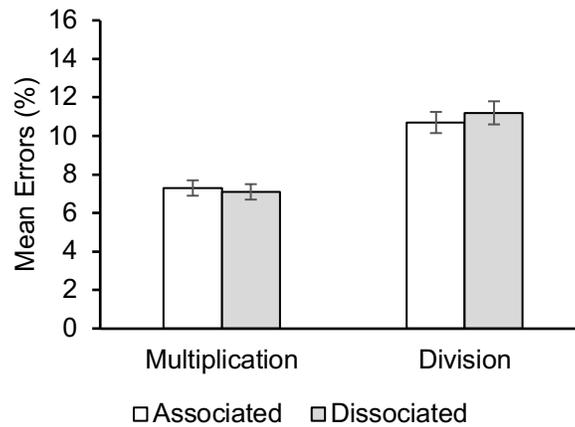


Figure 1. Mean multiplication errors (%) as a function of association and interference. Error bars are standard errors of the mean.

**Comparison across operations.** To compare our results to those of Jarosz and Jaeger (2019), story problem errors were analyzed in a 2 (association: associated, dissociated) x 2 (operation: multiplication, division) repeated measures ANOVA (see Figure 2). Replicating Jarosz and Jaeger, there was a significant main effect of operation,  $F(1, 204) = 17.63, p < .001, \eta_p^2 = .08$ . People made more errors on division problems ( $M = 10.9\%$ ) than on multiplication problems ( $M = 7.2\%$ ). The estimated Bayes factor,  $BF_{10} = 1.220e+5$ , indicates decisive evidence for the effect of operation.

In contrast, there was no significant main effect of association condition ( $M = 9.0\%$  for associated;  $M = 9.1\%$  for dissociated),  $F(1, 204) = 0.07, p = .80, \eta_p^2 = .001$ . The estimated Bayes factor,  $BF_{01} = 12.7$ , indicates strong evidence for the null hypothesis. Furthermore, in contrast to Jarosz and Jaeger, the interaction of association and operation was not significant,  $F(1, 204) = 0.33, p = .57, \eta_p^2 = .002$ . The estimated Bayes factor,  $BF_{01} = 7.7$ , indicates substantial evidence for the null hypothesis. Thus, we replicated the effect of operation, but did not find evidence that dissociated contexts influenced problem solving for division problems.



*Figure 2.* Mean story problem errors (%) as a function of association and operation. Error bars are standard errors of the mean.

## Discussion

We failed to replicate the findings of Mattarella-Micke and Beilock (2010) and thus did not find support for the foregrounding hypothesis. Neither textual association nor numerical interference significantly affected story problem performance for multiplication (cf. Figures 1 and 2 in Mattarella-Micke and Beilock vs. Figure 1 in the present study). We also did not replicate the findings of Jarosz and Jaeger (2019) who found that dissociative contexts resulted in more errors on division problems than associative contexts (cf. Figures 1 to 3 in Jarosz & Jaeger vs. Figure 2 in the present study).

The stimuli used in the present study were constructed using the same criteria as Mattarella-Micke and Beilock (2010) and Jarosz and Jaeger (2019), but they were not identical. We assumed that, if association and interference truly influenced story problem solving, then the findings should be replicable with similar stimuli. We ensured that the story problems were counterbalanced such that each story appeared in every condition. It was not possible to

determine if Mattarella-Micke and Beilock used this stringent criterion based on the information they provided, however, Jarosz and Jaeger did counterbalance stories across conditions. Careful problem construction and counterbalancing greatly reduced the possibility that one or a few poorly constructed story problems in a particular condition might drive the pattern of results.

### **Study 2**

In Study 2, we attempted to provide a test of the foregrounding hypothesis that would maximize the opportunity for an interfering number in an associative context to influence problem solving. Thus, Study 2 was identical to Study 1 with one exception; instead of presenting story problems that included either a highly-interfering or less-interfering number, the associated and dissociated multiplication problems contained either a highly-interfering number (as in Study 1) or a non-numeric word. Division problems all had non-numeric scenarios.

According to Mattarella-Micke and Beilock (2010) and the foregrounding hypothesis, only numerical information should interfere with problem solving. Thus, we directly compared numerical interference (i.e., the same number used in the highly-interfering story problems in Study 1) to non-numerical interference (i.e., an interfering word as opposed to a number). For example, “She put/removed nail polish on/from 5 fingers and then headed to work” was compared with “She put/removed nail polish on/from *some* fingers and then headed to work”. Note that, in their Study 3, Jarosz and Jaeger (2019) found an interaction between operation and association, even without any numerical interference. They concluded that the interaction was driven by the relation between the protagonists’ actions and the required operation, not by numerical interference. Thus, Study 2 provides a partial conceptual replication of the effects for division problems reported by Jarosz and Jaeger (Study 3).

## Method

**Participants.** Four hundred and twenty-five undergraduate students at a Canadian university accepted the invitation to participate in the study. As in Study 1, participants were excluded if they obtained a total score of less than 50% on the mathematical part of the story problems ( $n = 64$ ). Participants who failed to answer at least 8 of the last 10 story problems were also excluded ( $n = 2$ ), on the assumption that they had decided not to complete the study. After exclusion criteria were applied, data for 359 participants (84.5%) were analyzed (240 females;  $M_{age} = 20.1$  years,  $SD = 3.8$ ). All retained participants reported speaking fluent English, with 81.3% reporting English as their first language. After English, the most common first languages reported were Mandarin (5.0%), Arabic (3.6%), and French (3.6%). The remaining 6.5% of participants reported other first languages with a frequency of less than 1%.

**Measures.** With the exception of the changes made to the interference component of the story problems, the measures for Study 2 were the same as those reported in Study 1.

**Procedure.** The procedure for Study 2 was the same as that reported in Study 1.

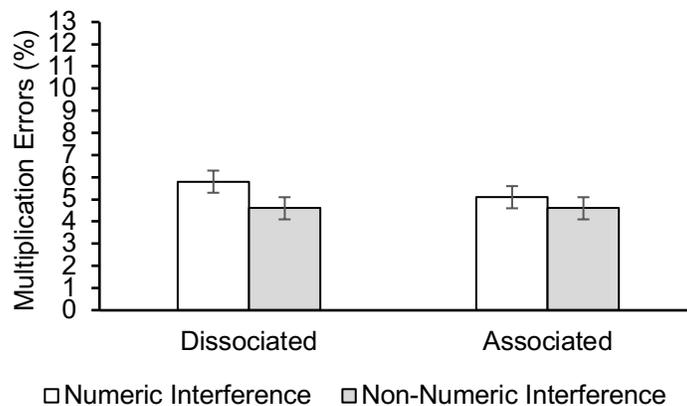
## Results

**Story problems.** Story problem comprehension scores ( $M = 88.8\%$ ,  $SD = 9.9\%$ ) suggested that participants were reading the story problems in their entirety. Participants made errors on 5.0% ( $SD = 8.1\%$ ) of multiplication problems and 12.4% ( $SD = 19.3\%$ ) of division problems. These means are similar to those reported in Study 1, and in Mattarella-Micke and Beilock (2010) and Jarosz and Jaeger (2019).

**Multiplication problems.** Percentage of errors for multiplication story problems was analyzed in a 2 (association: associated, dissociated) by 2 (interference: numeric, non-numeric) repeated measures ANOVA (see Figure 3). There was a significant effect of interference,  $F(1,$

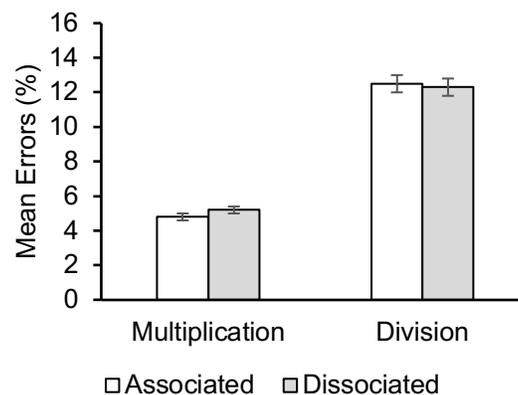
358) = 4.28,  $p = .04$ ,  $\eta_p^2 = .012$ . People made more errors on problems that contained numeric ( $M = 5.4\%$ ) than non-numeric ( $M = 4.6\%$ ) scenarios. However, the estimated Bayes factor,  $BF_{10} = 0.13$ , indicates anecdotal evidence for the null hypothesis. Thus, there is not enough evidence to support the hypothesis that the presence of extraneous numbers made the problems more difficult.

The effect of association was not significant,  $F(1, 358) = 0.83$ ,  $p = 0.36$ ,  $\eta_p^2 = 0.002$ . The estimated Bayes factor,  $BF_{01} = 11.8$ , indicates strong evidence for the null hypothesis. Furthermore, the interaction between interference and association was not significant,  $F(1, 358) = 0.93$ ,  $p = .34$ ,  $\eta_p^2 = 0.003$ . The estimated Bayes factor,  $BF_{01} = 7.7$ , indicates substantial evidence for the null hypothesis. Thus, there was fairly strong evidence in favour of the null hypothesis for the effects of association and the interaction between association and interference. We again failed to replicate the findings of Mattarella-Micke and Beilock (2010).



*Figure 3.* Mean multiplication errors (%) as a function of association and interference. Error bars are standard errors of the mean.

**Comparison across operations.** To compare the results to those of Jarosz and Jaeger (2019), percentages of errors were analyzed in a 2 (association: associated, dissociated) x 2 (operation: multiplication, division) repeated measures ANOVA (see Figure 4). There was a significant main effect of operation,  $F(1, 359) = 63.34, p < 0.001, \eta_p^2 = 0.15$ . People made more errors on division problems ( $M = 12.4\%$ ) than on multiplication problems ( $M = 5.0\%$ ). The estimated Bayes factor,  $BF_{10} = 6.427e+25$ , indicates decisive evidence for the effect of operation. However, the effect of association was not significant ( $M = 8.7\%$  for associated;  $M = 8.8\%$  for dissociated),  $F(1, 358) = 0.03, p = 0.87, \eta_p^2 = 0.001$ . The estimated Bayes factor,  $BF_{01} = 16.7$ , indicates strong evidence for the null hypothesis. The association by operation interaction was also not significant,  $F(1, 358) = 0.38, p = 0.54, \eta_p^2 = 0.001$  (see Figure 4). The estimated Bayes factor,  $BF_{01} = 11.1$  indicates strong evidence for the null hypothesis. Thus, there is decisive evidence that the model should include an effect of operation and strong support for the null hypothesis for all other effects.



*Figure 4.* Mean story problem errors (%) as a function of association and operation. Error bars are standard errors of the mean.

**Summary of accuracy analyses.** We again failed to replicate the patterns reported by Mattarella-Micke and Beilock (2010). Although we did find a small effect of an interfering number compared to a non-numerical word, the associated Bayes factor indicated that there was not much support for the model that included the effect of interference over the null model. Furthermore, although we found an effect of operation, we did not replicate the interaction between operation and association reported by Jarosz and Jaeger (2019). Thus, the results of this study did not support either the foregrounding or the inconsistent-operations hypothesis. To provide further insight into the possible effects of manipulations of operation, association, and interference on performance, we looked more closely at the types of errors that participants made.

**Error type analysis.** The foregrounding and inconsistent-operations hypotheses make specific predictions about the types of errors that people will make when they solve story problems that contain associative language and numerical interference. Thus, to better understand the types of errors individuals made, both in general and across each condition, errors for each story problem were analyzed, combining across Study 1 and Study 2.

Table 2 shows the error classifications for 1,463 out of the 2,112 errors participants made in both studies. The remaining 649 errors could not be classified into categories – they were either left blank or appeared to be mistyping errors (e.g.,  $7 \times 9 = 634$ ). Classifiable errors fell into two main categories: wrong operation errors and calculation errors. The foregrounding and inconsistent-operations hypotheses predict that participants will make *specific* wrong operation errors – using addition on multiplication problems with associated contexts and using subtraction on division problems with dissociated contexts. As shown in Table 2, however, addition and

subtraction errors were infrequent (1.2% and 0.7% of total classifiable errors) and were approximately equally distributed across the story problem conditions. Instead, people frequently chose the wrong operation on division problems – multiplying instead of dividing (48.3% of classifiable errors). They rarely divided on multiplication problems (3.1% of classifiable errors), instead, most errors on multiplication problems were arithmetic miscalculations.

The arithmetic calculation errors observed in the present studies are common in research on simple arithmetic (Campbell, 1995; Campbell & Xue, 1 C.E.) *Table-related multiplication errors* occurred when solvers used a multiplication fact that was one larger or smaller than the correct response (17.4% of classifiable errors; e.g.,  $9 \times 6 = 48$ , the product of  $8 \times 6$ ). *Table-related division errors* occurred when solvers' answers for division problems were one unit off of the correct answer. For example, if the division question was  $21 \div 3$ , participants responded with 8 or 6 instead of 7 (10.8% of classifiable errors). *Other numerical errors* occurred when solvers' responses were numerically close to the correct response but did not meet the criteria for being a table-related error (18.5% of classifiable errors). For example, they responded to  $9 \times 4$  with 34 (i.e., an operand intrusion error; Campbell, 1994).

Operation errors were more common than calculation errors for division (77% vs. 23%) but less common than calculation errors for multiplication (12% vs. 88%),  $\chi^2(1, N = 1,463) = 564.98, p < .001$ . Participants were probably biased to select multiplication because there were twice as many multiplication problems as division problems. Interestingly, only 31% of multiplication operation errors occurred on problems where both operands were less than 12, suggesting that even when problems had a large operand (e.g.,  $63 \div 7$ ) participants still multiplied when they should have divided.

Among division errors, the percentage of errors on associative versus dissociative story problems did not differ for either operation errors (47% vs. 53%) or calculation errors (53% vs. 47%),  $\chi^2(1, N = 932) = 2.43, p = .12$ . Similarly, among multiplication errors, the percentage of errors did not differ for associative or dissociative story problems for operation errors (57% vs. 43%) or calculation errors (50% vs. 50%),  $\chi^2(1, N = 531) = 1.09, p = .30$ . These patterns of errors suggest that the association/dissociation manipulation did not influence the types of errors solvers made for either multiplication or division problems.

With respect to the foregrounding and inconsistent-operations hypotheses, Table 2 shows that participants rarely made addition and subtraction errors. Thus, there is little support for the idea that story problem content primes specific operations that would lead to specific types of errors. Instead, the majority of multiplication errors were the result of miscalculations and the majority of division errors were the result of choosing the wrong operation. In summary, there was no evidence to suggest that the association/dissociation portion of the story problem influenced error patterns.

Table 2. Summary of common errors made across each condition across Studies 1 and 2 (see text for error descriptions).

Error Type	Conditions								Total
	Division				Multiplication				
	Associated		Dissociated		Associated		Dissociated		
			Interfering		Non-Interfering		Interfering		
Wrong Operation Errors									
• Multiplication	333	373	--	--	--	--	--	--	<b>706</b>
• Division	--	--	13	13	8	12			<b>46</b>
• Addition	5	3	4	2	3	1			<b>18</b>
• Subtraction	1	0	3	2	3	1			<b>10</b>
Calculation Errors									
• Table-related multiplication	--	--	57	73	61	64			<b>255</b>
• Table-related division	91	67	--	--	--	--			<b>158</b>
• Other numerical errors	25	34	57	46	57	51			<b>270</b>
<b>Total Classifiable Errors</b>	<b>455</b>	<b>477</b>	<b>134</b>	<b>136</b>	<b>132</b>	<b>129</b>			<b>1463</b>

### General Discussion

In two studies, we examined adults' performance on math story problems that included either associative or dissociative relations between objects and a protagonist. We tested the foregrounding hypothesis with a conceptual replication of two studies by Mattarella-Micke and Beilock (2010). We tested the inconsistent-operations hypothesis with a partial conceptual replication of three studies by Jarosz and Jaeger (2019). Additionally, in Study 1, we examined the effect of numeric interference (i.e., highly- vs. less-interfering) on problem solving whereas in Study 2 we examined the effect of numeric versus non-numeric interference on problem solving.

The results did not support either the foregrounding hypothesis (Mattarella-Micke & Beilock, 2010) or the inconsistent-operations hypothesis (Jarosz & Jaeger, 2019). The foregrounding hypothesis predicts that highly-interfering numerical information that is associated with the protagonist of the story will result in increased error rates on multiplication problems, presumably causing solvers to use an incorrect operation (i.e., addition instead of multiplication). Mattarella-Micke and Beilock found this pattern of increased errors in two studies, however, in the studies reported here, we found no support for the foregrounding hypothesis. The inconsistent-operations hypothesis (Jarosz & Jaeger, 2019) predicted that associative or dissociative activities of the protagonist would interfere with multiplication or division, respectively – that is, when the operation implied by the protagonists' activities was inconsistent with the operation required in the word problem, performance would be worse. In three studies reported in their paper, they found partial support for their hypothesis (i.e., only with division and dissociative actions). In the present research, however, we did not find evidence for the inconsistent-operations hypothesis with either multiplication or division.

In addition to the analyses of accuracy reported in the two prior studies, we looked more closely at the types of errors that participants made across story problem conditions. We found that participants rarely made the specific types of errors predicted by either the foregrounding or inconsistent-operations hypotheses, that is, using addition or subtraction instead of multiplication or division. In contrast, participants frequently made calculation errors on multiplication problems and chose the incorrect operation on division problems. Thus, the present studies showed no empirical evidence for either the foregrounding or the inconsistent-operations hypothesis, suggesting that information being associated with or dissociated from the protagonist of a story problem does not affect math word problem solving on these computationally simple problems.

There were four important differences between our studies and the two studies reported by Mattarella-Micke and Beilock (2010). First, in the present research we used story problems that were counterbalanced across the experimental conditions. In contrast, it appears that Mattarella-Micke and Beilock assigned specific story problems and numerical values to specific conditions. Thus, it is possible that specific story problems were more difficult than others in their study, resulting in the small differences they observed across conditions. Second, Mattarella-Micke and Beilock had very small sample sizes for both studies (i.e., 38 and 21, respectively) whereas we had much larger sample sizes. Our data provide more confidence that effects are not due to the presence of extreme scores in small biased samples.

The third methodological difference between the previous research and the current work was that we presented each story problem as a unit whereas Mattarella-Micke and Beilock (2010; see also Jarosz & Jaeger, 2019) presented the associative/dissociative context prior to the remainder of the problem. We chose this approach to increase the ecological validity of the

problem-solving process and minimize the possibility that participants would notice that the initial foregrounding context was irrelevant to the arithmetic problem. The fourth difference between the previous work and ours was that we used comprehension questions to ensure that participants had read the whole text, whereas in Mattarella-Micke and Beilock, participants rated the clarity of the text in comparison to other texts. Using comprehension questions allowed us to determine whether participants had read the associative/dissociative component of the story problem. We expected that our revised procedures would increase the chance of finding the predicted patterns if the foregrounding hypothesis was correct, because the critical text was presented simultaneously with the rest of the problem and the comprehension questions focussed on that content. In summary, despite the minor methodological differences across studies, we are confident that our design decisions, inclusion of Bayesian statistics, and large sample provided a reasonable test of the foregrounding hypothesis.

In comparing our results to those of Jarosz and Jaeger (2019), we found some similar patterns across operations (i.e., participants made more errors on division than multiplication problems) but did not replicate their partial support for the inconsistent-operations hypothesis. Although Jarosz and Jaeger included interfering numerical information in division problems and we did not, they found no effect of that manipulation, so these design differences are presumably not the reason for the different results. As in the current work, Jarosz and Jaeger counterbalanced problems across conditions, ensuring that there was no confounding of specific story elements with associative versus dissociative activities. Their sample sizes were larger than those of Mattarella-Micke and Beilock (i.e., 66, 94, and 100 participants in three studies), although still smaller than those used in the present research. Once again, we are confident that our design

decisions, inclusion of Bayesian statistics, and large sample sizes provided a reasonable test of the inconsistent-operations hypothesis.

In summary, the results of our study do not support the foregrounding or inconsistent-operations hypotheses. In fact, Bayesian analyses suggested that we have strong evidence in support of the null hypotheses for effects of association and interfering numbers. Even in the original study, Mattarella-Micke and Beilock (2010) found only a 5% increase in multiplication errors in the critical condition. In general, across the two previous papers and in the current studies the overall percentages of error were similar: Participants quite accurately solving simple arithmetic story problems. Patterns of errors were consistent with studies in which multiplication and division problems were mixed, resulting in cross-operation confusion effects (e.g., Bell et al., 1984, 1981; Fischbein et al., 1985; Graeber et al., 1989). However, the text manipulations (i.e., associative vs. dissociative language) did not influence problem solving for our participants.

Mattarella-Micke and Beilock (2010) and Jarosz and Jaeger (2019) also examined the relations between individual differences in working memory and performance on story problems. In some of their studies, effects of association were moderated by working memory skill. In their Study 2, Mattarella-Micke and Beilock found that the increase in errors for foregrounded, interfering problems was related to individual differences in working memory for participants with lower working memory capacity. Jarosz and Jaeger, in their Study 1, found that participants with high working memory capacity made fewer errors than participants with lower working memory capacity, but they did not find interactions between working memory and association or interference. In their Study 2, participants were informed that they might be tested for their memory on any number that appeared in the story problem. Under these instructions, participants with low working memory capacity made more errors on dissociative division problems. In their

Study 3, which was identical to the first study but did not contain numeric interference, they did not find any significant effects of working memory capacity.

Taken together, the patterns found in the two previous studies suggest that individual differences in working memory capacity may sometimes influence problem solving performance, consistent with many other studies (e.g., Hambrick & Engle, 2003; Wiley & Jarosz, 2012), but there was no clear pattern of results related to the associative/dissociative or interfering manipulations. Because patterns of working memory effects were not consistent across Mattarella-Micke and Beilock and Jarosz and Jaeger, as described above, individual differences in working memory capacity seemed unlikely to be the source of the differences across studies. Accordingly, in the present studies we chose to focus on the effects of associative/dissociative text and interfering numbers on story problem performance.

### **Limitations**

One limitation of the present studies was that we were unable to obtain the original stimuli used by Mattarella-Micke and Beilock (2010). If we had access to the original stimuli, we could have performed a closer replication. However, we carefully followed the description of the stimuli outlined by Mattarella-Micke and Beilock to design similar scenarios. Furthermore, if the foregrounding hypothesis was correct, it should be supported across a range of similar stimuli, as opposed to a specific set of story problems. Future studies that are able to directly test the stimuli used in the original Mattarella-Micke and Beilock article may be able to provide more insights into why we failed to replicate their findings.

The online administration of this study is a second potential limitation. To ensure that participants were fully engaged in the study, strict exclusion criteria were implemented, which resulted in a relatively high exclusion rate (i.e., exclusion rates of 17.0% and 15.5% for Studies 1

and 2 vs. 5.2% and 4.8% in Studies 1 and 2 of Mattarella-Micke & Beilock). However, Jarosz and Jaeger also excluded a substantial number of participants in Study 1 (26.7%) and Study 3 (13.0%), so exclusion rates are unlikely to explain the different findings. The online administration was also a strength because it allowed for a large sample size and full counterbalancing across all conditions. There is considerable evidence that online psychology studies produce similar results for a variety of effects originally observed in lab studies (Chuah, Drasgow, & Roberts, 2006; Cipora, Soltanlou, Reips, & Nuerk, 2019; Germine et al., 2012; Hilbig, Moshagen, & Zettler, 2016; Ihme et al., 2009). Furthermore, with more educators implementing classroom websites and using online assessments, the online environment may be representative of the typical environment in which students encounter word problems.

### **Implications and Conclusions**

In two studies, we failed to replicate patterns of textual and numerical interference in math story problems that were reported in two previous papers (Mattarella-Micke & Beilock, 2010; Jarosz & Jaeger, 2019). Bayesian statistics suggested strong support for the null hypotheses. Our findings emphasize the importance of large sample sizes, careful manipulations, and proper counterbalancing of problem features. Additionally, our examination of the specific errors made by participants revealed further information about underlying cognitive processes involved in problem solutions. Although linguistic features of math word problems undoubtedly influence problem solving in many situations (Daroczy et al., 2015; Fuchs, Fuchs, Compton, Hamlett, & Wang, 2015; Neshet, Hershkovitz, & Novotna, 2003; Thevenot & Barrouillet, 2015), our findings suggest that manipulating the associative relation between the protagonist and problem elements, specifically as irrelevant details unrelated to the situation model of the story problem, has no effect on adults' ability to solve simple word problems. In summary, the present

results suggest boundary conditions on the extent to which textual manipulations influence adults' problem-solving performance.

**Funding:** This project was funded by Natural Sciences and Engineering Research Council of Canada through a Discovery Grant and an Equipment Grant to Jo-Anne LeFevre and a Social Sciences and Humanities Research Council doctoral fellowship to Sabrina Di Lonardo Burr

**Declaration of Conflicting Interests:** The authors declare that there is no conflict of interest.

### References

- Bassok, M., Pedigo, S. F., & Oskarsson, A. T. (2008). Priming addition facts with semantic relations. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *34*(2), 343–352. <https://doi.org/10.1037/0278-7393.34.2.343>
- Bell, A., Fischbein, E., & Greer, B. (1984). Choice of operation in verbal arithmetic problems: The effects of number size, problem structure and context. *Educational Studies in Mathematics*, *15*(2), 129–147. <https://doi.org/10.1007/BF00305893>
- Bell, A., Swan, M., & Taylor, G. (1981). Choice of operation in verbal problems with decimal numbers. *Educational Studies in Mathematics*, *12*(4), 399–420. <https://doi.org/10.1007/BF00308139>
- Butterworth, B., Zorzi, M., Girelli, L., & Jonckheere, A. R. (2001). Storage and retrieval of addition facts: The role of number comparison. *The Quarterly Journal of Experimental Psychology Section A*, *54*(4), 1005–1029. <https://doi.org/10.1080/713756007>
- Campbell, J. I. D. (1994). Architectures for numerical cognition. *Cognition*, *53*, 1–44.
- 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., & Metcalfe, A. W. S. (2007). Arithmetic rules and numeral format. *European Journal of Cognitive Psychology*, *19*(3), 335–355. <https://doi.org/10.1080/09541440600717610>
- Campbell, J. I. D., & Timm, J. C. (2000). Adults' strategy choices for simple addition: Effects of retrieval interference. *Psychonomic Bulletin & Review*, *7*(4), 692–699. <https://doi.org/10.3758/BF03213008>

- Campbell, J. I. D., & Xue, Q. (1 C.E.). Cognitive arithmetic across cultures. *Journal of Experimental Psychology. General*, *130*(2), 299–315.
- Chuah, S. C., Drasgow, F., & Roberts, B. W. (2006). Personality assessment: Does the medium matter? No. *Journal of Research in Personality*, *40*(4), 359–376.  
<https://doi.org/10.1016/j.jrp.2005.01.006>
- Cipora, K., Soltanlou, M., Reips, U.-D., & Nuerk, H.-C. (2019). The SNARC and MARC effects measured online: Large-scale assessment methods in flexible cognitive effects. *Behavior Research Methods*, *51*(4), 1676–1692. <https://doi.org/10.3758/s13428-019-01213-5>
- Daroczy, G., Wolska, M., Meurers, W. D., & Nuerk, H.-C. (2015). Word problems: A review of linguistic and numerical factors contributing to their difficulty. *Frontiers in Psychology*, *06*. <https://doi.org/10.3389/fpsyg.2015.00348>
- Fischbein, E., Deri, M., Nello, M. S., & Marino, M. S. (1985). The role of implicit models in solving verbal problems in multiplication and division. *Journal for Research in Mathematics Education*, *16*(1), 3. <https://doi.org/10.2307/748969>
- Fuchs, L. S., Fuchs, D., Compton, D. L., Hamlett, C. L., & Wang, A. Y. (2015). Is word-problem solving a form of text comprehension? *Scientific Studies of Reading*, *19*(3), 204–223.  
<https://doi.org/10.1080/10888438.2015.1005745>
- Fuchs, L. S., Gilbert, J. K., Fuchs, D., Seethaler, P. M., & Martin, B. L. (2018). Text comprehension and oral language as predictors of word-problem solving: Insights into word-problem solving as a form of text comprehension. *Scientific Studies of Reading*, *22*(2), 152–166. <https://doi.org/10.1080/10888438.2017.1398259>
- Germine, L., Nakayama, K., Duchaine, B. C., Chabris, C. F., Chatterjee, G., & Wilmer, J. B. (2012). Is the web as good as the lab? Comparable performance from Web and lab in

- cognitive/perceptual experiments. *Psychonomic Bulletin & Review*, *19*(5), 847–857.  
<https://doi.org/10.3758/s13423-012-0296-9>
- Glenberg, A., Meyer, M., & Lindem, K. (1987). Mental models contribute to foregrounding during text comprehension. *Journal of Memory and Language*, *26*, 69–83.
- Graeber, A. O., Tirosh, D., & Glover, R. (1989). Preservice teachers' misconceptions in solving verbal problems in multiplication and division. *Journal for Research in Mathematics Education*, *20*(1), 95. <https://doi.org/10.2307/749100>
- Hambrick, D. Z., & Engle, R. W. (2003). The role of working memory in problem solving. In J. E. Davidson & R. J. Sternberg (Eds.), *The Psychology of Problem Solving* (1st ed., pp. 176–206). Cambridge University Press. <https://doi.org/10.1017/CBO9780511615771.007>
- Hilbig, B. E., Moshagen, M., & Zettler, I. (2016). Prediction consistency: A test of the equivalence assumption across different indicators of the same construct: Prediction consistency. *European Journal of Personality*, *30*(6), 637–647.  
<https://doi.org/10.1002/per.2085>
- Ihme, J. M., Lemke, F., Lieder, K., Martin, F., Müller, J. C., & Schmidt, S. (2009). Comparison of ability tests administered online and in the laboratory. *Behavior Research Methods*, *41*(4), 1183–1189. <https://doi.org/10.3758/BRM.41.4.1183>
- Jarosz, A. F., & Jaeger, A. J. (2019). Inconsistent operations: A weapon of math disruption. *Applied Cognitive Psychology*, *33*(1), 124–138. <https://doi.org/10.1002/acp.3471>
- Jarosz, A. F., & Wiley, J. (2014). What are the odds? A practical guide to computing and reporting bayes factors. *The Journal of Problem Solving*, *7*(1).  
<https://doi.org/10.7771/1932-6246.1167>
- Jeffreys, H. (1961). *Theory of probability* (3rd ed.). Oxford, UK: Oxford University Press.

- Kintsch, W., & van Dijk, T. A. (1978). Toward a model of text comprehension and production. *Psychological Review*, 85(5), 363–394. <https://doi.org/10.1037/0033-295X.85.5.363>
- Kouba, V. L. (1989). Children's solution strategies for equivalent set multiplication and division word problems. *Journal for Research in Mathematics Education*, 20(2), 147. <https://doi.org/10.2307/749279>
- 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*, 125(3), 284–306. <https://doi.org/10.1037/0096-3445.125.3.284>
- LeFevre, J.-A., Smith-Chant, B. L., Hiscock, K., Daley, K. E., & Morris, J. (2003). Young adults' strategic choices in simple arithmetic: Implications for the development of mathematical representations. In A. J. Baroody & A. Dowker, *The Development of Arithmetic Concepts and Skills: Constructing Adaptive Expertise* (pp. 203–228). Mahwah, NJ: Erlbaum.
- Mattarella-Micke, A., & Beilock, S. L. (2010). Situating math word problems: The story matters. *Psychonomic Bulletin & Review*, 17(1), 106–111. <https://doi.org/10.3758/PBR.17.1.106>
- Mayer, R. E., & Hegarty, M. (1996). The process of understanding mathematical problems. In R. J. Stemberg & T. Ben-Zeev, *The Nature of Mathematical Thinking* (pp. 29–53). Mahwah, NJ: Erlbaum.
- Mulligan, J. T., & Mitchelmore, M. C. (1997). Young children's intuitive models of multiplication and division. *Journal for Research in Mathematics Education*, 28(3), 309. <https://doi.org/10.2307/749783>
- Nesher, P., Hershkovitz, S., & Novotna, J. (2003). Situation model, text base and what else? Factors affecting problem solving. *Educational Studies in Mathematics*, 52(2), 151–176.

- Schley, D. R., & Fujita, K. (2014). Seeing the math in the story: On how abstraction promotes performance on mathematical word problems. *Social Psychological and Personality Science*, 5(8), 953–961. <https://doi.org/10.1177/1948550614539519>
- Siegler, R. S. (1988). Strategy choice procedures and the development of multiplication skill. *Journal of Experimental Psychology*, 117, 258–275.
- Thevenot, C., & Barrouillet, P. (2015). Arithmetic word problem solving and mental representations. In R. Cohen Kadosh & A. Dowker, *The Oxford handbook of numerical cognition* (pp. 158–179). Oxford: Oxford University Press.
- Wiley, J., & Jarosz, A. F. (2012). How working memory capacity affects problem solving. In *Psychology of Learning and Motivation* (Vol. 56, pp. 185–227). Elsevier. <https://doi.org/10.1016/B978-0-12-394393-4.00006-6>
- Zwaan, R. A., & Radvansky, G. A. (1998). Situation models in language comprehension and memory. *Psychonomic Bulletin*, 123, 162–185.