Situating math word problems: The story matters

ANDREW MATTARELLA-MICKE AND SIAN L. BEILOCK
University of Chicago, Chicago, Illinois

In two experiments, we explored how the situation model of a math story problem impacts math problem performance. Participants completed multiplication story problems in which a set of objects was associated with or dissociated from a protagonist, making them more or less accessible in memory during answer retrieval. On the basis of previous findings that the sum of two numbers interferes with retrieval of their product, the number of objects in the math problem was either highly interfering ("9" for 4 × 5) or less interfering ("8" for 4 × 5) for multiplication retrieval in the problem. Participants made more errors in problem solving when highly interfering numerical content was associated with the protagonist and, thus, foregrounded. Moreover, the lower one's working memory, the bigger this effect. In sum, small changes in the situation model of a math story problem can harm performance. These changes shift the balance of factors that influence math performance away from math knowledge and toward individual differences in general cognitive capacity.

Studies of discourse processing suggest that people build a mental simulation—a *situation model* (van Dijk & Kintsch, 1983)—of the text that they read. Situation models represent the characters, their behavior, and the objects the characters interact with inside a text. They allow readers to reason about the text and predict likely outcomes. In short, situation models are the basis for text comprehension (Zwaan & Radvansky, 1998).

Although situation models are most often studied in the context of reading comprehension, they carry implications for other domains as well. Consider the following problem:

> Walter went to the park to have a barbecue with his friends. He brought 5 cans of soda and a package of hot dogs from his car when he got to the park. If there were 3 people at the barbecue, and each person ate 2 hot dogs, how many hot dogs were eaten in total?

This type of problem is often used to gauge mathematical knowledge (Arizona Department of Education, 2006). But completing a problem like this also requires comprehension of the text in which it is housed. If the solver constructs a mental model of the text to answer this problem, the situation model evoked by the text could have implications for math performance. Indeed, situation models that contradict readers' expectations about the mathematical operation needed to complete a problem can impair problem solving (Coquin-Viennot & Moreau, 2007). Here, we show that situation models do not have to mislead a reader to hurt performance. Rather, even small changes to the wording of a math problem can alter its situation model in a way that changes the accessibility of numerical content within the problem and negatively impacts performance.

Foregrounding in Math Story Problems

Work in reading comprehension suggests that shifts in the situation model change the accessibility of content through the highlighting, or *foregrounding*, of important story elements (Glenberg, Meyer, & Lindem, 1987). For instance, an object that is associated (picked up, carried) with a protagonist becomes foregrounded, relative to items that are dissociated (set down, discarded, etc.). Associated objects are highly accessible in memory—they are recognized and read more quickly than dissociated objects (Glenberg et al., 1987)—and are more likely to influence the reader's expectations (Cook, Gueraud, Was, & O'Brien, 2007). Similar effects occur when protagonists meet up versus separate (Gerrig & McKoon, 2001).

We hypothesized that foregrounding carries implications for math story problems because math answer retrieval depends on the accessibility of problem content. If foregrounding raises the accessibility of content that interferes with answer retrieval in a math story problem, this interfering content might be harmful for problem solving. Siegler's (1988) distribution of associations (DOA) model describes a relationship between content accessibility and math answer retrieval that supports our hypothesis.

The Distribution of Associations Model

According to the DOA model, when one is presented with a math problem (e.g., 4 × 5), both the correct answer (4 × 5 = 20) and incorrect answers (4 × 5 = 21) are activated and compete for retrieval. The efficacy of retrieval depends on the associative strength of the correct answer, relative to the incorrect answer. Thus, if an incorrect answer becomes more accessible, it should interfere with correct answer retrieval.

S. L. Beilock, beilock@uchicago.edu

Although most incorrect answers are weak competitors for retrieval, certain math operations draw upon overlapping retrieval structures that cause mutual interference. In particular, there is evidence that both multiplication and addition operations are automatically activated during either type of calculation and cause strong interference for one another (Zbrodoff & Logan, 1986). Campbell and Timm (2000) showed that when participants completed either multiplication or division problems prior to an addition task, multiplication selectively interfered with addition, resulting in reduced addition speed and accuracy, relative to the division condition. Conversely, multiplication fact retrieval is significantly slowed and less accurate when an addition answer is presented ($4 \times 5 = 9$; true or false?) versus when it is not ($4 \times 5 = 8$; Winkelman & Schmidt, 1974).

Thus, addition answers can be particularly interfering for multiplication, and vice versa. Of course, for adults who have been practicing multiplication and addition for years, under most conditions, even highly interfering answers (i.e., 7 for $3 \times 4 = ?$) are less activated than the correct answer (12). However, the automatic activation of these competing associations is sensitive to semantic context (Bassok, Pedigo, & Oskarsson, 2008). Therefore, if an interfering answer became foregrounded—and, thus, highly accessible in memory—it might disrupt answer retrieval. As a result, even slight shifts in the text of a story problem could affect math problem solving.

The Present Experiments

In two experiments, we used the association–dissociation paradigm from Glenberg et al. (1987) to demonstrate that foregrounding potentially interfering numerical content during a multiplication story problem can negatively impact performance. In Experiment 1, participants read a short passage that either associated a number of objects with the protagonist ("On his way there, Brian stopped at a department store and bought 7 shirts") or dissociated them ("On his way there, Brian stopped at a department store and returned 7 shirts"). After this passage, the participants read the rest of the math problem ("Brian bought 3 boxes of lightbulbs. Each box contained 4 lightbulbs. How many lightbulbs did Brian buy?"). On half the trials, the number of objects related to the protagonist corresponded to the highly interfering addition solution, selected for the mutual interference between multiplication and addition retrieval (e.g., for the above 3×4 problem, 7, because $3 + 4 = 7$; Winkelman & Schmidt, 1974; Zbrodoff & Logan, 1986). For the other half of trials, the number of objects was a control, selected to produce relatively low retrieval interference.

If numeric content associated with a set of objects is foregrounded, the elicited situation model should affect math performance. Specifically, when highly interfering numerical content is associated with the protagonist and, thus, made more accessible, this content should compete with retrieval of the correct answer, causing errors in problem solving. If these increased errors are, in fact, due to interference, foregrounding should relate to poor math performance only when the numeric content inter-

feres with multiplication retrieval. This is a critical test of our hypothesis: If foregrounding causes failure simply by increasing cognitive load (a manipulation shown to impair math performance; DeStefano & LeFevre, 2004), associating any set of objects with the protagonist should cause performance decrements. But if increased errors occur selectively in conditions in which the foregrounded number is highly interfering for the calculation needed to be performed (e.g., an addition answer to a multiplication problem), this will provide evidence that numerical interference is the key source of performance loss. As a preview, this is exactly what we found. Association impeded multiplication performance, but only on trials that included highly interfering numerical content.

Furthermore, if foregrounding leads to math errors through retrieval interference, individual differences in the ability to inhibit interfering associations should relate to this performance decrement. Working memory (WM) is one such individual-difference variable that is linked with the ability to inhibit competing associations (Conway & Engle, 1994). If answer retrieval depends on the ability to inhibit interference, variation in WM should correspond to math performance when interfering content (i.e., the addition answer) is foregrounded. Experiment 2 explored this possibility. If true, such a finding would suggest that math problems not constructed with findings from the situation model literature in mind might inadvertently measure solvers' WM capacity, rather than their math knowledge exclusively. This seems problematic if the primary goal of a test is to assess what math students have learned in school, rather than individual differences in general cognitive capacity.

EXPERIMENT 1

Method

Participants. Thirty-eight participants (27 females) were recruited from the participant pool at the University of Chicago (mean age = 21.14 years, $SD = 2.87$) for a "cognitive tasks study." During recruitment, there was no reference to math or sentence comprehension. The participants received \$10 or course credit for participation.

Procedure. The participants first filled out a consent form and then were told that they would be performing a computerized task involving math story problems. The participants read through a series of instructions and began the task.

The participants completed 72 problems in a completely within-subjects design. Half of the problems were multiplication problems of interest, and half were filler division problems. Division problems were designed to prevent the participants from skimming the material for numbers and immediately multiplying them, rather than thoroughly reading the information, and were not analyzed.

Multiplication problems were evenly divided in a 2 (numerical interference: high, low) $\times$ 2 (association: associated, dissociated) design. On each trial, a particular math fact (e.g., 8×7) paired with a story context was randomly sampled. Math facts were paired with a particular story to ensure story plausibility (e.g., for 8×7, scenarios such as "Bryan picked up the 15 bowling balls and went to the counter" were avoided because they could not be physically realized). This story was then randomly placed in one of the four conditions by changing the sentences and numbers (see below). Thus, for every math fact from 2×2 to 9×9 (i.e., 2×2, 2×3, etc.), we generated four problem versions that corresponded to the four conditions in our 2×2 design. This allowed us to equate problem size across conditions. Each participant randomly saw one version

of the problem, so that they received an equal number of problems in each condition.

The *association factor* was varied by changing one to three words in the sentence that either associated the set of objects with the protagonist or dissociated it from the protagonist. The *numerical interference factor* was varied by changing the number mentioned in the introduction. In the high-interference condition, this number was the highly interfering addition answer. In the low-interference condition, the number was randomly selected from the same range as the interfering numbers (4–18, corresponding to the addition solutions 2 + 2 through 9 + 9), with the only restriction being that this randomly selected number was not the highly interfering addition answer. We did not control for other potential sources of interference in the low-interference condition (e.g., interference from adjacent multiplication solutions), because these occurred rarely and would only work against finding an impact of our numerical interference factor. Condition order was randomized within and across participants.

Each problem began with a two-sentence opening passage. The first sentence introduced the scenario ("Alexis, an architect, was busy designing a new office building downtown"). The second sentence was either an associated condition ("She picked up the 4 sketches she had been working on and took the elevator to a meeting where she would deliver her ideas") or a dissociated condition ("She set aside the 4 sketches she had been working on and took the elevator to a meeting where she would deliver her ideas"). This screen was displayed separately from the main problem, so the participants were obligated to read each passage for potentially relevant information. This also allowed us to obtain introduction passage reading times. The participants then pressed the space bar to continue the problem.

The arithmetic problem itself was either a multiplication problem ("Alexis' elevator took her 9 floors down. If each floor is 8 feet tall, how many feet down did she go?") or a filler division problem ("Each floor of Alexis' building is 8 feet tall. If she went 72 feet down in the elevator, how many floors down did she go?").

The story problem remained on the screen until the participants spoke their answer into the microphone. A text box then appeared, and the participants typed in their previously stated answer. Accuracy feedback was then displayed for 1,000 msec. Lastly, the participants rated the introduction sentences for clarity ("How clear was the passage?") and similarity to other passages ("How similar was the passage?"). This instruction was included to further encourage the participants to read the opening passage completely.

Results

Multiplication problem errors and reaction time (RT) served as the dependent measures of interest. Data from 2 participants were removed because their introduction passage reading time or their math problem accuracy was not within three SDs of the group average.

We first analyzed reading times for the introductory sentence, the section of the text prior to the story problem. The participants took around 5 sec ($M = 5.49$ sec, $SD = 2.23$) to read these sentences. Sentences contained 27 words on average ($M = 27.03$, $SD = 5.6$) and depicted objects that were either associated with or dissociated from a protagonist. Reading times for sentences depicting objects dissociated from the protagonist showed a trend toward being longer than reading times for sentences associating objects with the protagonist [$t_1(35) = 2.14$, $SE = 103$, $p = .04$; $t_2(71) = 1.80$, $SE = 135$, $p = .08$].

Next, we analyzed multiplication errors with a 2 (numerical interference: high, low) × 2 (association: associated, dissociated) ANOVA using both participants (F_1) and items (F_2) as random variables. Neither the main ef-

fect of interference [$F_1(1,35) = 1.56$, $MS_e = .010$, $p = .22$; $F_2(1,35) = 1.79$, $MS_e = .011$, $p = .19$] nor that of association [$F_1(1,35) = 0.39$, $MS_e = .002$, $p = .54$; $F_2(1,35) = 0.003$, $MS_e < .001$, $p = .96$] was reliably significant. However, the interference × association interaction was significant across both participants [$F_1(1,35) = 4.33$, $MS_e = .031$, $p < .05$] and items [$F_2(1,35) = 4.98$, $MS_e = .024$, $p < .05$].

As is shown in Figure 1, in the associated condition, performance for the high-interference number problems was significantly worse than that for the low-interference problems [$t_1(35) = 2.38$, $SE = .019$, $p < .05$; $t_2(35) = 2.73$, $SE = .016$, $p < .01$]. In the dissociated condition, this difference was not significant [$t_1(35) = 0.63$, $SE = .020$, $p = .54$; $t_2(35) = -0.468$, $SE = .019$, $p = .64$]. These results suggest that the participants constructed a situation model of the story problem and that the content of this mental model had an effect on problem solving by modulating the impact of numerical interference on correct answer retrieval.

Multiplication RTs were analyzed using the same 2 (numerical interference: high, low) × 2 (association: associated, dissociated) ANOVA. Neither the main effect of interference [$F_1(1,35) = 1.81$, $MS_e = 1,766,708$, $p = .19$; $F_2(1,35) = 1.54$, $MS_e = 1,988,944$, $p = .22$] nor that of situation [$F_1(1,35) = 0.09$, $MS_e = 112,952$, $p = .76$; $F_2(1,35) = 0.572$, $MS_e = 534,903$, $p = .45$], nor the interference × association interaction, was reliably significant [*associated*, low interference, $M = 5,679$ msec, $SE = 345$; high interference, $M = 5,108$ msec, $SE = 342$; *dissociated*, low interference, $M = 5,386$ msec, $SE = 310$; high interference, $M = 5,513$ msec, $SE = 337$; $F_1(1,35) = 3.57$, $p = .07$, $MS_e = 4,384,874$; $F_2(1,35) = 7.48$, $MS_e = 7,493,115$, $p < .01$].

One might note that the interference × association interaction in RTs approached significance in the participant analysis. We thus felt it was important to verify
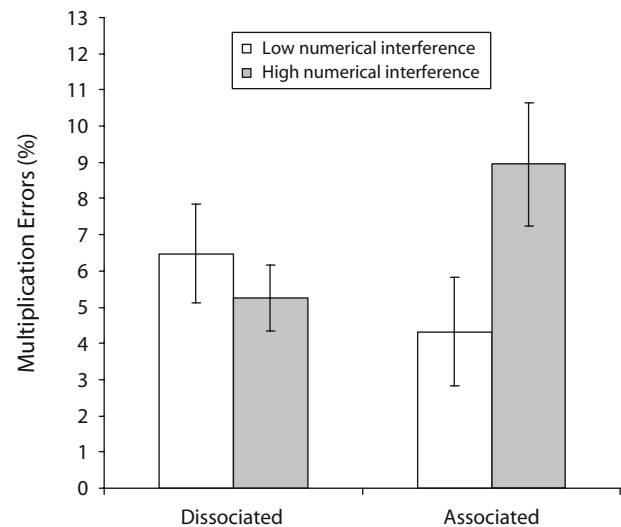

Figure 1. Mean story problem errors in Experiment 1. Error bars denote standard errors.

that this was not indicative of a trade-off with our error rates—especially for our most critical comparison in the associated condition of high-interference versus low-interference errors. This does not appear to have been the case. First, in the associated high-interference condition, multiplication errors were *not* related to RT such that the participants who made more errors were faster to solve the problems [$r(36) = .06, p = .72$]. Moreover, low-interference errors were still significantly lower than high-interference errors in the associated condition even when difference in RT across these conditions (associated–low-interference RT − associated–high-interference RT) was covaried out [$F_1(1,34) = 4.78$, $MS_e = .034, p < .05; F_2(1,34) = 5.49, MS_e = .025$, $p < .05$]. Finally, the omnibus interference × association interaction in accuracy also remained significant [$F_1(1,34) = 4.15, MS_e = .030, p < .05; F_2(1,34) = 4.83$, $MS_e = .024, p < .05$] when differences in RT across the associated conditions were covaried out.

Discussion

Participants solved math story problems that were prefaced by seemingly harmless passages about the main character. Yet, when a passage associated a set of objects with the protagonist, performance was affected by the numerical content of that set *if* the number of objects corresponded to the addition solution (an association that interferes with multiplication retrieval; Winkelman & Schmidt, 1974; Zbrodoff & Logan, 1986). This increase in errors was not seen for noninterfering numbers of objects or when these same objects were divorced from the main character. Manipulation of the situation model can affect math problem solving—specifically, by varying the accessibility of potentially interfering content.

If Experiment 1's results were due to retrieval interference, it follows that individual differences in WM should predict performance when interference is at its highest—that is, when interfering numerical content is foregrounded—because WM is critical for the inhibition of retrieval interference (Conway & Engle, 1994). Experiment 2 tested this notion by replicating Experiment 1 and examining performance as a function of individual differences in WM.

EXPERIMENT 2

Method

Participants. Twenty-one participants (9 males) who did not take part in Experiment 1 were recruited from the participant pool at the University of Chicago (mean age = 21.33 years, $SD = 2.51$), using Experiment 1's methods.

Individual differences in WM were assessed using the automated reading span (RSPAN). The RSPAN involves reading a series of sentences, followed by letters (e.g., "On warm sunny afternoons, I like to walk in the park. ? F"). Participants judge whether each sentence makes sense by clicking "True" or "False" on the computer. At the end of a series of these letter–sentence pairs, the participants recall the sequence of letters. Each series contains two to five sentence–letter pairs. Length order is determined randomly. Individuals are tested on three series of each length. We calculated RSPAN scores on the basis of the number of letters recalled in the correct order on any trial, regardless of whether the entire sequence was correct (Con-

way et al.'s [2005] recommended partial-credit scoring procedure). RSPAN scores ranged from 34 to 74 ($M = 59.22, SE = 2.49$).

Procedure. The arithmetic task was identical to that in Experiment 1, with the exception that each participant completed twice as many problems. Following this task, the participants completed the RSPAN.

Results

As in Experiment 1, the participants' data were analyzed only if their reading times on the introduction passages and their accuracy on the math problems were within three *SD*s of the group average. One participant was removed because she did not meet these criteria. Two participants were not included in the WM analysis, because they arrived late and did not have time to complete the RSPAN.

As in Experiment 1, sentence reading times averaged about 5 sec ($M = 5.09$ sec, $SD = 2.17$). There were no significant differences in reading times across the associated and dissociated condition sentences [$t_1(19) = 1.18, SE = 108, p = .25; t_2(143) = 1.19, SE = 135, p = .24$].

Multiplication errors were investigated in a 2 (numerical interference: high, low) × 2 (association: associated, dissociated) ANOVA (Figure 2). Neither the main effect of association [$F_1(1,19) = 3.70, MS_e = .005, p = .07$; $F_2(1,71) = 2.49, MS_e = .031, p = .12$] nor the main effect of interference [$F_1(1,19) = 3.39, MS_e = .019, p = .08$; $F_2(1,71) = 3.19, MS_e = .042, p = .08$] was significant. However, as in Experiment 1, there was a significant interference × association interaction [$F_1(1,19) = 10.50$, $MS_e = .019, p < .005; F_2(1,71) = 4.81, MS_e = .046$, $p < .05$]. The impact of numerical interference on performance depended on whether these numbers (represented as a set of objects) were associated with the protagonist or not. In the associated condition, performance on the high-interference number problems was significantly worse than that on the low-interference problems [$t_1(19) = 2.83$, $SE = .021, p = .01; t_2(71) = 2.33, SE = .021, p < .05$]. In the dissociated condition, this difference was not sig-
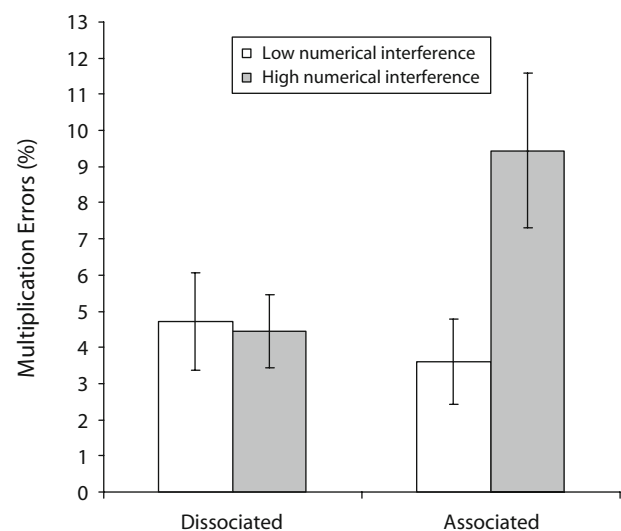

Figure 2. Mean story problem errors in Experiment 2. Error bars denote standard errors.

nificant [$t_1(19) = -0.20$, $p = .85$, $SE = .014$; $t_2(71) = -0.09$, $SE = .013$, $p = .93$].

Multiplication RTs were analyzed with a similar 2×2 ANOVA. Although there was a significant main effect of interference [$F_1(1,19) = 4.07$, $MS_e = 2,789,257$, $p = .06$; $F_2(1,71) = 5.44$, $MS_e = 18,375,403$, $p < .05$], neither the main effect of association [$M = 9$ msec, $SE = 139$; $F_1(1,19) = 0.003$, $MS_e = 1,442$, $p = .96$; $F_2(1,71) = 0.07$, $MS_e = 207,198$, $p = .79$] nor the interference × association interaction was significant [*associated*, low interference, $M = 5,912$ msec, $SE = 428$; high interference, $M = 6,312$ msec, $SE = 483$; *dissociated*, low interference, $M = 5,930$ msec, $SE = 402$; high interference, $M = 6,276$ msec, $SE = 503$; $F_1(1,19) = 0.03$, $MS_e = 14,508$, $p = .86$; $F_2(1,71) = 0.19$, $MS_e = 462,179$, $p = .668$].

Finally, we looked at how individual differences in WM related to math performance when the situation model foregrounded potentially interfering numerical content. To do this, we computed an *interference score* for each participant in both association conditions by subtracting low-interference math errors from high-interference math errors. This interference score represents the performance loss resulting from the addition of high-interfering content, using the low-interference condition as a baseline. A higher score means more interference-related performance loss for a given condition.

Using this interference score, we examined the relationship between interference-related errors and WM. There was a significant correlation between RSPAN and interference scores in the associated condition [$r(18) = -.65$, $p < .005$] (Figure 3A). When the set of objects was associated with the protagonists, the lower the participant's WM, the more errors he or she made as a result of numerical interference. This same correlation was not significant in the dissociated condition [$r(18) = -.11$, $p = .64$], when interfering content was not foregrounded (Figure 3B).

These correlations were significantly different from each other [$z(18) = -1.82$, $p < .05$, one-tailed]. Thus, when interfering material is foregrounded, WM is related to math errors in simple math story problems. When the potential for interference is minimized because inferring information is not foregrounded, this relation disappears.

Discussion

We replicated our finding of an increase in math errors for foregrounded, interfering material, as compared with nonforegrounded and noninterfering numerical content. Furthermore, we showed that this increase in errors was related to individual differences in WM. The lower the participant's WM, the more the participant's math-problem-solving accuracy was impacted when the interfering numerical information was foregrounded. Given that standardized tests are often designed to measure student learning rather than general cognitive ability, this finding provides an important caveat for test design: Features of a test that make interfering content more accessible, such as foregrounding in a math story problem, not only impair performance, but also alter what predicts success. Under these conditions, individual differences in WM play a greater role in performance, reducing the relative contribution of other factors, such as math knowledge.

GENERAL DISCUSSION

Across two experiments, participants read multiplication story problems in which sets of objects were either associated with or dissociated from a protagonist. When the number of objects in the set was also the sum of the two numbers to be multiplied (a highly interfering number for multiplication retrieval; Winkelman & Schmidt, 1974; Zbrodoff & Logan, 1986), performance suffered, but only
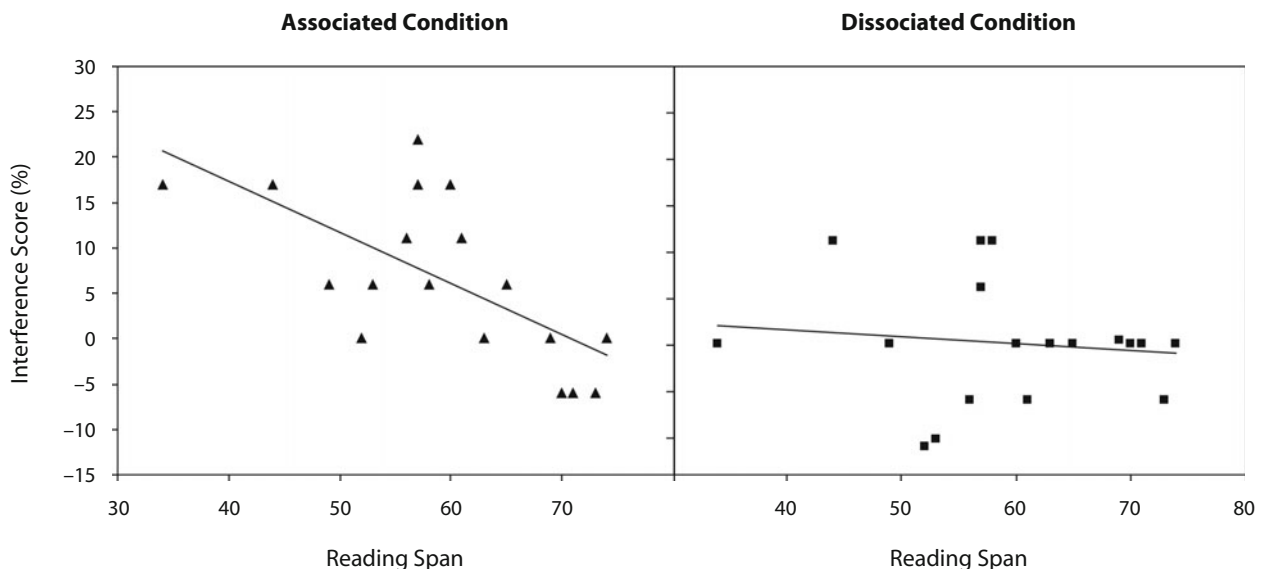

Figure 3. Relation between working memory and interference score in the associated (left) and dissociated (right) conditions in Experiment 2. A higher interference score means more interference-related performance loss.

when those objects were associated with the protagonist. Foregrounding, the process of highlighting important story elements, can act on the associative content of the story to the potential disadvantage of the problem solver.

We also looked at how the impact of foregrounded, interfering material might vary as a function of WM. Because individual differences in WM are related to the ability to retrieve content under conditions of interference (Conway & Engle, 1994), math story performance should be sensitive to WM differences when interference is at its strongest. In Experiment 2, we found that, when interfering numerical content was associated with the protagonist, the lower the participant's WM, the more errors on the task.

One might wonder why we chose the RSPAN as our WM measure. We did this because (1) the RSPAN has been shown to be a reliable predictor of a domain-general WM construct (Conway et al., 2005) and (2) we did not want to involve any math in our WM measure. Using a WM measure such as the operation span (OSPAN), which is similar to the RSPAN, except that participants solve math equations instead of making sentence judgments, would leave open an alternative explanation for our WM–interference correlation—that is, that decreased interference with increasing WM could have been due to OSPAN's relation to math ability and was not necessarily a result of the inhibition of interference. A similar critique could be applied to the RSPAN, but it would make the opposite prediction. If reading span is related to a superior construction of the situation model (although previous research suggests that it is not; Radvansky & Copeland, 2001), a higher reading span would lead to increased interference effects as a result of foregrounding. This alternative explanation would not account for (and indeed might hide the full magnitude of) our observed negative relationship between WM and interference errors.

The DOA model (Siegler, 1988) suggests that retrieval efficacy is based on the association of a problem with the correct answer, relative to alternative associations. Therefore, one might expect that priming an incorrect addition answer would result in intrusions related to those alternative associations (responding "7" to 3 × 4). But, in the present data, these errors accounted for only about 5% of incorrect answers across all multiplication conditions (5.6% in Experiment 1 and 4.1% in Experiment 2). The priming of the addition association appeared to be robust enough to interfere with multiplication accuracy—producing an answer that would lead to an incorrect question on a test—but not robust enough to produce significant addition intrusions. Interestingly, however, the bulk of incorrect multiplication responses were smaller than the correct solution (Experiment 1, 79.1%; Experiment 2, 83.5%), suggesting that answers were potentially biased by the size of the primes (which were, on average, lower than the multiplication answers). Future work examining how the specific incorrect answers produced on multiplication problems prefaced by foregrounded addition solutions come about will shed further light on our mental representation of arithmetic facts and incorrect associations in math story problems.

In conclusion, we show that situation models can affect problem solving—and enhance the role of WM in performance—simply by changing the accessibility of content within a math story problem. For tests designed to capture the math that students learn in school, rather than their general cognitive abilities, the situation model elicited by a math story problem matters.

REFERENCES

ARIZONA DEPARTMENT OF EDUCATION (2006). *Arizona's Instrument to Measure Standards Sample Test*. Retrieved April 7, 2009, from www.ideal.azed.gov/system/files/Grade4SampleTestFinal.pdf.

BASSOK, M., PEDIGO, S. F., & OSKARSSON, A. T. (2008). Priming addition facts with semantic relations. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, **34**, 343-352.

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

CONWAY, A. R. A., & ENGLE, R. W. (1994). Working memory and retrieval: A resource-dependent inhibition model. *Journal of Experimental Psychology: General*, **123**, 354-373.

CONWAY, A. R. A., KANE, M. J., BUNTING, M. F., HAMBRICK, D. Z., WILHELM, O., & ENGLE, R. W. (2005). Working memory span tasks: A methodological review and user's guide. *Psychonomic Bulletin & Review*, **12**, 769-786.

COOK, A. E., GUERAUD, S., WAS, C. A., & O'BRIEN, E. J. (2007). Foregrounding effects during reading, revisited. *Discourse Processes*, **44**, 91-111.

COQUIN-VIENNOT, D., & MOREAU, S. (2007). Arithmetic problems at school: When there is an apparent contradiction between the situation model and the problem model. *British Journal of Educational Psychology*, **77**, 69-80.

DESTEFANO, D., & LEFEVRE, J.-A. (2004). The role of working memory in mental arithmetic. *European Journal of Cognitive Psychology*, **16**, 353-386.

GERRIG, R. J., & MCKOON, G. (2001). Memory processes and experiential continuity. *Psychological Science*, **12**, 81-85.

GLENBERG, A. M., MEYER, M., & LINDEM, K. (1987). Mental models contribute to foregrounding during text comprehension. *Journal of Memory & Language*, **26**, 69-83.

RADVANSKY, G. A., & COPELAND, D. E. (2001). Working memory and situation model updating. *Memory & Cognition*, **29**, 1073-1080.

SIEGLER, R. S. (1988). Strategy choice procedures and the development of multiplication skill. *Journal of Experimental Psychology: General*, **117**, 258-275.

VAN DIJK, T. A., & KINTSCH, W. (1983). *Strategies in discourse comprehension*. New York: Academic Press.

WINKELMAN, J. H., & SCHMIDT, J. (1974). Associative confusions in mental arithmetic. *Journal of Experimental Psychology*, **102**, 734-736.

ZBRODOFF, N. J., & LOGAN, G. D. (1986). On the autonomy of mental processes: A case study of arithmetic. *Journal of Experimental Psychology: General*, **115**, 118-130.

ZWAAN, R. A., & RADVANSKY, G. A. (1998). Situation models in language comprehension and memory. *Psychological Bulletin*, **123**, 162-185.