Working memory span and situation model processing

GABRIEL A. RADVANSKY and DAVID E. COPELAND
University of Notre Dame

This study looked at how comprehension and memory processing at the situation model level is related to traditional measures of working memory capacity, including the word span, reading span, operation span, and spatial span tests. Issues of particular interest were the ability to remember event descriptions, the detection and memory of functional relationships, the detection of inconsistencies, sensitivity to causal connectivity, and memory for surface form, textbase, and situation-specific content. There was little evidence that traditional measures of working memory span were directly related to processing at the situation model level. However, working memory span was related to our few textbase-level tests.

One consistent claim from research on working memory is that people who have a larger working memory capacity are better at language comprehension and memory (Daneman & Merikle, 1996). However, many of these studies have not looked at the influence of capacity on situation model processing. Situation models (van Dijk & Kintsch, 1983; Zwaan & Radvansky, 1998), or mental models (Johnson-Laird, 1983), are complex mental representations that can simulate the situation described by a text. The information we receive as descriptions conveys important cues about the structure of a situation, how it is changing over time, and perhaps that it may be inconsistent with what has been stated before. Because model processing entails the active manipulation of information, working memory is intimately involved. Our aim was to assess the degree to which processing information at the situation model level is affected by individual differences in working memory.

When people read for understanding, they create three types of representations: the surface form, the textbase, and the situation model (van Dijk & Kintsch, 1983). The surface form is a representation of the exact words and syntax. This is a verbatim memory. Knowledge of this sort is short-lived (Sachs, 1967). At a deeper level is the textbase. This is a more abstract representation of the basic idea units, or propositions, in a text. Unlike the surface form, the textbase is more durable and can be more easily retrieved from memory. Finally, further removed is the situation model. This is a representation of the situation described by the text, not

AMERICAN JOURNAL OF PSYCHOLOGY
Summer 2004, Vol. 117, No. 2, pp. 191–213

the text itself (Glenberg, Meyer, & Lindem, 1987). The focus here is on several aspects of situation model processing, including the processing and remembering of functional relationships, the detection of inconsistencies, and general memory for information about a described event. We assessed individual differences using traditional measures of working memory capacity.

Working memory capacity

Working memory processing often is operationalized in terms of capacity or "span," that is, the number of items that can be held over a period of time with the additional load of a second task, such as reading or solving math problems. A number of studies have related working memory span to language comprehension and memory. The basic idea is that people with greater working memory capacity are better at a number of comprehension and memory tasks. There is a great deal of support for this idea. Working memory span is related to anaphoric resolution (Light & Anderson, 1985), lexical ambiguity resolution (Jackson & McClelland, 1979), syntactic complexity resolution (King & Just, 1991), and recall ability (Daneman & Carpenter, 1980). Working memory span is also related to standardized test performance, such as the verbal Scholastic Achievement Test (SAT; Daneman & Hannon, 2001).

What working memory spans actually measure is a matter of some debate (see Miyake, 2001). The traditional view is that span scores reflect how much information a person can maintain during processing. This assumes some sort of resource metaphor (Baddeley, 1986). Others have suggested that span scores reflect attentional control mechanisms, such as inhibition (Conway & Engle, 1994; Kane, Bleckley, Conway, & Engle, 2001) or the management of interference (Lustig, May, & Hasher, 2001; May, Kane, & Hasher, 1999). Regardless of the view taken, the focus in working memory span measures is on the retention of certain elements during active processing.

A number of working memory span tests have been developed. The most basic are simple span tasks, such as digit span or word span. In these tasks, participants are given sets of items (i.e., digits or words, respectively) and are tested to determine the largest set size they can recall. Daneman and Merikle (1996) showed that the word span is superior to the digit span as a measure of language processing. Because our concern is with language comprehension and memory, we focus on the word span.

More recently, the focus has been on measures that include a processing task in addition to the need to hold a set of items in memory. Tasks with both processing and storage components are called complex span tasks. We consider three such tasks here. One of the more popular is the Daneman and Carpenter (1980) reading span test. This is regarded as a

measure of working memory span for language processing. In this task, participants read aloud a set of sentences, such as, "His head sat on his shoulders like a pear on a dish." After each set, participants report back the last word of each sentence in the most recent set. Performance on this measure is correlated with a number of measures of language processing and memory (see Daneman & Merikle, 1996, for a review).

Another measure is the operation span test, developed by Turner and Engle (1989). This task has been promoted as being more domain independent. For this measure, participants are given a two-operation math problem (e.g., $[7 \times 1] + 8 = 16$) along with a word (e.g., *horse*). The problem and solution are read aloud, and the participant indicates whether the solution is correct. Then the word is read aloud. At the end of each set, a participant recalls as many of the words from that set as he or she can. Performance on this measure has been related to language processing (Engle, Cantor, & Carullo, 1992) and memory retrieval (Cantor & Engle, 1993).

More recently, Shah and Miyake (1996) developed a spatial span test that is aimed more at the processing and retention of spatial information. The structure of this test is modeled after the reading span test. In this test participants are presented with a series of rotated letters with the task of indicating whether each is normal or mirror reversed. After each set, the participant reports the location of the top of each letter (among eight equally spaced locations laid out in a circle) in the most recent set. Performance on this test has been related to spatial language processing (Friedman & Miyake, 2000).

There is no question that working memory span is related to comprehension and memory. However, almost all of this research has focused on the surface form or textbase level. Studies looking at memory for specific words, lexical access, and vocabulary ability are aimed at the surface form. In addition, studies looking at memory for propositional content, paraphrasing ability, and ability to follow directions are aimed at the textbase level. Studies that use standardized tests often emphasize the surface form, textbase, or general world knowledge. Although some portions of a verbal SAT score may reflect situation model processing, it is difficult to say which portion of a verbal SAT score is related to span. With the notable exception of a study by Friedman and Miyake (2000), which is detailed in the general discussion, very few studies have looked at working memory capacity and its relationship to processing that involves primarily the situation model level.

One of these studies was done by Lee-Sammons and Whitney (1991), who compared memory span (as measured by the operation span test) with performance on a memory paradigm developed by Anderson and Pichert (1978). This paradigm tests how reader perspective affects text

memory. Specifically, participants read a text that gave a description of a house. Half of the participants were asked to read from the perspective of a home buyer and half from the perspective of a burglar. After reading, participants were given two recall tests. The first simply asked them to recall the text. Importantly, on the second recall, they were asked to adopt either the original or an alternative perspective. Recalls were scored based on their propositional content. Span had no influence on memory when there was no shift in perspective. However, when there was a shift, the greater a participant's span score, the more propositions were recalled.

This study is interesting because perspective is a quality more of the situation model than of the textbase. One interpretation is that people with greater spans were better able to switch perspectives and recover knowledge that is more pertinent to building a new situation model. If so, this would demonstrate that working memory capacity is related to situation model processing. However, it is not clear that a perspective shift necessarily results in people creating a new situation model. It may only discredit the original model. If so, people would be less likely to use the situation model during recall and would rely more on the textbase. Under these circumstances these results would reflect differences in memory for the textbase, not the situation model.

Other suggestive studies have looked at inference verification. Work by Dixon, LeFevre, and Twilley (1988) and Masson and Miller (1983) compared performance on a number of language processing and working memory tasks. The language tasks of most interest here had participants verify inferences that entailed information from the text along with general knowledge (Dixon et al., 1988) or the combination of text elements (Masson & Miller, 1983). Performance in these studies was related to reading span scores. However, it is important to note that these reading span tests departed from the standard Daneman and Carpenter (1980) test. Rather than simply reading the sentences, participants also indicated whether the sentences made sense (Dixon et al., 1988) or performed a cloze task for fragmented sentences (Masson & Miller, 1983). These tests entail more complex processing that might involve situation models, whereas simply reading aloud may not. Therefore, it is unclear how well these tasks measure working memory capacity as it has been traditionally been operationalized.

It is also important to note that some evidence suggests that memory span may not be strongly related to situation model processing. In a study by Singer, Andrusiak, Reisdorf, and Black (1992), participants were told information about various ordered relationships, such as "A JAL is larger than a TOC," "A beaver is larger than a CAZ," and "A TOC is larger than a pony." Afterward, participants were presented with probes, such

as "A TOC is larger than a beaver." This may involve situation models because participants had to make inferences about the various relationships. Importantly, however, in this study no relationship between span and performance was observed.

Situation model use

This section considers a number of situation model processing measures. This includes memory for general event descriptions, the comprehension and memory of functional aspects of a text, the detection of situational inconsistencies, the influence of causal connectivity on reading, and memory for situation-specific information.

Situation identification. As stated earlier, a situation model is a representation of the situation described by a text, not the text itself. When using situation models to make memory decisions, people are able to identify information that is consistent with a previously described situation even if that explicit information had never been encountered before. Sometimes, using situation models to guide memory can cause people to select information that is situationally consistent with a previous description instead of selecting the actual description (Garnham, 1981; Radvansky, Gerard, Zacks, & Hasher, 1990). For our purposes, we wanted a method that required participants to use situation models to identify statements that are consistent with previous descriptions but have not been read before. To this end, we used a situation identification test in which participants first read a series of sentences. Afterward, they were presented with a test in which they were to select one of six options that best described the same situation as the original sentence. The ability to do this was used as a measure of situation model memory use.

If working memory span is related to situation model processing, participants with higher span scores should have higher situation identification scores. Having greater capacity would allow them to encode and store information in long-term memory more effectively. This would be reflected in a greater ability to remember what was read earlier and determine which of the alternatives best matched the original.

Functionality. An important role of the situation model is to represent the functional relationships between entities. Functional relationships are those that involve the meaningful interaction between entities and define the situation. These relationships often convey a typical interaction between two entities, such as between a hammer and a nail. However, it may be possible for a functional relationship to be atypical, such as using a rock to pound a nail. These atypical but functional relationships are identified only when it is possible for the entities to meaningfully interact, typically based on the affordances of the entities to one another (Glenberg, 1997). Functional relationships are important for situation model

processing. People find it easier to encode and remember functional as compared with nonfunctional spatial relationships (Radvansky & Copeland, 2000). In addition, people are able to identify pictures of objects more quickly when the orientation of the pictured object matched how an object was described as functioning in a particular context (Stanfield & Zwaan, 2001).

If working memory capacity is related to situation model processing, a reasonable expectation is that people with higher span scores will show larger functionality effects. Their greater capacity would allow them to access and use more information from long-term memory. With more information available about the referents in the text, it would be easier to note how the elements in a situation are interacting. Thus, high-span people would be in a better position to take advantage of functional relationships.

Inconsistencies. If a person comprehends a text sufficiently, he or she will notice any inconsistencies. The term *inconsistency* here refers to information that describes conditions in the world that cannot coexist. For example, a person cannot be in Georgia and Indiana at the same time. People are sensitive to situational inconsistencies. In one study, O'Brien and Albrecht (1992) had participants read texts in which the location of a character was described. For example, a sentence might be, "As Kim stood *inside/outside* the health club she felt a little sluggish," where Kim's location is varied between conditions. A later sentence in the text would be, "She decided to go outside and stretch her legs a little." If Kim was initially described as being inside, this second sentence would be consistent with the previously described situation. However, if she was already outside the health club, this second sentence would be inconsistent with the described events. Reading times for critical sentences are slower in the inconsistent than the consistent condition. People can be sensitive to the internal consistency of the described situation.

Working memory span may be involved in this process, with participants with larger capacities showing greater inconsistency effects. The detection of inconsistencies requires a person to have available both the current information in the text and the information from the earlier portion of the text that makes it inconsistent. The greater a person's working memory span, the more likely this information is to be available.

Causal connectivity. One characteristic of described situations that is not explicitly conveyed in the structure of a text itself is the causal relationship. There is some variability in the degree to which various elements are causally related to one another. This is called causal connectivity. The more causal connections there are, the greater the causal connectivity. Causal connectivity has an important influence on comprehension and memory. Specifically, information that is integrated into the causal chain

of events and is higher in causal connectivity is better remembered (Trabasso & van den Broek, 1985) and is rated as being more important (Trabasso & Sperry, 1985). The current study uses an approach in which the elements of a text, such as clauses or sentences, are first coded in terms of the degree of causal connectivity. In our case, we look only at prior causal antecedents to the current text unit because participants cannot be expected to know what will follow in the text they are reading. Then reading times for these elements are analyzed in a regression analysis, with the number of causal connections as one of the predictor variables.

If working memory span is involved in this process, people with larger capacities will show a greater influence of causal connectivity. Influence by causal connectivity requires that a person have available the previous portions of the text that are causally relevant to the information that is being read. The greater a person's working memory span, the more likely this information is to be available.

Memory for situation-specific information. As described earlier, people typically create three levels of representation: the surface form, the textbase, and the situation model. People may vary in the degree to which they emphasize these different representations. Therefore, it makes sense to separate out these components using a method developed by Schmalhofer and Glavanov (1986) that has been successfully used by others (Fletcher & Chrysler, 1990; Kintsch, Welsch, Schmalhofer, & Zimny, 1990; Radvansky, Zwaan, Curiel, & Copeland, 2001; Zwaan, 1994). In this paradigm, participants read a text. They are then given a series of probe sentences. The task is to indicate whether a sentence was read before. There are four classes of probes: verbatims, paraphrases, inferences, and incorrects. Using signal detection analysis, the proportion of "yes" responses to each of these probe types can be used to determine the strength of the three representational levels.

EXPERIMENT

At this point it is fairly clear that traditional measures of working memory span are related to comprehension and memory at the surface form and textbase levels. However, the relationship to the situation model level is not clear. It may be that working memory span is a general-purpose cognitive resource. If so, one would expect it to be related to situation model processing as well. The greater one's capacity, the better one is able to construct, and therefore remember, a coherent situation model of the described events. Alternatively, it may be that traditional measures of working memory span are more confined to surface form and textbase levels of comprehension and memory and do not have a strong, direct

relationship to the situation model level. The current experiment tested these ideas, with a focus on how working memory capacity relates to situation model–level processing.

METHOD

Participants

We tested 160 participants drawn from participant pools at the University of Notre Dame and Indiana University at South Bend who were given partial class credit. All were native English speakers. The data from one participant were excluded because his situation identification score was at chance (4 out of 24) and his reading times very short, suggesting that he was not actually reading the stories.

Materials and procedure

Memory span tests. There were four tests of working memory capacity. The most basic was a word span test. For this test, participants were presented with a series of words in increasing set sizes of three to eight, with three sets at each set size, and an additional three practice trials at set size three. The words were presented one at a time on a computer screen for 1 s each. At the end of each set, a series of question marks was displayed, one for each item in the set. The task was to recall the words in the order in which they were presented. Participants responded by typing their responses into the computer. They typed "DK" ("don't know") for the items they could not recall. The order of the words was randomized for each participant.

For the Daneman and Carpenter (1980) reading span test participants were presented with a series of sentences in increasing sets of two to six, with five sets at each set size, except for set size six, for which there were only three sets. The sentences were presented on a computer screen, and the task was to read each sentence aloud. After each sentence was finished, the experimenter advanced to the next sentence by pressing the spacebar. When a blank screen appeared, the task was to recall the final word from each sentence in the current set. The experimenter typed these responses into the computer. Again, participants were encouraged to recall these words in the order in which they were presented. The order of the sentences was randomized for each participant.

For the Turner and Engle (1989) operation span test participants were presented with a series of math problems followed by a word, such as "$(9 \times 1) + 1 = 9$" then "BOAT" in increasing sets of two to seven, with three sets at each set size, except for set size two, for which there were six sets (the first three were considered practice). The problems and words were presented on a computer screen, and the task was to read each problem aloud, indicate whether the answer was correct by pressing one of two buttons on a computer mouse, and then read the word aloud. When a blank screen appeared, the task was to recall the words from that set. The experimenter typed these responses into the computer. Again, participants were encouraged to recall these words in the order in which they

were presented. The order of the problems and words was randomized for each participant.

For the Shah and Miyake (1996) spatial span test participants were presented with a series of rotated letters (i.e., "F," "J," "L," "P," and "R") in increasing sets of two to six letters, with five sets at each set size, except for only three at set size six. The letters were displayed on index cards, and the processing task was to state whether the letter was normal or reversed. Responses were recorded by the experimenter. After each set, a card was presented with eight locations mapped out around a circle at 45° intervals (0°, or upright, was never used). The task was to indicate the position of the top of the letters in the current set in the order in which they were encountered. Note that the same location was never repeated within a set. The experimenter recorded these responses. Participants were encouraged to guess if they were unsure.

For the scoring of all span tests, if a set was recalled correctly, then the score was equal to the number of words in that set. The scores for all of the correctly recalled sets were then totaled (Conway & Engle, 1994).

Situation identification test. In this task, participants were given a series of 24 sentences, which were drawn from or inspired by the materials used by Garnham (1981). These sentences were presented one at a time on a computer screen. During this initial presentation, the task was to rate the sentences for pleasantness. Responses were entered into the computer using a scale ranging from 1 (*extremely unpleasant*) to 7 (*extremely pleasant*). Because this was a cover task, these ratings were not recorded.

After all the sentences were rated, participants were given a surprise identification task. On each trial, six alternatives were presented, which were variations of the original sentence. These six alternatives were presented in a random order on each trial for each participant. The task was to select the item that most closely described the same situation as the sentence read and rated earlier. Participants were informed that the original sentences would never appear and to type their responses into the computer. No feedback was provided. The six alternatives were altered prepositional phrase, altered direct object, altered verb, altered prepositional phrase and direct object, altered prepositional phrase and verb, and altered direct object and verb. An example of six sentences for the original sentence, "The man lost a hand of poker at the card shark's," are presented here. Sentence 2 is the correct response.

1. The man lost a hand of poker like the card shark.
2. The man lost some money at the card shark's.
3. The man won a hand of poker at the card shark's.
4. The man lost some money like the card shark.
5. The man won a hand of poker like the card shark.
6. The man won some money at the card shark's.

Reading and memory task. For this study participants were asked to read a series of eight narratives. These stories were 31 to 45 sentences long ($M = 38$). An example of one of these texts is presented in Appendix A. These stories were presented one clause at a time on an IBM-compatible computer in white on a

black background. Participants pressed the spacebar with their left hand to advance to the next clause. Reading times were collected. They rested their right hand on the computer mouse to respond to the comprehension questions at the end of each story. The left button was marked with a "Y" for "yes," and the right button was marked with an "N" for "no." A number of measures were gathered during story reading. A few memory measures were presented using paper and pencil after all stories were read. We discuss each of these measures in turn.

Functionality. To assess the degree to which participants were sensitive to functional aspects of the described situation, each story contained two sentences that described a spatial relationship. For two of the stories, both relationships were functional. For two stories, both were nonfunctional. Finally, for four stories one was functional and the other nonfunctional. The combinations of conditions to stories were rotated across participants. To illustrate the difference between the functional and nonfunctional sentence versions, in the story presented in Appendix A, the functional version was "David was standing below an old bridge." This is functional because it allows David to get out of the rain. The nonfunctional version was "David was standing next to a lamppost." This is nonfunctional because this spatial relationship plays no role in the story. In all cases, we changed the functional version to a nonfunctional version by altering both the spatial relationship and the reference object. To assess whether functionality had an effect on comprehension, we used the reading time for these sentences. Previous research has shown that functional sentences are read more quickly than nonfunctional sentences (Radvansky & Copeland, 2000). Presumably, people are influenced by how well the information fits into the causal chain of the described events.

In addition, after participants had read all the stories, they were given a recognition test for this functional information. They were presented with each original sentence along with three distractors. The distractor sentences altered the located object in the sentence, the spatial relationship between the two, or both. For example, if the original sentence was "David was standing below an old bridge," the distractor sentences would be "David was standing next to an old bridge," "David was standing below a lamppost," and "David was standing next to a lamppost." Thus, the functional and nonfunctional versions were always present. These options were randomized, and the order in which the various trials occurred was also randomized. Participants responded by circling the letter next to each sentence (*a, b, c,* or *d*) that they thought corresponded to the sentence they had read earlier.

Inconsistency. To assess the ability to detect inconsistencies, embedded in the text were three sentences that described actions that were either consistent or inconsistent with the protagonist's current location. Each story had either one consistent and two inconsistent sentences or two consistent and one inconsistent sentences. The versions were rotated in each story across participants. To illustrate the difference between consistent and inconsistent sentence versions, in the story presented in Appendix A there is a statement that David is walking along the banks of a river. Several sentences later, a sentence occurs that is either consistent ("David walked further down the river") or inconsistent ("David walked outside to the river") with the previously described location. Reading time for

this target sentence was recorded and used as the dependent measure. Across all stories there were 24 consistency sentences.

Causal connectivity. To assess the influence on causal connectivity, we first coded each of the experimental stories' causal structure. Specifically, for each clause that was read, we assessed whether it was causally related to subsequent clauses. We scored two story clauses as being causally connected if they met the following criteria: The first event had started before the second event, the first event was still in operation when the second event started, the occurrence of the first event was necessary for the second event, and the first event was sufficient for the second event, either by itself or in conjunction with other narrative events. As an example from the story in Appendix A, the sentence "The nights could be bitterly cold" causes "He hugged his thin, dust-covered coat around him." In addition to this standard scoring procedure, we also scored two narrative events as being causally related, but in a backward manner, if the second event preceded the first. As an example from the story in Appendix A, the sentence "The townsfolk treated you with contempt . . ." is caused by ". . . if you didn't speak the native Arabic or French."

After the narratives were scored, we used causal connectivity as a predictor variable in a regression analysis, with reading times for the various clauses as the dependent variable. The basic idea was that if causal connectivity is important to comprehension, then the more causal connections there are, the easier it should be to read the clause and the faster the reading time. In addition to causal connectivity, we included a number of text variables in the regression analyses that are known to affect reading times: number of syllables, word frequency, serial position, and new arguments. Reading times for the inconsistency and functionality items were excluded form this analysis because their relationship to the causal structure of the text varied depending on which version a participant read.

Comprehension questions. Although they were not aimed at a particular level of processing, after each story a pair of general comprehension questions were presented, such as, "Did the travel agent lie to David?" and, "Did David like the coffee at the hotel?" Their purpose was to encourage participants to actively read the texts. These comprehension questions required "yes" or "no" responses that were recorded by having participants press one of two buttons on the computer mouse.

Levels of representation. Using the Schmalhofer and Glavanov (1986) paradigm, after they read a text, we gave participants a recognition test. Sixteen sentences were selected from each experimental text. There were four types of probes. *Verbatim* probes were sentences that had appeared in the text (e.g., "The driver filled up the gas tank"). *Paraphrase* probes contain the same propositions, expressed differently. Sentences were reworded so that they retained the propositional content of the original. This was done using synonyms or altering word order (e.g., "The gas tank was being filled by the driver"). *Inference* probes were information that was not mentioned but was consistent with and important to the described situation (e.g., "The gas tank of the car was near empty"). Finally, *incorrect* probes were composed of information that was not mentioned and was unlikely to be inferred. However, the information was globally consistent with the passage's theme (e.g., "The driver washed the windows of the car").

The task was to indicate whether a sentence had been read earlier. Participants were warned that the sentences might contain slight wording changes. The probes were blocked by story, with the title appearing before the probes. This was done so that participants knew to which passage the items referred. The story order was the same as during reading. The order of the probes within a story block was randomized. Participants responded by circling either a "Y" or an "N" located next to the probe sentence to indicate whether it was an old or new sentence. The type of probe (i.e., verbatim, paraphrase, inference, or wrong) for a particular sentence was rotated across participants.

The ability to discriminate verbatim from paraphrase probes is an index of the surface representation. The difference between these is that one matches the original surface characteristics, and the other does not. They are equivalent in how they map onto the textbase and situation model. Similarly, the ability to discriminate between paraphrase and inference probes is an index of the textbase. The difference between these is that one matches the original in propositional content, and the other does not. Also, both are inconsistent with the surface structure and consistent with the described situation. Finally, the ability to discriminate between inference and incorrect probes provides an index of the use of situation models. The difference between these is that one corresponds to the described situation, and the other does not. Both are inconsistent with the surface structure and textbase.

A' scores (Donaldson, 1992), a nonparametric signal detection measure, were calculated as a discrimination measure for each of the aforementioned comparisons. For the surface form measure, verbatims were considered hits, and paraphrases were considered false alarms. For the textbase measure, paraphrases were considered hits, and inferences were considered false alarms. Finally, for the situation model measure, inferences were considered hits, and incorrects were considered false alarms.

Experimental context. These measures were embedded in the context of a larger study. In addition to the tasks described earlier, we included measures of processing speed, vocabulary, spatial updating, logical reasoning, and a fan effect task. The results of these tasks are considered elsewhere (Copeland & Radvansky, in press; Radvansky & Copeland, 2001). A certain order of tasks was set. This order was (a) vocabulary, (b) speed, (c) situation identification, (d) word span, (e) reading span, (f) updating task, (g) spatial span, (h) long text comprehension and recognition, (i) operation span, (j) logical reasoning, and (k) fan effect. Testing was done across two 1½-hour sessions or three 1-hour sessions. If done over 2 days, Tasks a–h were done on the first day and Tasks i–k were done on the second day. However, if testing was done over 3 days, tasks a–f were done on the first day, tasks g–i were done on the second day, and tasks j–k on the third day.

Data treatment. All of the reading time data, including the functionality and inconsistency items, were trimmed by first eliminating any clearly deviant times (less than 50 ms per syllable or more than 1,500 ms per syllable). In addition, for sentences in the functional and inconsistency analyses, the fastest and slowest reading times per condition per participant were dropped (cf. Rinck & Bower, 1995). Finally, we calculated Cronbach's alphas for each of our indices.

RESULTS

As an overview, for all of our measures of situation model processing, the expected effects were observed. However, there was no clear evidence that performance on the working memory span tests was related to comprehension and memory at the situation model level. However, consistent with the majority of the published research, working memory span seemed to be more related to performance on our few textbase-level measures.

Working memory span

Summary data for these tests in our experiment are presented in Table 1; the correlations between the various span measures are presented in Table 2. There was a similar level of reliability among the various span scores: word span Cronbach's $\alpha = .71$, reading span Cronbach's $\alpha = .74$, operation span Cronbach's $\alpha = .75$, and spatial span Cronbach's $\alpha = .82$. As can be seen, the span tests were moderately correlated with one another. Therefore, we elected to compute a composite span score. To do this we calculated z scores for each of the individual span measures and then took the average of these z scores. The mean and standard deviation of this composite score are listed in Table 1. For the remainder of the results, we discuss only the relationship of our memory and comprehension variables to the composite score. Relevant means and analyses for the individual span scores are presented in Appendix B.

Situation identification

The situation identification test scores ranged from 7 to 23 (out of 24 possible), with a mean of 17.7 ($SD = 3.3$). Thus, participants could make judgments about the described situation fairly well based on their memory of what they had read earlier. The rate of correctly identifying previously described situations was used as an index of performance, Cronbach's $\alpha = .70$. Performance on this test was not correlated with a participant's composite working memory score, $r = .12$, $p = .13$.

Table 1. Summary of mental ability test scores

	Mean	*SD*	Minimum	Maximum	Scoring
Word span	35.2	13.1	9	77	Count score
Reading span	22.5	12.6	4	69	Count score
Operation span	21.6	11.1	0	54	Count score
Spatial span	29.0	19.2	0	84	Count score
Composite	0.0	0.8	−1.6	2.9	Mean z score

Table 2. Correlations between the span tests

	Word	Reading	Operation	Spatial
Word span	—			
Reading span	.59	—		
Operation span	.49	.61	—	
Spatial span	.40	.55	.39	—

Note. All correlations were significant, $p < .001$, with a Bonferroni correction.

Functionality

Reading time data were converted to milliseconds per syllable to account for differences in length of sentences. None of the reading time data were trimmed for exceeding the long criterion, but 1.7% of the data were trimmed for being too fast (less than 50 ms per syllable). The fastest and slowest reading times in each condition were trimmed as described in the *Method* section. Overall, people read the spatial relationship sentences faster when they conveyed a functional relationship (181 ms per syllable, $SD = 48$) than when they conveyed a nonfunctional relationship (198 ms per syllable, $SD = 54$), $F(1, 158) = 33.86$, $MSE = 658$, $p < .001$. This is consistent with previous research (Radvansky & Copeland, 2000). The difference between the reading times in the two conditions (16 ms per syllable, $SD = 36$) was used as an index of the functionality effect to compare with the memory span tests, Cronbach's $\alpha = .93$. There was no significant relationship between working memory span and this measure, $r = -.08$, $p = .35$. If anything, the correlation is nominally in the wrong direction.

For the recognition data, people identified the sentences better when they conveyed a functional relationship (82%, $SD = 15$) than a nonfunctional relationship (57%, $SD = 22$), $F(1, 158) = 214.17$, $MSE = .024$, $p < .001$. The difference between the recognition rates in the two conditions (25 ms per syllable, $SD = 22$) was used as an index of the functionality effect to compare with the memory span tests, Cronbach's $\alpha = .30$. Although there was a significant relationship with working memory span, $r = -.22$, $p = .007$, the direction of this relationship is the opposite of what was predicted. Specifically, the greater a participant's reading span score, the smaller the functionality effect.

Closer inspection of the recognition data revealed an interesting relationship between the memory span measures and performance on the functional and nonfunctional items. Specifically, working memory span was related to performance on the nonfunctional items, $r = .32$, $p < .001$, but not to the functional items, $r = .14$, $p = .09$. This is consistent with the idea that span tests tap more into propositional textbase-level memories.

This is because the nonfunctional items are not well integrated into the situation model, and so memory for this information is more likely to reflect memory for the textbase.

Inconsistencies

Again, reading time data were converted to milliseconds per syllable to account for differences in length of sentences. Overall, participants read consistent sentences faster (195 ms per syllable, $SD = 50$) than inconsistent ones (220 ms per syllable, $SD = 64$), $F(1, 158) = 67.09$, $MSE = 747$, $p < .001$. Thus, comprehension was disrupted by information that was inconsistent with the current state of affairs. This is consistent with previous research (e.g., O'Brien & Albrecht, 1992). The difference between the reading times in the two conditions (25 ms per syllable, $SD = 40$) was used as an index of the inconsistency effect to compare with the memory span tests, Cronbach's $\alpha = .91$. Again, there was no significant relationship with working memory span with this measure, $r = -.08$, $p = .33$. If anything, the relationship is in the wrong direction.

Causal connectivity

The degree of causal connectivity was significantly related to reading time, with a mean beta-weight of $-.049$ ($SD = .039$), $t(159) = -15.79$. Thus, participants were sensitive to the causal structure of the texts, as predicted by situation model theory. The causal beta-weight from the regression analyses was used as an index of sensitivity to causal connectivity to compare with the memory span tests, Cronbach's $\alpha = .33$. Again, there was no significant relationship with the composite working memory span measure, $r = -.08$, $p = .30$, with the correlation nominally in the wrong direction.

Comprehension questions

Overall, performance on the comprehension questions was good, at 93% ($SD = 7$) correct. The accuracy rate was used as an index of comprehension understanding, Cronbach's $\alpha = .38$. Despite the low reliability of this measure, working memory span was significantly related to performance, $r = .19$, $p = .017$. It should be noted that these questions asked about detailed information that would not necessarily have been retained in the situation model alone. The surface form and textbase levels of representation could have been used as well. Therefore, some influence of propositional textbase memory would be expected.

Schmalhofer and Glavanov analysis

The recognition test data revealed higher A' discrimination scores at the situation model level (mean $A' = .79$, $SD = .09$), less so at the textbase

level (A' =.68, SD = .13), and lowest at the surface form level (A' = .58, SD = .10). All of these values are significantly different from one another, all $ps < .001$, and all were significantly greater than chance, all $ps < .001$. Thus, in this case, memory for the texts was dominated by the situation model level. Not surprisingly, there was no reliability in the surface level measure, Cronbach's $\alpha = .03$, given that people are performing near chance anyway. Reliability was better for the textbase, Cronbach's $\alpha = .39$, and situation model measures, Cronbach's $\alpha = .56$.

Not surprisingly because they were so close to chance (.5), surface form A's were unrelated to the composite span measure, $r = .01$, $p = .95$. However, for the textbase A's, performance was significantly related to the composite span measure, $r = .23$, $p = .003$. Finally, the situation model A' was not significantly related to the composite span measure, $r = -.10$, $p = .21$, and was in the wrong direction.

DISCUSSION

This study looked at performance on a variety of situation model comprehension and memory processes and their relationship to standard working memory span measures. However, no strong pattern emerged. On the whole, there was very little observed relationship between working memory span and performance on situation model–level tasks.

As suggested earlier, memory span tests may be more sensitive to processing at the textbase level than at the situation model level. Performance on the textbase A' measure and general comprehension questions clearly involves this level of representation. Also, memory for functional relationships could be viewed as involving a textbase component because this task involves recognition of what was actually in the text. This idea is further reinforced by the finding that span is related only to the nonfunctional items, which are generally remembered more poorly overall. This nonfunctional information would not be as well integrated into the situation model, so performance on these items would be more a reflection of textbase memory.

The idea that working memory span tests are more related to textbase-level processing is reinforced by the large number of prior studies that found significant relationships between span and performance on particular tasks. For example, studies of anaphor resolution (e.g., Light & Anderson, 1985) involve resolving the reference of a text anaphor to an earlier text element, a process that focuses on the textbase level. Tasks that involve resolving syntactic complexities (e.g., King & Just, 1991) are focused on textbase processing in that participants must untangle the various components of a sentence. Finally, many tests of long-term memory that have been correlated with span, such as recall ability (e.g., Dane-

man & Carpenter, 1980), have focused on scoring individual propositions, which are the components of the textbase, not necessarily the described situation.

We will consider each of the memory span tests in turn. For the word span test, although it was rarely better than the other measures, its relationship to performance was generally consistent with the more complex span measures. The reading span test has been argued to be a measure of general language processing ability (Daneman & Carpenter, 1980). Therefore, it would be expected that this measure might fare better than most of the others. However, this was not what was observed. It was significantly correlated to only a single performance measure, the recognition of spatial descriptions, and then only to the nonfunctional ones, like all of the other span tests. The operation span test has been promoted as a measure of more general-purpose cognition (Kane et al., 2001). As such, it might be a better measure of general cognitive ability. Consistent with this idea, this measure was related to two performance measures, comprehension question accuracy and textbase A', whereas the other span tests were related to only one measure at most. If working memory span tests are best viewed as measures of textbase processing, the operation span score does this task better than the others.

Finally, the spatial span test has been put forward as tapping into spatial processing abilities (Shah & Miyake, 1996). Moreover, Friedman and Miyake (2000) suggested that this measure also picks up on the processing of spatial information during language comprehension. This is important in the context of the current study because many of our situation model measures involved spatial information. The functionality measures were assessing spatial functionality. The inconsistency measure tapped into spatial inconsistencies. However, in all cases, spatial span did no better than the other span tests. The only time spatial span distinguished itself from the others was on the situation model A' measure. However, even here the relationship is in the opposite direction of what would be expected if spatial working memory were important. The pattern of results is consistent with the idea that even this span test reflects lower-level processing.

Although it appears that traditional memory span measures are more in tune with processing at the textbase level than the situation model level, this is not to say that span is unimportant for language comprehension. Obviously it is. All we are saying is that these indices appear to be measuring processing at low or intermediate levels. Furthermore, we are not trying to argue that memory span has no implications for situation model processing. If processing is sufficiently disrupted at lower levels, this will complicate the ability to create coherent and accurate situation models that are built from this information. Thus, memory span may have

indirect influences on situation model processing that were not observed here.

This idea that memory span is related more to the textbase level than the situation model level is also in line with some aging research. It is well known that older adults score lower on working memory span measures than younger adults (e.g., Radvansky et al., 1990). However, across a number of tasks there has been a noticeable absence of age-related differences in the ability to process information at the situation model level (Radvansky, 1999). For example, in a study directly addressing the issue of levels of processing and aging, Radvansky et al. (2001; Radvansky, Copeland, & Zwaan, 2003) found that although older adults performed more poorly than younger adults on a textbase memory measure, they did as well as, if not better than, the younger adults on a situation model measure. To reinforce our conclusions about the relationship between working memory span and textbase- but not situation model-level processing, we looked at the correlation between these measures. In Radvansky et al. (2003) the correlations between span and surface, textbase, and situation model levels were $r = .08$, $r = .29$, and $r = -.04$, respectively, for Experiment 1 and $r = .10$, $r = .22$, and $r = -.06$ for Experiment 2.

Situation model dimensions and working memory

In a pair of experiments by Friedman and Miyake (2000), performance on reading span and spatial span tests was compared with situation model processing during reading. That study focused on the dimensions of causality and space. Participants read texts that described characters moving about in buildings. During reading, they were interrupted with probes to assess causal or spatial information. The causal probes were inferences that were verified or rejected. The spatial probes were maps of the building with one of the rooms highlighted. The task was to indicate whether that room was the protagonist's current location.

Friedman and Miyake (2000) interpreted their results as showing that the spatial and causal dimensions were processed using different components of working memory, with the reading span being more related to the causal dimension and the spatial span being more related to the spatial dimension. Performance seemed to vary as a function of the amount of capacity in the verbal and spatial portions of working memory for each of these dimensions, respectively.

However, a closer examination of the data calls this interpretation into question. For the causal probes, reading span was correlated with accuracy but not response times in the first experiment and with neither in the second experiment. Moreover, although causal processing was unrelated to spatial span in the first experiment, it was related to spatial accuracy in the second experiment unless two outliers were removed. Thus,

the relationship between reading span and causal processing was weak. Furthermore, it is unclear to what degree this causal task depended on the retrieval of textbase information. It is possible that the significant relationship between reading span and causal probe accuracy in the one experiment tapped into the textbase level, not the situation model level.

For the spatial probes, in both experiments spatial span was not related to accuracy but was related to response time. Moreover, in the first experiment, spatial probe accuracy was related to reading span unless one outlier was removed. In the second experiment, spatial probe response time was related to reading span unless two outliers were removed. The evidence for the relationship between spatial span and situation model processing is better here, but it is confined to the response time data. Moreover, the extent to which this task taps situation model processing is unclear. Probe presentation disrupts comprehension and requires participants to do an explicitly spatial task involving a map. It may be that this task leads participants to coordinate information in the situation model with a more context-independent mental map. Thus the spatial span–response time relationship may reflect mental map, not situation model, processing. Given the absence of a relationship between spatial span and the situation model measures in the current study, it is plausible that Friedman and Miyake's (2000) results were due more to the nature of the probe task than to situation model processing per se.

CONCLUSION

The current study showed that working memory capacity, as traditionally operationalized, is a good predictor of success of some of the cognitive processes involved in language comprehension and memory. However, it does not have the broad scope that some researchers have ascribed to it. Indeed, it seems to be confined to lower levels of processing, such as the textbase. Many aspects of comprehension and memory are unrelated to memory span, including many processes involved in cognition at the situation model level.

Appendix A. Experimental text example

David walked along the banks of the river in town. Although the days were unbearably hot, the nights could be bitterly cold. He hugged his thin, dust-covered coat around him as he thought about how the travel agent had lied to him. Two months ago, David went to see the travel agent. She told him that North African Sahara towns were friendly and romantic. Now, everything so far suggested the opposite.

David walked further down the river. (consistent)

David walked outside to the river. (inconsistent)

A steady cold rain began to pour from the sky. If he stayed out much longer he would get soaked.

David was standing below an old bridge. (functional)

David was standing next to a lamppost. (nonfunctional)

He listened to the rain falling on the road as he took stock of their misfortunes so far. The townsfolk treated you with contempt if you didn't speak the native Arabic or French. His wallet and passport had been stolen. Maureen and he were shocked to find out how decrepit and dirty their hotel was. Even the coffee they were served was bad. David was sure that this trip would bring his troubled marriage to an end. Twenty minutes later, David saw a taxi and hailed it.

The driver stopped and David got out. (consistent)

The driver stopped and David got in. (inconsistent)

As he was scanning the drab city he saw an object that could free him. While driving through the merchant district, he saw an old black Ford. Although it was far from perfect, he thought that he could use it to escape this cursed place. David couldn't take his eyes off that car. The driver had just pulled into a gas station.

The old car was sitting to the left of a slick new gas pump. (functional)

The old car was sitting in front of a slick new Mercedes. (nonfunctional)

The contrast was striking. The driver filled up the gas tank.

David wished he had his wallet . . . (consistent)

David wished pulled out his wallet . . . (inconsistent)

. . . so that he could offer to buy that car. Maybe he would just steal it. How liberating it would be to cruise out of this town in that car. He didn't know where he would drive to, he just wanted out of here. Even sitting in a gas station it seemed to command his attention.

Appendix B

Relevant means for individual span scores

		Memory span tests		
		Sentence	Operation	Spatial
Situation identification	.05	.11	.07	.14
Functionality (reading times)	.04	−.03	−.12	−.11
Functionality (recognition accuracy)	−.15	−.20*	−.16	−.16
Inconsistencies (reading times)	−.08	−.05	−.08	−.06
Causal connectivity (beta-weight)	.00	−.06	−.02	−.14
Comprehension questions (accuracy)	.15	.14	.23*	.08
Surface form (A')	−.02	.02	−.09	.08
Textbase (A')	.16	.16	.23*	.19**
Situation model (A')	−.09	−.03	−.04	−.16

Note. Reading times were calculated as milliseconds per syllable.

Functionality recognition correlation analyses broken down by condition

	Memory span tests			
	Word	Sentence	Operation	Spatial
Nonfunctional items	.26*	.28*	.21*	.24*
Functional items	.14	.11	.05	.10

*$p < .05$ (Bonferroni corrected for each dependent variable). **$p \leq .10$.

Notes

We would like to thank Michael Wenger for his comments on an earlier version of this manuscript. We would also like to thank Diane Berish, Anthony Bishara, Jackie Curiel, Lacey Nielson, Kris White, and Sarah Wieber for their assistance in collecting the data. This research was supported in part by a grant from the Army Research Institute, ARMY-DASW01-99-K-0001.

Correspondence about this article should be addressed to Gabriel A. Radvansky, Department of Psychology, 107 Haggar Hall, University of Notre Dame, Notre Dame, IN 46556 (e-mail: Radvansky.1@nd.edu). Received for publication June 24, 2002; revision received April 15, 2003.

References

Anderson, R. C., & Pichert, J. W. (1978). Recall of previously unrecallable information following a shift in perspective. *Journal of Verbal Learning and Verbal Behavior, 17,* 1–12.

Baddeley, A. D. (1986). *Working memory.* Oxford: Oxford University Press.

Cantor, J., & Engle, R. W. (1993). Working-memory capacity as long-term memory activation: An individual differences approach. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 19,* 1101–1114.

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

Copeland, D. E., & Radvansky, G. A. (in press). Working memory and syllogistic reasoning. *Quarterly Journal of Experimental Psychology.*

Daneman, M., & Carpenter, P. (1980). Individual differences in working memory and reading. *Journal of Verbal Learning and Verbal Behavior, 19,* 450–466.

Daneman, M., & Hannon, B. (2001). Using working memory theory to investigate the construct validity of multiple-choice reading comprehension tests such as the SAT. *Journal of Experimental Psychology: General, 130,* 208–223.

Daneman, M., & Merikle, P. M. (1996). Working memory and language comprehension: A meta-analysis. *Psychonomic Bulletin & Review, 3,* 422–433.

Dixon, P., LeFevre, J., & Twilley, L. C. (1988). Word knowledge and working memory as predictors of reading skill. *Journal of Educational Psychology, 80,* 465–472.

Donaldson, W. (1992). Measuring recognition memory. *Journal of Experimental Psychology: General, 121,* 275–277.

Engle, R. W., Cantor, J., & Carullo, J. J. (1992). Individual differences in work-

ing memory and comprehension: A test of four hypotheses. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 18,* 972–992.

Fletcher, C. R., & Chrysler, S. T. (1990). Surface forms, textbases, and situation models: Recognition memory for three types of textual information. *Discourse Processes, 13,* 175–190.

Friedman, N. P., & Miyake, A. (2000). Differential roles for visuospatial and verbal working memory in situation model construction. *Journal of Experimental Psychology: General, 129,* 61–83.

Garnham, A. (1981). Mental models as representations of text. *Memory & Cognition, 9,* 560–565.

Glenberg, A. M. (1997). What memory is for? *Behavioral and Brain Sciences, 20,* 1–55.

Glenberg, A. M., Meyer, M., & Lindem, K. (1987). Mental models contribute to foregrounding during text comprehension. *Journal of Memory and Language, 26,* 69–83.

Jackson, M. D., & McClelland, J. L. (1979). Processing determinants of reading speed. *Journal of Experimental Psychology: General, 108,* 151–181.

Johnson-Laird, P. N. (1983). *Mental models: Towards a cognitive science of language, inference and consciousness.* Cambridge, MA: Harvard University Press.

Kane, M. J., Bleckley, M. K., Conway, A. R. A., & Engle, R. W. (2001). A controlled-attention view of working memory. *Journal of Experimental Psychology: General, 130,* 169–183.

King, J., & Just, M. A. (1991). Individual differences in syntactic processing: The role of working memory. *Journal of Memory and Language, 30,* 580–602.

Kintsch, W., Welsch, D., Schmalhofer, F., & Zimny, S. (1990). Sentence memory: A theoretical analysis. *Journal of Memory and Language, 29,* 133–159.

Lee-Sammons, W. H., & Whitney, P. (1991). Reading perspectives and memory for text: An individual differences analysis. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 17,* 1074–1081.

Light, L. L., & Anderson, P. A. (1985). Working-memory capacity, age, and memory for discourse. *Journals of Gerontology, 40,* 737–747.

Lustig, C., May, C. P., & Hasher, L. (2001). Working memory span and the role of proactive interference. *Journal of Experimental Psychology: General, 130,* 199–207.

Masson, M. E., & Miller, J. A. (1983). Working memory and individual differences in comprehension and memory of text. *Journal of Educational Psychology, 75,* 314–318.

May, C. P., Kane, M. J., & Hasher, L. (1999). The role of interference in memory span. *Memory & Cognition, 27,* 759–767.

Miyake, A. (2001). Individual differences in working memory: Introduction to the special section. *Journal of Experimental Psychology: General, 130,* 163–168.

O'Brien, E. J., & Albrecht, J. E. (1992). Comprehension strategies in the development of a mental model. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 18,* 777–784.

Radvansky, G. A. (1999). Aging, memory and comprehension. *Current Directions in Psychological Science, 8,* 49–53.

Radvansky, G. A., & Copeland, D. E. (2000). Functionality and spatial relations in memory and language. *Memory & Cognition, 28,* 987–992.

Radvansky, G. A., & Copeland, D. E. (2001). Working memory and situation model updating. *Memory & Cognition, 29,* 1073–1080.

Radvansky, G. A., Copeland, D. E., & Zwaan, R. A. (2003). Aging and functional spatial relations in comprehension and memory. *Psychology and Aging, 18,* 161–165.

Radvansky, G. A., Gerard, L. D., Zacks, R. T., & Hasher, L. (1990). Younger and older adults' use of mental models as representations of text materials. *Psychology and Aging, 5,* 209–214.

Radvansky, G. A., Zwaan, R. A., Curiel, J. M., & Copeland, D. E. (2001). Situation models and aging. *Psychology and Aging, 16,* 145–160.

Rinck, M., & Bower, G. H. (1995). Anaphor resolution and the focus of attention in situation models. *Journal of Memory and Language, 34,* 110–131.

Sachs, J. S. (1967). Recognition memory for syntactic and semantic aspects of connected discourse. *Perception & Psychophysics, 2,* 437–442.

Schmalhofer, F., & Glavanov, D. (1986). Three components of understanding a programmer's manual: Verbatim, propositional, and situational representations. *Journal of Memory and Language, 25,* 279–294.

Shah, P., & Miyake, A. (1996). The separability of working memory resources for spatial thinking and language processing: An individual differences approach. *Journal of Experimental Psychology: General, 125,* 4–27.

Singer, M., Andrusiak, P., Reisdorf, P., & Black, N. L. (1992). Individual differences in bridging inference processes. *Memory & Cognition, 20,* 539–548.

Stanfield, R. A., & Zwaan, R. A. (2001). The effect of implied orientation derived from verbal context on picture recognition. *Psychological Science, 12,* 153–156.

Trabasso, T., & Sperry, L. L. (1985). Causal relatedness and importance of story events. *Journal of Memory and Language, 24,* 595–611.

Trabasso, T., & van den Broek, P. W. (1985). Causal thinking and the representation of narrative events. *Journal of Memory and Language, 24,* 612–630.

Turner, M. L., & Engle, R. W. (1989). Is working memory task dependent? *Journal of Memory and Language, 28,* 127–154.

van Dijk, T. A., & Kintsch, W. (1983). *Strategies in discourse comprehension.* New York: Academic Press.

Zwaan, R. A. (1994). Effect of genre expectations on text comprehension. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 20,* 920–933.

Zwaan, R. A., & Radvansky, G. A. (1998). Situation models in language comprehension and memory. *Psychological Bulletin, 123,* 162–185.