Assessing Levels of Narrative Memory Over Time

STEPHEN G. CHRONISTER, ANDREA. K. TAMPLIN,
and GABRIEL A. RADVANSKY
University of Notre Dame

Memory for text is represented at 3 levels: the surface form, textbase, and situation model. Although prior research has shown that these levels of representation can be differentially emphasized at the time of encoding, the influence of retrieval instructions on the use of these levels has not been tested. Moreover, there has been no assessment of how this influence might change over time. We assessed memory under both verbatim- and consistency-based instructions. For the verbatim instructions, people indicated whether probe sentences were actually read in the text. For consistency instructions, people responded based on whether probe sentences were consistent with what had been read earlier. A transitory influence view suggests that instructions at retrieval would guide the information used immediately, but not after a delay, when some levels of representation have faded. In contrast, a stable influence view suggests that retrieval instructions would guide the information used both immediately and after a delay. The results revealed that the verbatim instructions emphasized surface form and textbase measures, but consistency instructions emphasized situation model measures. This pattern shifted somewhat over a 1-week delay, with surface form memory becoming equivalent but the differences at the other 2 levels remaining.

KEYWORDS: situation models, text comprehension, narrative memory, levels of representation

Memory for text is captured at multiple levels of representation, namely the surface form, textbase, and situation model levels (van Dijk & Kintsch, 1983). The surface form conveys memory for the verbatim text, the textbase conveys memory for the propositional idea units presented in the text, and the situation model conveys the events described by the text. These levels of representation are important because they describe different aspects of knowledge that are stored in memory. Most previous work on these different levels has used retrieval instructions that emphasize the surface form (i.e., "please indicate whether each of these sentences was actually read in the previous text"), although there are exceptions (Bohay, Blakely, Tamplin, & Radvansky, 2011). Moreover, work has also shown that the strengths of these levels are differentially affected by the passage of time (Fisher & Radvansky, 2018; Kintsch, Welsch, Schmalhofer, & Zimney, 1990). The aim of the current study was to assess the influence of retrieval instructions on measures of levels of representation and whether this influence changes over time.

Encoding and Retrieval Instructions

One factor that can affect the levels of representation during encoding (reading) of a text is the reading

American Journal of Psychology
Summer 2022, Vol. 135, No. 2 pp. 139–149 •

instructions. A wide range of studies have assessed reading comprehension and memory and how they are influenced by such instructions (e.g., McCrudden, Magliano, & Schraw, 2010; Tzeng, van den Broek, Kendeou, & Lee, 2005). For example, prereading instructions can emphasize some aspects of a text and can lead to better memory for that type of information, although the other elements of a text may be encoded as well (Postman & Senders, 1946). Moreover, work on the Landscape model has emphasized that information is activated in memory, and inferences are drawn, with respect to different standards of coherence (e.g., whether one is reading for facts or understanding) (e.g., Todaro, Millis, & Dandotkar, 2010; van den Broek, Lorch, Linderholm, & Gustafson, 2001). These instructions then affect the resulting memory of the text (Blanc, Kendeou, van den Broek, & Brouillet, 2008; Linderholm, Virtue, Tzeng, & van den Broek, 2004; McCrudden, 2011; 2019; Rapp & McCrudden, 2018; van den Broek, Young, Tzeng, & Linderholm, 1999). Although such work has emphasized the developing memory representation during comprehension and the stability of the resulting memory representation, it seems plausible that such models could be used to address how different levels of representation could be emphasized during postcomprehension retrieval processes. We are unaware of any research that has explicitly addressed this issue from these perspectives.

For example, some research has shown that encoding instruction can influence the strength of these different levels. For example, Zwaan (1994) found that specific instructions before reading resulted in stronger recognition measures of either the surface form or situation model levels. A stronger surface form measure was obtained when instructions conveyed that the text was literature (and that how something is written was important), whereas a stronger situation model measure was obtained when instructions conveyed that the text was a news article (and that the described event was important).

Thus, emphasizing one level of representation at encoding can bias what is remembered later. Emphasizing wording and syntax leads to better surface form memory, whereas emphasizing described events leads to better situation model memory. What is unknown is whether instructions at the time of retrieval influence the use of information from these different levels of representation per se. One possibility is that the relative strength of memories in storage are set before retrieval and that they continue to have differential influences despite the instruction type. Thus, the prediction is that the relative strength of the different level measures would be similar under different instruction conditions.

Alternatively, consistent with other research on retrieval instruction more generally (e.g., McCrudden et al., 2010; Tzeng et al., 2005), different instructions could place different emphases on types of stored knowledge, with instructions emphasizing information tied to different levels. Although the information may be encoded in a certain way, what is actually selected and used during retrieval may vary. Thus, the prediction is that the relative strength of the level measures varies.

There is some evidence that instructions can influence retrieval. Similarly, the content that is retrieved can be influenced by the perspective people are encouraged to use. Taking the perspective of a burglar or a homebuyer both before and after reading changed memory for descriptions of a house (Anderson & Pichert, 1978; Anderson, Pichert, & Shirey, 1983; Baillet & Keenan, 1986; Borland & Flammer, 1985; Borland, Flammer, & Wearing, 1987; Kaakinen, Hyönä, & Keenan, 2001; Kardash, Royer, & Greene, 1988; Schraw & Dennison, 1994). Moreover, text memory can be affected by whether the schemas used to originally encode the information are used at retrieval based on the instructions given after reading (Hasher & Griffin, 1978). Although there is a bias to use the same schema at retrieval that was used during encoding, this can be avoided by discrediting the aptness of the schema by telling people that the original schema was incorrect. For example, if people were told before reading that a story was about a deer hunter, they might then be told at retrieval that it was actually about an escaped convict. When people use the new and different schema, the intrusion of schema-consistent but unmentioned inference information declines. Thus, this approach can be viewed as being analogous to a change in retrieval instructions because people are instructed to remember using either the original story title or a different one.

What those studies do not show is how retrieval instructions influence how the different levels of representation are used at retrieval. That is, does the established strength of the levels override instructional emphasis, or do instructions play a larger role? Some work has shown that there are differences in retrieval as a function of whether people are given verbatim or gist instructions. Specifically, gist instructions at retrieval lead to a greater acceptance of inferences (Kintsch, 1974; Singer, 1979; Singer & Remillard, 2008). This is because answering gist-based questions does not depend on verbatim text. However, this is an assessment of the acceptance rates of different probe types, not an assessment of the different levels of representation per se. Thus, it is unknown how retrieval instructions will influence measures of the three levels of representation. Note that we see a difference between gist and consistency instructions. Specifically, gist instructions would ask people to respond based on information that maps onto idea units present in a text (and thus would be more in line with textbase-level processing), whereas consistency instructions would include this information and any inferences drawn that go beyond the idea units of the text per se (and thus would be more in line with situation model processing).

Memory over Time

When considering memory, it is important to gather data over multiple delays (Ebbinghaus, 1885). Work by Fisher and Radvansky (2018; see also Kintsch et al., 1990) shows that performance at the different levels of representation is lost at different rates. Specifically, memory at the surface form level is lost quickly. In comparison, memory at the textbase level is strong immediately after learning but drops 1 week later. Finally, memory at the situation model levels is strong immediately after learning and remains so even 1 week later.

In this context, we also tested whether differences with verbatim and consistency instructions would vary with testing delay. We did so by comparing performance immediately after reading and 1 week later. Based on the prior work, we would expect that if there are any differences as a function of instruction type at retrieval, these would be observed immediate after reading, when all the levels of representation are accessible in memory.

After a delay there are two possibilities. One is that, because situation models are more resistant to forgetting, and there is a loss of information at the surface form and textbase levels, any influence of retrieval instruction would be attenuated. This is because people are forced into placing more of an emphasis on the situation model level. Alternatively, although there is differential forgetting at the levels of representation, retrieval processes can still be focused on different types of information. If so, then the expectation would be that any differences in the pattern of retrieval observed immediately after reading would be present later as well.

Measuring Levels of Representation

The levels of text representation can be assessed via a method developed by Schmalhofer and Glavanov (1986), which has subsequently been used widely (Bohay et al., 2011; Fisher & Radvansky, 2018; Fletcher & Chrysler, 1990; Kintsch et al., 1990; Narvaez, Radvansky, Lynchard, & Copeland, 2011; O'Rear & Radvansky, 2021; Radvansky, Copeland, Berish, & Dijkstra, 2003; Radvansky, Copeland, & von Hippel, 2010; Radvansky, Copeland, & Zwaan, 2003; Radvansky, O'Rear, & Fisher, 2017; Radvansky, Zwaan, Curiel, & Copeland, 2001; Wasiuk, Radvansky, Greene, & Calandruccio, 2021; Wolfe & Woodwyk, 2010; Zwaan, 1994). In this approach, people read a series of texts and then complete a recognition test for that material. These recognition test data are assessed via signal detection analysis to index memory strength at the surface form, textbase, and situation model levels. Specifically, during recognition there are four types of probes: verbatim, paraphrase, inference, and wrong. The verbatim probes are the exact sentences that were read earlier in a text. The paraphrases are probes that capture the same propositional ideas that were in the text but involve a different wording. For example, if the text was "the boy was kissed by the girl," a paraphrase probe would be "the girl kissed the boy." The inference probes were unstated ideas derived from prior knowledge that were expected to be integrated into the situation model. Finally, wrong probes are sentences that are thematically consistent with a text but inconsistent with the described situation.

The data from the recognition test are submitted to a series of signal detection analyses. For the surface form, the proportion of "yes" responses to verbatim probes is the hit rate, and that of paraphrases is the false alarm rate. These two conditions are used to derive an index of surface form memory because they both capture ideas that were actually present in the text, but only the verbatim probes capture the actual wording. Similarly, to derive a measure of the textbase, the paraphrase "yes" rates are treated as hits, and the inference "yes" rates are treated as false alarms. Here, neither of these probe types were actually presented in the text, but the ideas conveyed by the paraphrases were actually in the text. Thus, the degree to which people more often respond "yes" to paraphrases compared with inferences provides an index of the strength of the textbase representation. Finally, to derive a measure of the situation model, the inference "yes" rates are treated as hits, and the wrong "yes" rates are treated as false alarms. Here, in both cases the ideas were not actually present in the text, but in the case of the inference probes, the ideas are consistent with the situations that were described. Thus, the degree to which people respond "yes" more often to the inference probes than the wrongs is an index of the situation model level.

Our assessment of different levels of representation and retrieval instruction types bears some similarity to work on phantom recollection with narrative materials (Singer & Remillard, 2008; Singer & Spear, 2015). In brief, phantom recollection occurs when retrieval involves a phenomenological experience of remembering for items that were not actually encountered. That work also uses verbatim and gist retrieval instructions with verbatim, inference, and incorrect probes. Importantly, that work also had a retrieval condition in which people were to respond "yes" to inference probes but not verbatims. The focus of those studies was on the phenomenological experience (whether something is remembered or known) (Brainerd, Reyna, & Mojardin, 1999), which assumes that there are dual processes involved in memory retrieval (Mandler, 1972). For that line of research, verbatim instructions led participants to rely more on unconscious familiarity, and gist instructions led them to rely more on conscious recollection.

Aim of the Current Study

The aim of the current study was to explore the influence of instructions at test on measures of the different levels of text memory and how it might change over time. The degree to which instructions influence what is derived from these levels of representation would result in different patterns of remembering, and changes in the accessibility of information at the different levels over time may alter this pattern.

Most of the work that has assessed text memory via the Schmalhofer–Glavanov procedure has used instructions in which people were told to respond "yes" only for the probes that were actually in the text. Thus, this is a verbatim criterion. However, in everyday life, people infrequently remember verbatim information, which typically is lost within a few seconds or minutes (Sachs, 1967, 1974). Instead, when people remember a text, such as a news article, they focus on the described events, or the situation model, which is retained longer (Kintsch et al., 1990; Radvansky, O'Rear, & Fisher, 2017). Thus, this is a consistency criterion in which people assess whether something is consistent with what was read before.

One study that used the Schmalhofer–Glavanov method with a consistency criterion was a study by Bohay et al. (2011). They used this instruction to assess the benefits of note taking on memory for a spoken lecture and a written text. The instructions were to indicate whether a probe was consistent with what had been read earlier. The results revealed that note taking improved performance largely at the situation model level.

The current study assessed the influence of instruction type during retrieval on the measures of the levels of representation. Verbatim instructions place a greater emphasis on memory for the text itself and emphasize surface form and textbase memory. In comparison, consistency instructions place a greater emphasis on what the text was about and so would emphasize the situation model level. People are more willing to respond "yes" to sentences under consistency than verbatim instructions (Reder, 1982; Singer, 1979). However, it is unknown to what degree the levels of representation are involved and how this pattern may change over time.

The current study assessed instructions at the time of retrieval on measures of memory for text at different levels of representation. This was done with two groups of people. The instructions for both were the same at reading. At test, people received either verbatim criterion or consistency criterion instructions. The recognition data were then analyzed via a standard Schmalhofer–Glavanov analysis.

EXPERIMENT

METHOD

Participants

In this study 130 participants were tested: 63 in the verbatim group and 67 in the consistency group. They were drawn from the subject pool in the Department of Psychology at the University of Notre Dame (n = 60) and were given partial course credit or were selected from the pool of Mechanical Turk workers (n = 70) and were given monetary compensation for their participation. The Notre Dame and MTurk workers were evenly distributed across the conditions. All participants were native English speakers. Participation was approved by the institutional review board at the University of Notre Dame.

Materials

Four texts were drawn from Radvansky et al. (2001). These texts were 58 to 85 sentences long and described historical events: the development of the marine chronometer in 18th-century England, the British Gunpowder Plot of 1605, the Australian Rum Rebellion of 1807, and the Dutch tulip craze of the 17th century. These texts had Fletcher–Kincaid grade levels from 7.4 to 11.0 (M = 9.0). They were chosen because participants were unlikely to have knowledge about them.

For the recognition test, eight sentences were selected from each text. Four memory probes were created from each of the selected sentences. The first was the verbatim probe, an actual sentence from the text (e.g., "The plot bitterly intensified Protestant suspicions of Catholics"). The second was the paraphrase probe, a rewording of the sentence that retained the propositional content of the original by using synonyms or altering word order (e.g., "The plot greatly heightened Protestant distrust of Catholics"). The third was the inference probe, information that was important to the description but was not explicitly mentioned (e.g., "The plot led to increased acts of persecution of Catholics"). Finally, the fourth was the incorrect probe, information that was not mentioned and was unlikely to be inferred. However, the information was globally consistent with the passage's theme (e.g., "After the plot, donations to Protestant churches rose dramatically").

Procedure

For this task, all the texts were read in a sentence-by-sentence manner. After reading all the texts, participants were given a recognition test. On each trial, they viewed a single sentence and had to provide a response. Each person saw only one version (verbatim, paraphrase, inference, or wrong) of each test sentence. These were half of the recognition probes. Thus, there were 16 probes, four of each type. The sentence versions were rotated across participants. The two groups in this study differed in their retrieval instructions. For the verbatim group, participants were told to indicate whether a probe sentence was one that they had read before. In comparison, for the consistency group, participants had to indicate whether a probe sentence was consistent with the text. The various texts were presented in a different random order for each participant.

The recognition memory instructions for the verbatim group were as follows:

> In this section, you will be presented with a series of sentences, one at a time. Your task is to decide, yes or no, whether that exact sentence was presented in the stories you read earlier. If it was, press the "Y" key. If it was not, press the "N" key.

The instructions for the consistency group were as follows:

> In this section, you will be presented with a series of sentences, one at a time. Your task is to decide, yes or no, whether each sentence was consistent with the stories you read earlier. If it was, press the "Y" key. If it was not, press the "N" key.

A week later, participants returned for the second memory test, like the first, but with different items. These were the other half of the recognition probe items. There was no opportunity to review the texts. This allowed us to assess any memory change over a 1-week interval.

RESULTS

The recognition data are summarized in Table 1 and Figure 1, with Table 1 reporting the "yes" response rates and Figure 1 reporting the results of the signal detection analysis. The "yes" response rate data were submitted to an A' signal detection analysis (Snodgrass & Corwin, 1988) to gain measures of the surface form, textbase, and situation model levels (following Schmalhofer & Glavanov, 1986). A' is a nonparametric measure that we have used extensively in our

TABLE 1. Recognition Rate (*SE*) of Responding "Yes" on the Immediate and Delayed Tests for Each of the 4 Memory Probe Types at the 2 Testing Intervals

		Immediate test		
Instructions	**Verbatim**	**Textbase**	**Inference**	**Wrong**
Verbatim	.64 (.03)	.46 (.04)	.32 (.04)	.21 (.03)
Consistency	.74 (.02)	.73 (.03)	.66 (.04)	.37 (.04)
		1-Week delay test		
Instructions	**Verbatim**	**Textbase**	**Inference**	**Wrong**
Verbatim	.54 (.03)	.51 (.03)	.46 (.03)	.38 (.04)
Consistency	.60 (.03)	.57 (.03)	.65 (.03)	.41 (.03)

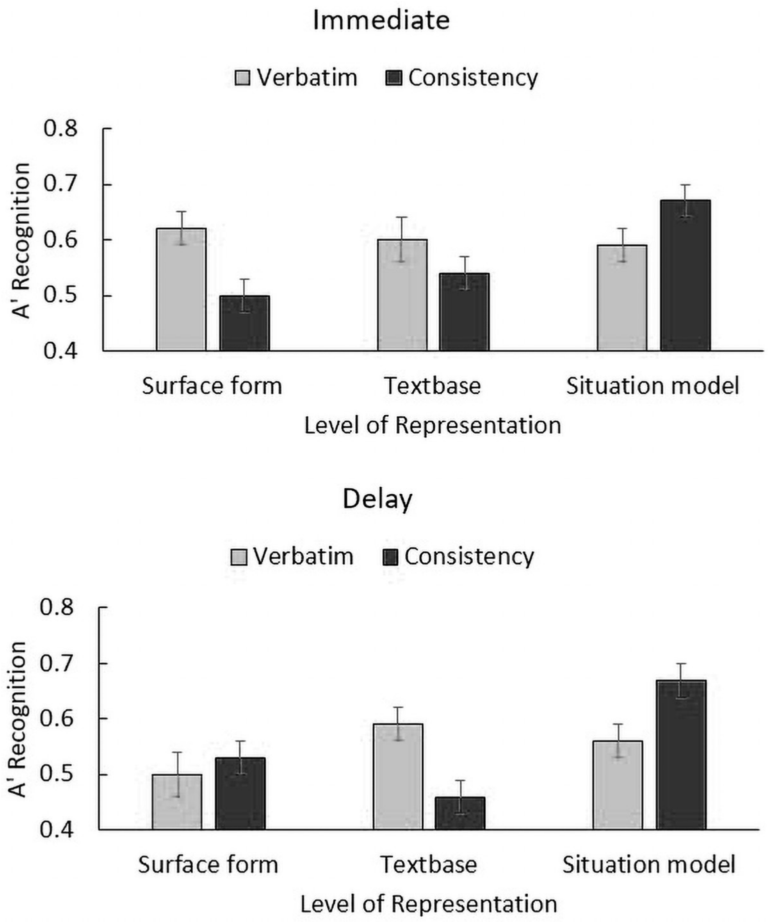

FIGURE 1. *A′* recognition scores for the immediate and delayed recognition tests for each of the three levels of representation. Chance is .5.

prior work of this type.[1] We first assessed whether these three measures were significantly different from chance. For the verbatim group, all three measures in the immediate condition were greater than chance (.5), $t(62) = 3.98$, $p < .001$; $t(62) = 2.77$, $p = .007$; and $t(62) = 2.86$, $p = .006$, for the surface form, textbase, and situation model levels, respectively. However, after a 1-week delay, performance was not different from chance for the surface form measure, $t(62) = 0.12$, $p = .90$, but was for the textbase measure, $t(62) = 2.89$, $p = .005$, and marginally so for the situation model measure, $t(62) = 1.81$, $p = .08$. This is consistent with the idea that a verbatim retrieval instruction places greater emphasis on memory for a text itself rather than the situations described by a text. In comparison, for the consistency group, only the situation model measure was significant, and this was true both immediately, $t(66) = 0.18$, $p = .86$; $t(66) = 1.04$, $p = .30$; and $t(66) = 5.18$, $p < .001$, for the surface form, textbase, and situation model levels, respectively, and after a 1-week delay, $t(66) = 0.85$, $p = .40$; $t(66) = 1.30$, $p = .20$; and $t(66) = 6.80$, $p < .001$. Thus, consistency instructions place a greater emphasis on memory for the situations described by the texts rather than the texts themselves.

The A' scores at each level of representation were submitted to a 2 (group) × 2 (delay) mixed ANOVA, with the first variable being between participants and the second within. For the surface form, neither main effect of group nor the main effect of delay was significant, $F(1, 128) = 2.32$, $MSE = .061$, $p = .13$, $\eta_p^2 = .02$, and $F(1, 128) = 1.99$, $MSE = .069$, $p = .16$, $\eta_p^2 = .02$, respectively. However, there was a significant interaction, $F(1, 128) = 5.97$, $MSE = .069$, $p = .02$, $\eta_p^2 = .05$. Simple effects tests revealed that performance differed for the two groups for immediate testing, $F(1, 128) = 9.74$, $MSE = .053$, $p = .002$, $\eta_p^2 = .07$, but not for delayed testing, $F < 1$. This is consistent with the idea that the verbatim instructions placed a greater emphasis on memory of the text itself and that surface form memory, though more available immediately, is lost quickly (Sachs, 1967, 1974). This is then followed by a loss of this type of information for later text memory retrieval.

For the textbase, there was a main effect of group, $F(1, 128) = 7.70$, $MSE = .074$, $p = .006$, $\eta_p^2 = .06$, with people in the verbatim group scoring higher than those in the consistency group. Congruent with the idea that the retrieval instructions emphasize different levels of representation, the verbatim group had superior memory at the textbase level because of the greater emphasis on the text itself. The main effect of delay was not significant, $F(1, 128) = 1.88$, $MSE = .065$, $p = .17$, $\eta_p^2 = .02$, with people scoring similarly on the immediate and delayed tests. The interaction was also not significant, $F(1, 128) = 1.09$, $MSE = .065$, $p = .30$, $\eta_p^2 = .01$. Thus, although some studies have suggested that there is a decline in the accessibility of textbase-level information, it was still present to a sufficient degree to be differentially selected and used based on the retrieval instructions.

Finally, for the situation model there was a main effect of group, $F(1, 128) = 10.66$, $MSE = .064$, $p = 001$, $\eta_p^2 = .08$, with people in the consistency group scoring higher than those in the verbatim group. This is congruent with the idea that the retrieval instructions place different emphases on different memory representations, with the consistency instructions placing a greater emphasis on the situation model level. In addition, neither the main effect of delay nor the interaction was significant, both Fs < 1. This is consistent with prior work showing that there is little forgetting at the situation model level (Kintsch et al., 1990; Radvansky et al., 2017). This is also consistent with the idea that retrieval instructions continue to have an influence on the type of information that is retrieved from memory of the text.

DISCUSSION

The aim of this study was to explore the degree to which retrieval instructions influence measures of the different levels of text memory (van Dijk & Kintsch, 1983) and how these change over time. Although reading instructions can influence levels of representation, it is unclear what effect retrieval instructions might have and whether this will change as the accessibility of the different levels changes over time (Fisher & Radvansky, 2018). As expected, during immediate retrieval, when the instructions emphasized the text itself (verbatim), people scored higher on both the surface form and textbase measures. In contrast, when the instructions emphasized the events described by the text (consistency), people scored higher on the situation model measure.

Over time, this pattern exhibited evidence of both change and stability. In terms of change, although there was greater use of the surface form level with verbatim instructions, there was no difference between the groups a week later. This probably occurred because the surface form information was no longer accessible in memory, meaning that although the verbatim instructions asked people to use this information, this could no longer be done, and so there was no difference between the groups.

In terms of stability, although prior research has reported a decline in memory strength for the textbase level (Fisher & Radvansky, 2018; Kintsch et al., 1990), no such decline was observed in this study. People who received verbatim instructions placed a greater emphasis on this level than people who received the consistency instructions. Moreover, this did not change over the delay. For the situation model levels, there was no decline in memory strength over time, which is consistent with prior work. There was also no difference in the degree to which information at this level was used by people, and this did not change over time.

Thus, overall, there was some support for both of the possibilities outlined at the beginning of this article. There was some evidence of a change in emphasis at the surface form level, largely due to forgetting at this level. However, whenever information remained accessible in memory, how it was used was influenced by the retrieval instructions. Extrapolating from this, it would be easy to see that as the textbase memory declines further, as it is likely to do, then all that would be left would be the situation model level, and performance would probably be similar for the different instruction types.

CONCERNS AND LIMITATIONS

One concern that could be raised about the current study is whether signal detection analysis is appropriate here. We think that it is. Specifically, in both the verbatim and consistency conditions, people are given the task of indicating whether a probe sentence conforms to a memory, regardless of whether the instructions are more strict or lenient. For our surface form measure, the verbatim probes would more closely match memories than the paraphrases. For the textbase measure, the paraphrases would match better than the inferences. Finally, for the situation model measure, the inferences would match better than the wrongs. Given this setting, one would expect there to be an impact of instructions on the bias to say "yes."[2]

Given the nature of the information accessible from memory for the different probe types, it is straightforward to understand the results in terms of how this information is applied during the memory test. Moreover, it can also be seen that the information emphasized by these instructions is used, but only when it is accessible, and this accessibility changes over time.

As mentioned at the beginning of this article, one way of thinking about the current results is in terms of how different instructions relate to the type of retrieval process involved. Specifically, the question is whether there is more of an emphasis on familiarity or recollection (Singer & Remillard, 2008; Singer & Spear, 2015). Prior work has suggested that verbatim instructions encourage participants to place more emphasis on unconscious familiarity, whereas gist instructions have more of an emphasis on conscious recollection. At first blush this seems counterintuitive. After all, verbatim memory requires a more precise set of information than does gist memory. However, further consideration reveals why this is not the case. Specifically, verbatim memory requires surface form information, which is lost quickly. Thus, people need to rely more on familiarity for more impoverished memory representations. In comparison, situation model memory is more enduring. Thus, it is more likely that one would be able to better recollect the events described by a text. In this way, the current results line up with the prior work on phantom recollection. However, the current data cannot be used to separate out such memory processes, and the phantom recollection data have not assessed memory at these multiple levels of representation. The next step would be to combine these two approaches.

In addition to instructions given to participants, memory retrieval tasks in general can also influence performance. For example, in a study by Curiel and Radvansky (2002), people memorized a map of a research center and then either read narratives about characters moving about the building (Rinck & Bower, 1995) or took a recognition test to assess priming of map locations (McNamara, 1986). Importantly, map memorization was done so that spatial and temporal proximity were deconfounded[3] (Clayton & Habibi, 1991; Curiel & Radvansky, 1998). Whereas the recognition priming data gave evidence of temporal but not spatial priming (an emphasis on temporal memory), the reading time data gave evidence of a spatial gradient of accessibility (an emphasis on spatial memory). Thus, these studies illustrate that both instructions and task can influence the content of the retrieved memories.

There are some other more general limitations to our study. First, we did not collect demographic characteristics of our samples that might reveal interesting differences. For example, when instructions are manipulated before comprehension, this can result in dramatic changes in memory performance, with people better remembering the type of information that is emphasized by those instructions (Bohn-Gettler & McCrudden, 2018; McCrudden, 2011; 2019; McCrudden et al., 2010; McCrudden & Schraw, 2007, 2009). However, it is important to note that our data were collected from a homogeneous high-achieving academic population, and such an individual difference measure would probably have too little variance to provide interesting insights into the mechanisms at work here.

Another concern is that participants may have interpreted the term *consistent* in various ways. Thus, there may be some variation in how the instructions were interpreted across participants. That said, there are clear differences in the pattern of performance in the two retrieval instruction conditions.

Overall, for assessing knowledge, if the primary aim is to assess whether people have learned a set of facts, then a verbatim criterion is better. In contrast, if the aim is to assess a broader understanding beyond the text itself, then the memory at the situation model level is better emphasized by consistency instructions. This is likely to be the type of instruction that would be useful in most cases. When people read, they often try to understand the events being described. This often involves a parsing and segmenting of the stream of activity, an understating of the structural relations between entities involved in an event, an establishment of the causal structure that explains why different things happen, and so on. This understanding is captured in the situation model, not in the surface form or textbase levels. Verbatim assessments are more likely to miss such an understanding and provide a shallower evaluation of what people comprehend from a text.

NOTES

Open practices statement: The data and materials for this experiment are available at https://osf.io/5pf9k/?view_only=c14fa00213284e20a970d328094130c0.

This research was supported by an Undergraduate Research Opportunities grant from the Institute for Scholarship in the Liberal Arts at the University of Notre Dame, awarded to Steve Chronister.

Address correspondence about this article to G. A. Radvansky, 366 Corbett Hall, Department of Psychology, University of Notre Dame, Notre Dame, IN 46556 (email: gradvans@nd.edu).

1. When hit and false alarm rates were either both 0 or 1, *A′* values were set to .5, chance, on this principle that memory was essentially absent.

2. This is confirmed in our online supplement.

3. That is, map items that were near each other in space were encountered far from each other in the temporal order of learning, and map items that were near one another in the temporal order during learning were far from one another on the map.

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.

Anderson, R. C., Pichert, J. W., & Shirey, L. L. (1983). Effects of the reader's schema at different points in time. *Journal of Educational Psychology, 75,* 271–279.

Baillet, S. D., & Keenan, J. M. (1986). The role of encoding and retrieval processes in the recall of text. *Discourse Processes, 9,* 247–268.

Blanc, N., Kendeou, P., Van Den Broek, P., & Brouillet, D. (2008). Updating situation models during reading of news reports: Evidence from empirical data and simulations. *Discourse Processes, 45,* 103–121.

Bohay, M., Blakely, D. P., Tamplin, A. K., & Radvansky, G. A. (2011). Note-taking, review, memory, and comprehension. *American Journal of Psychology, 124,* 63–73.

Bohn-Gettler, C. M., & McCrudden, M. T. (2018). Effects of task relevance instructions and topic beliefs on reading processes and memory. *Discourse Processes, 55,* 410–431.

Borland, R., & Flammer, A. (1985). Encoding and retrieval processes in memory for prose. *Discourse Processes, 8,* 305–317.

Borland, R., Flammer, A., & Wearing, A. J. (1987). Text memory: Recalling twice, using different perspectives. *European Journal of Psychology of Education, 2,* 209–217.

Brainerd, C. J., Reyna, V. F., & Mojardin, A. H. (1999). Conjoint recognition. *Psychological Review, 106,* 160–179.

Clayton, K., & Habibi, A. (1991). Contribution of temporal contiguity to the spatial priming effect. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 17,* 263–271.

Curiel, J. M., & Radvansky, G. A. (1998). Mental organization of maps. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 24,* 202–214.

Curiel, J. M., & Radvansky, G. A. (2002). Mental maps in memory retrieval and comprehension. *Memory, 10,* 113–126.

Ebbinghaus, H. (1885). *Memory: A contribution to experimental psychology* (H. A. Ruger & C. E. Bussenius, Trans.). Columbia University, Teacher's College.

Fisher, J. S., & Radvansky, G. A. (2018). Patterns of forgetting. *Journal of Memory and Language, 102,* 130–141.

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

Hasher, L., & Griffin, M. (1978). Reconstructive and reproductive processes in memory. *Journal of Experimental Psychology: Human Learning and Memory, 4,* 318–330.

Kaakinen, J. K., Hyönä, J., & Keenan, J. M. (2001). Individual differences in perspective effects on text memory. *Current Psychology Letters: Behaviour, Brain & Cognition, 5,* 21–32.

Kardash, C. A., Royer, J. M., & Greene, B. A. (1988). Effects of schemata on both encoding and retrieval of information from prose. *Journal of Educational Psychology, 80,* 324–329.

Kintsch, W. (1974). *The representation of meaning in memory*. Erlbaum.

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

Linderholm, T., Virtue, S., Tzeng, Y., & Van den Broek, P. (2004). Fluctuations in the availability of information during reading: Capturing cognitive processes using the Landscape model. *Discourse Processes, 37,* 165–186.

Mandler, G. (1972). Organization and recognition. In E. Tulving & W. Donaldson (Eds.), *Organization of memory* (pp. 139–166). Academic Press.

McCrudden, M. T. (2011). Do specific relevance instructions promote transfer appropriate processing? *Instructional Science, 39,* 865–879.

McCrudden, M. T. (2019). The effect of task relevance instructions on memory for text with seductive details. *Applied Cognitive Psychology, 33,* 31–37.

McCrudden, M. T., Magliano, J. P., & Schraw, G. (2010). Exploring how relevance instructions affect personal reading intentions, reading goals and text processing: A mixed methods study. *Contemporary Educational Psychology, 35,* 229–241.

McCrudden, M. T., & Schraw, G. (2007). Relevance and goal-focusing in text processing. *Educational Psychology Review, 19,* 113–139.

McCrudden, M. T., & Schraw, G. (2009). The effects of relevance instructions and verbal ability on text processing. *Journal of Experimental Education, 78,* 96–117.

McNamara, T. P. (1986). Mental representations of spatial relations. *Cognitive Psychology, 18,* 87–121.

Narvaez, D., Radvansky, G. A., Lynchard, N. A., & Copeland, D. E. (2011). Are older adults more attuned to morally charged information? *Experimental Aging Research, 37,* 398–434.

O'Rear, A. E., & Radvansky, G. (2021). Encoding and referent event influence on retrospective memory. *Quarterly Journal of Experimental Psychology, 74,* 1117–1127.

Postman, L., & Senders, V. L. (1946). Incidental learning and generality of set. *Journal of Experimental Psychology, 36,* 153–165.

Radvansky, G. A., O'Rear, A. E., & Fisher, J. S. (2017). Event models and the fan effect. *Memory & Cognition, 45,* 1028–1044.

Radvansky, G. A., Copeland, D. E., Berish, D. E., & Dijkstra, K. (2003). Aging and situation model updating. *Aging, Neuropsychology, and Cognition, 10,* 158–166.

Radvansky, G. A., Copeland, D. E., & von Hippel, W. (2010). Stereotype activation, inhibition, and aging. *Journal of Experimental Social Psychology, 46,* 51–60.

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

Rapp, D. N., & McCrudden, M. T. (2018). Relevance before, during, and after discourse experiences. In J. Strassheim & H. Nasu (Eds.), *Relevance and irrelevance: Theories, factors and challenges* (pp. 141–160). De Gruyter.

Reder, L. M. (1982). Plausibility judgments versus fact retrieval: Alternative strategies for sentence verification. *Psychological Review, 89,* 250–280.

Rinck, M., & Bower, G. H. (1995). Anaphora 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.

Sachs, J. S. (1974). Memory in reading and listening to discourse. *Memory & Cognition, 2,* 95–100.

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.

Schraw, G., & Dennison, R. S. (1994). The effect of reader purpose on interest and recall. *Journal of Reading Behavior, 26*(1), 1–18.
Singer, M. (1979). Temporal locus of inference in the comprehension of brief passages: Recognizing and verifying implications about instruments. *Perceptual & Motor Skills, 49,* 539–550.
Singer, M., & Remillard, G. (2008). Veridical and false memory text: A multiprocess analysis. *Journal of Memory and Language, 59,* 18–35.
Singer, M., & Spear, J. (2015). Phantom recollection of bridging and elaborative inferences. *Discourse Processes, 52,* 356–375.
Snodgrass, J. G., & Corwin, J. (1988). Pragmatics of measuring recognition memory: Applications to dementia and amnesia. *Journal of Experimental Psychology: General, 117,* 34–50.
Todaro, S., Millis, K., & Dandotkar, S. (2010). The impact of semantic and causal relatedness and reading skill on standards of coherence. *Discourse Processes, 47,* 421–446.
Tzeng, Y., van den Broek, P., Kendeou, P., & Lee, C. (2005). The computational implementation of the Landscape model: Modeling inferential processes and memory representations of text comprehension. *Behavior Research Methods, 37,* 277–286.
van den Broek, P., Lorch, R. F., Linderholm, T., & Gustafson, M. (2001). The effects of readers' goals on inference generation and memory for texts. *Memory & Cognition, 29,* 1081–1087.
van den Broek, P., Young, M., Tzeng, Y., & Linderholm, T. (1999). The Landscape model of reading: Inferences and the online construction of a memory representation. In H. van Oostendorp & S. R. Goldman (Eds.), *The construction of mental representations during reading* (pp. 71–98). Erlbaum.
van Dijk, T. A., & Kintsch, W. (1983). *Strategies in discourse comprehension*. Academic Press.
Wasiuk, P. A., Radvansky, G. A., Greene, R. L., & Calandruccio, L. (2021). Spoken narrative comprehension for young adult listeners: Effects of competing voices and noise. *International Journal of Audiology, 60,* 711–722.
Wolfe, M B. W., & Woodwyk, J. M. (2010). Processing and memory of information presented in narrative or expository texts. *British Journal of Educational Psychology, 80,* 341–362.
Zwaan, R. A. (1994). Effect of genre expectations of text comprehension. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 20,* 920–933.