The relative roles of visuospatial and linguistic working memory systems in generating inferences during visual narrative comprehension

Joseph P. Magliano, Adam M. Larson, Karyn Higgs & Lester C. Loschky

Memory & Cognition

ISSN 0090-502X

Mem Cogn
DOI 10.3758/s13421-015-0558-7

Springer

Mem Cogn
DOI 10.3758/s13421-015-0558-7

The relative roles of visuospatial and linguistic working memory systems in generating inferences during visual narrative comprehension

Joseph P. Magliano[1] · Adam M. Larson[2] · Karyn Higgs[1] · Lester C. Loschky[3]

Abstract This study investigated the relative roles of visuospatial versus linguistic working memory (WM) systems in the online generation of bridging inferences while viewers comprehend visual narratives. We contrasted these relative roles in the *visuospatial primacy* hypothesis versus the *shared* (*visuospatial & linguistic*) *systems* hypothesis, and tested them in 3 experiments. Participants viewed picture stories containing multiple target episodes consisting of a beginning state, a bridging event, and an end state, respectively, and the presence of the bridging event was manipulated. When absent, viewers had to infer the bridging-event action to comprehend the end-state image. A pilot study showed that after viewing the end-state image, participants' think-aloud protocols contained more inferred actions when the bridging event was *absent* than when it was present. Likewise, Experiment 1 found longer viewing times for the end-state image when the bridging-event image was absent, consistent with viewing times revealing online inference generation processes. Experiment 2 showed that both linguistic and visuospatial WM loads attenuated the inference viewing time effect, consistent with the shared systems hypothesis. Importantly, however, Experiment 3 found that articulatory suppression did not attenuate the inference viewing time effect, indicating that (sub)vocalization did not support online inference generation during visual narrative comprehension. Thus, the results support a shared-systems hypothesis in which both visuospatial and linguistic WM systems support inference generation in visual narratives, with the linguistic WM system operating at a deeper level than (sub)vocalization.

✉ Joseph P. Magliano
jmagliano@niu.edu

[1] Department of Psychology, Northern Illinois University, DeKalb, IL 60115, USA

[2] Findlay College, Findlay, OH, USA

[3] Kansas State University, Manhattan, KS, USA

Keywords Language comprehension · Working memory

An inherent feature of all narratives is that only a selected portion of the events that comprise the plot are explicitly conveyed, and therefore events that bridge the gaps between the explicitly conveyed events must be inferred (e.g., Graesser, Singer, & Trabasso, 1994; Magliano, Baggett, Johnson, & Graesser, 1993; Singer & Halldorson, 1996). This feature of narratives is particularly evident in sequential narratives, such as comic strips and graphic novels (Cohn, 2014; McCloud, 1993). Consider a sequential, two-panel excerpt from the graphic story *The Fantom of the Fair* (Gustavson, 1939; see Fig. 1). The first panel shows Fantoman and a woman rising to the surface of a bay that they jumped into in order to escape an attempt on their lives. The next panel shows the woman and Fantoman on land, with Fantoman leaping away from the woman. Presumably, comprehending the panels would require viewers to infer the missing events (e.g., "They emerged from the water"), as is well documented for text (for extensive reviews, see Graesser et al., 1994; Magliano & Graesser, 1991). These are referred to as "bridging inferences" because they establish how two or more elements of a text are semantically related.

Given that bridging inferences are assumed to be essential for text comprehension (e.g., Graesser et al., 2004; McNamara & Magliano, 2009), it is highly likely that they are also important for the comprehension of visually based narratives (Cohn & Wittenberg, 2015; Magliano, Loschky, Clinton, & Larson, 2013; Nakazawa, 2005; Saraceni, 2001, 2003; West & Holcomb, 2002). However, the study of how people

Published online: 08 October 2015

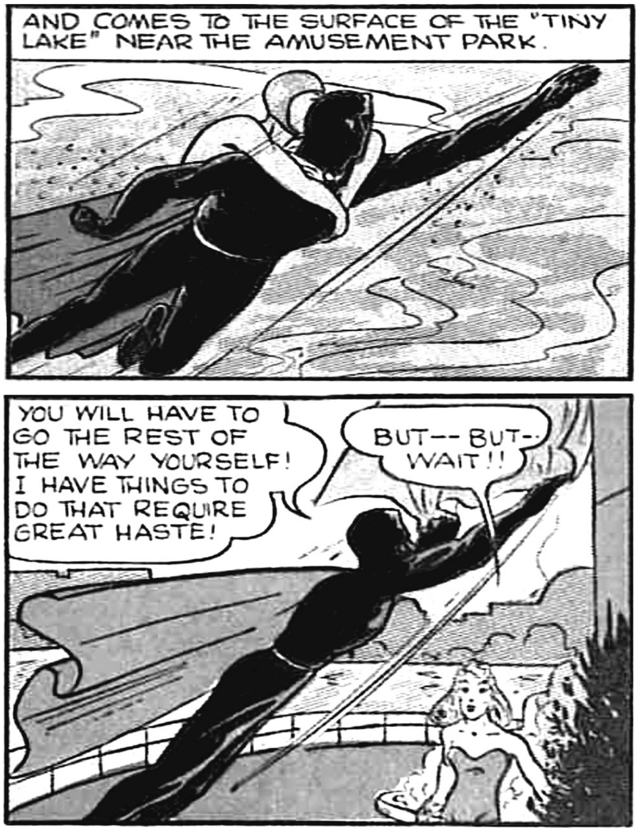
AND COMES TO THE SURFACE OF THE "TINY LAKE" NEAR THE AMUSEMENT PARK.
YOU WILL HAVE TO GO THE REST OF THE WAY YOURSELF! I HAVE THINGS TO DO THAT REQUIRE GREAT HASTE!
BUT-- BUT- WAIT!!

Fig. 1 Sequential, two-panel excerpt from the graphic story *The Fantom of the Fair*

comprehend and process visual narratives (film or v narratives) is in its nascent stages, and to our knowledge, the field is lacking studies that explore *how* bridging inferences are constructed during the moment-to-moment processing of visually based narratives (see Cohn & Wittenberg, 2015, for a recent exception).

While there is some evidence for medium-specific knowledge (i.e., learned through repeated experiences with the medium) that supports the processing of visually based narratives (Nakazawa, 2005; Schwan & Ildirar, 2010), there is also good evidence for a relatively high degree of overlap in the nature of higher level representations (e.g., mental models) for text-based and visually based narratives (Baggett 1975, 1979; Gernsbacher, 1990; Gernsbacher, Varner, & Faust, 1990; Magliano et al., 2013; Magliano, Radvansky, & Copeland, 2007). Moreover, working memory resources should be involved in the generation of inferences in sequential narratives. A plethora of studies shows that WM resources (e.g., WM capacity) predict text readers' generation of bridging inferences (Allbritton, 2004; Calvo, 2001, 2005; Just & Carpenter, 1992; Linderholm & van den Broek, 2002; Rai, Loschky, Harris, Peck, & Cook, 2011, Rai, Loschky & Harris, 2014; Reichle & Mason, 2007; St. George, Mannes, & Hoffman, 1997). For example, readers with lower WM capacity are less likely to generate bridging inferences (Calvo, 2001; Linderholm & van den Broek, 2002), and when they do they require more processing time to do so (Estevez & Calvo, 2000, 2011, Rai et al., 2014). Furthermore, when WM resources are consumed, for example by a WM load task or by anxiety, it exacerbates these relationships (Darke, 1988; Rai et al., 2011, 2014).

However, this does not necessarily imply that the same working memory resources and knowledge systems are drawn upon to *generate* functionally similar inferences across media. Given that sequential narratives involve the presentation of visuospatial content, it makes sense that perceptual and visuospatial working memory (WM) processes would support the computation of bridging inferences in this medium. A less obvious possibility is that linguistic resources could also support the computation of bridging inferences during sequential narrative comprehension. As we will discuss below, Cohn (2013a, b, 2014; Cohn, Paczynski, Jackendoff, Holcomb, & Kuperberg, 2012) has argued that sequentila narratives share basic features with language, and therefore some of the same WM and cognitive systems (lexical, semantic, and syntactic systems) that support language comprehension processes may also support comprehension of sequential narratives.

The goals of this study were therefore (1) to determine whether the generation of bridging inferences while comprehending wordless sequential narratives can be revealed by analyzing picture viewing times in a manner similar to the analysis of sentence reading times in text comprehension (e.g., Clark, 1977), and, if so, then (2) to identify the possible roles of linguistic and perceptual (visuospatial) WM resources in the moment-to-moment computation of those inferences. However, addressing Goal 2 is arguably the most important contribution of the present study.

A case for a role of visuospatial working memory processes in the comprehension of sequential narratives

Given that sequential narratives differ from text-based narratives, at a minimum, in terms of their use of the pictorial medium, the study of inferences in sequential narratives should consider the role of perceptual processes. To this end, we have recently proposed the Scene Perception and Event Comprehension theory (SPECT; Loschky, Hutson, Magliano, Larson, & Smith, 2014; Loschky, Smith, & Magliano, 2015). A key distinction in SPECT is between the front end, which involves visual processing of information within single fixations, and the back end, which involves processing in memory across multiple fixations. Front-end processes involve visual perceptual processing, of both a scene as a whole and the individual objects within it. For example, there is evidence that recognizing the semantic category of a scene at the superordinate level (e.g., indoors) precedes recognizing it at the basic level (e.g., kitchen), and both precede recognizing basic-level actions carried out by a person within the scene (e.g., cooking; Larson & Loschky, submitted). According to SPECT, such information (i.e., entities [a man], their actions [cooking], and locations [kitchen]) are processed in WM in order to create a semantic representation of the currently viewed event, which leads to an updating of the mental model of the narrative in episodic long-term memory (LTM) to incorporate that event (e.g., Gernsbacher, 1990; Kintsch, 1998; Zwaan & Radvansky, 1998). The process of updating the mental model in WM and LTM would fall under the category of back-end processes (Loschky et al., 2014, 2015; Magliano et al., 2013).

Of critical importance to the present study, according to SPECT, the perceptual processes in the front end have implications for back-end processes that support mental model construction, such as bridging inferences. For example, perception of visuospatial relationships, their storage in WM, and their comparison across panels may be important for indicating when a bridging inference is required. Consider the change in the spatial relationships between the entities shown in Panels 1 and 2 of Fig. 1. The drastic changes in the spatial orientations and relationships between characters across the panels likely indicates that there were intervening events that caused those changes. However, to perceive such changes, visuospatial relationships between entities in Panel 1 must have been stored in WM, so that they could be compared to the scene depicted in Panel 2. In this way, visuospatial relationships in WM may play an important role in generating bridging inferences while comprehending sequential narratives.

If such visuospatial WM resources play a critical role in generating bridging inferences while comprehending sequential narratives, one may ask whether they actually differ from linguistic WM resources. There are plausible reasons to believe that such visuospatial WM representations are distinct from verbally coded spatial representations. First, there is a wealth of WM research showing clear distinctions between visuospatial and verbal WM (e.g., Baddeley, 2001; Cocchini, Logie, Della Sala, MacPherson & Baddeley, 2002; Shah & Miyake, 1996; Smith & Jonides, 1997; Smith, Jonides & Koeppe, 1996). Second, vision and its concomitant visuospatial representations clearly must have evolved far earlier than language. Thus, if such visuospatial WM resources are used to generate bridging inferences for visual scenes and events, they likely were used by humans prior to the evolution of language, and by other nonlinguistic animals to comprehend their environments (Herman & Austad, 1996).

A case for a role of language in the comprehension of sequential narratives

Are language systems involved in comprehending wordless sequential narratives (e.g., comics, children's picture stories)? One answer to this question is proposed in Cohn's (2013a, 2014) visual language theory (VLT), which explains how sequential narratives are processed and comprehended. A central assumption of VLT is that sequential narratives share important features with language in that they both have semantic content (conveyed by words and/or pictures) and sequencing rules (e.g., grammar) that govern how that content can be combined to form meaningful sequences. Moreover, both text-based and sequential narratives involve conveying a sequence of events that comprise a hierarchically structured plot (e.g., structured around goal plans; Magliano et al., 2013; Trabasso & Nickels, 1992). As such, VLT assumes that there are shared cognitive and brain systems that support the processing of both linguistic discourse and sequential narratives. The present study provides a test of this assumption of VLT, specifically in terms of assessing linguistic WM resources that potentially support the construction of bridging inferences in sequential narratives.

What evidence is there that language supports the processing of wordless sequential narratives? Much of the evidence comes from the research testing the assumption of VLT that the sequencing of panels in sequential narratives follows a *visual narrative grammar* (Cohn, 2013a, 2014; Cohn, Jackendoff, Holcomb, & Kuperberg, 2014). The sequential narrative grammar specifies that panels fit into categorical roles that can be combined to form *visual phrases*, which function similarly to linguistic phrases (e.g., noun and verb phrases) in a sentence. Violations of visual phrase structure in a sequential narrative disrupts processing and has a similar event related potential (ERP) signature to violations of linguistic phrase structure in sentence processing (i.e., N400 responses to structural violations; Cohn et al., 2014). Additionally, panel sequences that are consistent with a sequential narrative grammar are judged to be more coherent than inconsistent sequences, and in a panel-ordering task, randomly presented panels are ordered consistently with rules of the sequential narrative grammar (Cohn, 2014).

Additionally, under certain circumstances, viewers spontaneously activate linguistic knowledge when processing visual images (Brandimonte, Hitch, & Bishop, 1992a, b; Hitch, Brandimonte, & Walker, 1995; Meyer, Belke, Telling, & Humphreys, 2007). Brandimonte et al. (1992a) showed viewers ambiguous visual images, and found that viewers naturally engaged in linguistic processing of them, producing more abstract mental representations at the expense of perceptually veridical ones (c.f., verbal overshadowing; Schooler & Engstler-Schooler, 1990). While verbal processing in sequential narratives could similarly change the memory representation, it would likely do so to support constructing a coherent mental model.

Overview of the current study

The present study assessed the role of both linguistic and visuospatial WM resources in the computation of bridging inferences during sequential narrative processing using a dual-task paradigm, similarly to Fincher-Kiefer and D'Agostino (2004) who explored the effects of verbal and visuospatial WM loads on inference generation during text comprehension (cf. Fincher-Kiefer, 2001). The authors had participants engage in verbal or visuospatial load tasks while reading simple narratives, and they used a lexical decision task to measure inferences activation. Only the visuospatial load impaired predictive inferences, and they concluded that bridging inferences do not require WM. However, a plethora of studies have shown the importance of WM resources in drawing inferences during text comprehension (Allbritton, 2004; Calvo, 2001, 2005; Just & Carpenter, 1992; Linderholm & van den Broek, 2002; Rai et al., 2011, 2014; Reichle & Mason, 2007; St. George et al., 1997). Moreover, as discussed in Experiment 2, we used arguably more demanding WM load tasks than those used in Fincher-Kiefer and D'Agostino (2004). Thus, it seems premature to conclude that bridging inferences do not require WM resources (linguistic or visuospatial).

The current study describes three experiments, in each of which participants viewed six wordless sequential narratives. Each story contained sets of three-picture narrative action sequences depicting (1) a beginning state, (2) a bridging event, and (3) an end state. For example, Fig. 2 shows the event of a boy falling into a pond while trying to catch a frog. Picture 1 shows the boy starting to run down the hill; picture 2 shows him tripping on a tree root; picture 3 shows his feet sticking out of the water. We chose sequences such that if the bridging-event picture (e.g., the boy tripping on the root) was missing, the bridging event could be readily inferred when viewing the end state (e.g., the boy's feet sticking out of the water). We then manipulated whether the bridging-event pictures were missing or not.

Experiment 1 was conducted to verify that viewers would infer the bridging event when the corresponding picture was missing, and that we could measure that using viewing times, similar to text reading times (Clark, 1977). Experiments 2 and 3 explored the types of WM resources used to generate bridging inferences while viewing sequential narratives. We did this by assessing the effects of concurrent WM loads on bridging inference generation as reflected by viewing times. Experiment 2 contrasted verbal and visuospatial loads. Experiment 3 assessed whether subvocal processes are important for inference generation by using concurrent articulatory suppression.

According to the *visuospatial primacy hypothesis*, perceptual systems are primarily involved in computing mental models for events (Zwaan, 2014). Thus, a visuospatial load

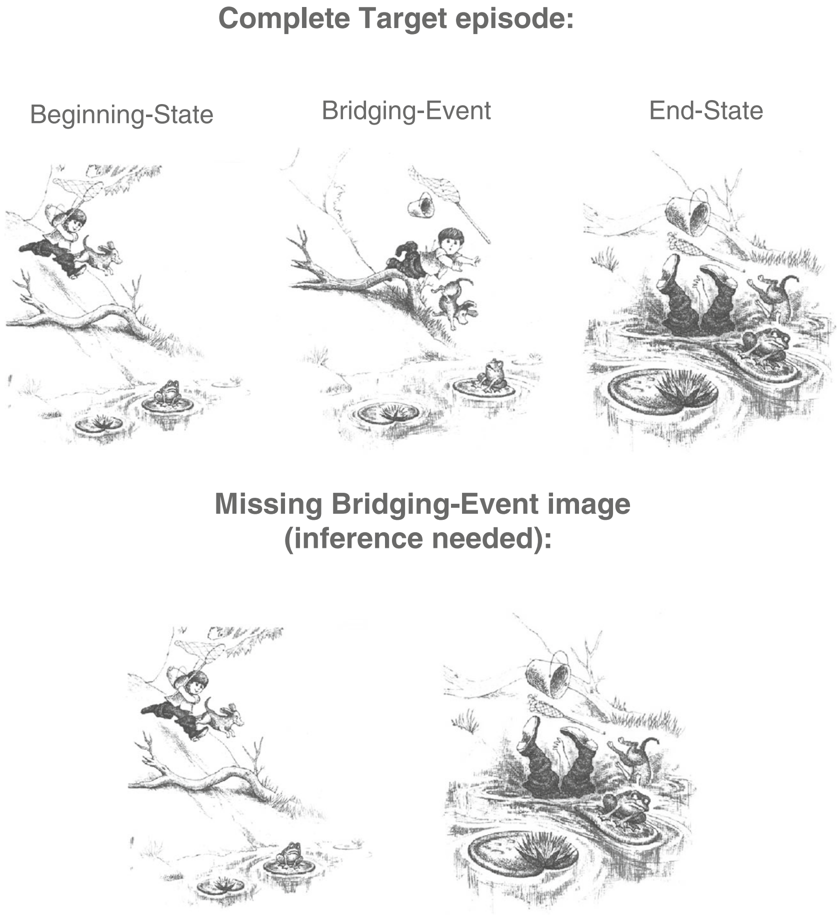

Fig. 2 Example episodes with bridging event missing and present

should impair bridging inference generation, but a verbal load should not. Alternatively, according to the *shared systems hypothesis*, both linguistic and perceptual WM resources are used to construct mental models of sequential narratives (Cohn, 2013a, 2014; Cohn et al., 2014), so both WM load types should impair bridging inference generation. Similarly, if subvocalization supports bridging inference generation, articulatory suppression, which prevents subvocalizing, should impair that. The shared systems hypothesis is consistent with dual code perspectives on information processing (Louwerse, 2008; Louwerse & Zwaan, 2009; Paivio, 1990; Schnotz, 2002; Schnotz & Bannert, 2003).

Experiment 1

The purpose of Experiment 1 was to demonstrate that picture viewing times are sensitive to natural inference processes, as is the case with sentence reading times (e.g., Clark, 1977). A recent study showed that panel viewing times were sensitive to inference process when viewers were cued that information was missing. Cohn and Wittenberg (2015) showed participants comic strips in which a critical panel was or was not preceded by a panel depicting a symbol (an action star), which signaled that an action had occurred that was not depicted. The critical panel was processed more slowly if it was preceded by the action symbol than when it was not, suggesting that the symbol cued participants to infer a missing action. This raises the question of whether such inferences are spontaneously generated without being explicitly cued that an action is missing. Experiment 1 did just that, using a viewing time paradigm in which participants viewed picture stories one picture at a time, both with and without missing bridging events, while their viewing times were recorded. If, when bridging events are missing, viewers generate bridging inferences while viewing the end-state images, then their end-state image viewing times should be longer when bridging events were missing than when present. Importantly, as a control, one would *not* expect such viewing time differences for the picture *immediately following* the end-state picture, and so we also analyzed viewing times for those.

Method

Participants Forty participants (female = 20) at Kansas State University participated in the study for course credit.

Materials We used six picture stories (ranging from 24–26 images each) from the *Boy, Dog, Frog* series (Mayer, 1967, 1969, 1973, 1974, 1975; Mayer & Mayer, 1971), with each of the six stories containing four target episodes, for a total of 24 target episodes. As shown in Fig. 2, target episodes were three-image sequences that contained (1) a beginning state, (2) a bridging event, and (3) an end state. A pilot study (described briefly below) validated that when the bridging events were missing, they were readily inferred. In each story, two target episodes showed the bridging-event action and two did not, and this was counterbalanced (described below). It is important to note that the original images were altered. Specifically, they contained a considerable amount of background detail, but the amount of detail varied from image to image. In an effort to control for the amount of details in the panels, much of the background details were removed so that the images focused on the characters and events depicted in them. Images were presented on a white background at a 1, 024 × 768 pixel resolution.

Design Experiment 1 used a 2 (bridging-event presence: present vs. absent) × 2 (end-state picture: end state vs. end state + 1) within-participants design. Viewing times on the target images were used as the dependent variable.

The assignment of the target episodes to the bridging-event-presence conditions was counterbalanced as follows. If we label "bridging-event present" as "A" and "bridging-event absent" as "B," then for each set of four episodes in a given story there are six possible orders: (1) AABB, (2) ABAB, (3) ABBA, (4) BBAA, (5) BABA, and (6) BAAB. The six orders of bridging-event presence were combined with six story-presentation orders (first–sixth) in a 6 × 6 Latin square, producing 36 order combinations. Each of these 36 orders was randomly assigned to a single subject. Thus, across 36 subjects, all six possible bridging-event-presence orders (described above) occurred equally often for each of the six stories, and all six stories were presented equally often as the first to sixth story in the experiment.

Pilot study We conducted a pilot study to verify whether participants would infer the missing bridging events. Thirty-six undergraduates (female = 19) at Kansas State University participated in the pilot study for course credit. Following the three-pronged method (Magliano & Graesser, 1991), participants were prompted to "think aloud" after each end-state picture (though participants were not informed of this systematic contingency). Participants were told they would see six picture stories and their task was to comprehend them. Participants advanced at their own pace through story images, one at a time, by pressing a button labeled "NEXT." After each end-state image, participants were prompted to "think aloud" by typing their understanding of the picture that they just saw into a text box that appeared on the screen, and then pressed the "end" key to view the next image. This was repeated for all six picture stories.

We predicted that participants would mention the bridging event (e.g., the boy tripped) in the bridging-event-*absent* condition more often than the bridging-event-present condition, because having to infer the action would more highly activate it in WM than simply seeing the action. For each bridging-event picture, we constructed a list of verb phrases describing the depicted action (e.g., tripped, fell). Two raters independently judged if each verbal protocol, for each end-state picture, contained any target phrases and produced acceptable interrater reliability (Cohen's kappa = .80). Thus, all discrepancies between the judges were resolved through discussion to produce the final coding of all protocols. As predicted, bridging events were more likely than not to be inferred in the bridging-event-*absent* condition (59 %) but not in the *present* condition (42 %), $\chi^2 = 25.69$, $p < .001$.

Procedure The task instructions and procedures for Experiment 1 were identical to those in the pilot study, with the following exceptions. Participants simply progressed through the images at their own pace until they reached the last image of the story. After viewing the last image, they were prompted to type a brief (3–4 sentence) summary of the story, which was done to motivate participants to comprehend the stories, and thus the summary protocols were not analyzed. Picture viewing times were recorded with ms accuracy using a Cedrus button box, as the time between picture onset and pressing the "NEXT" button.

Results and discussion

We cleaned the data using an initial criterion-based trimming procedure followed by normative trimming. The minimum acceptable viewing time for an image was set at 480 ms, which is sum of (1) the average scene viewing eye fixation duration (330 ms; Rayner, 1998) plus (2) a very short estimate of the minimum simple manual RT (150 ms; Teichner & Krebs, 1972). Because this was unlikely to provide sufficient time to comprehend a picture, viewing times below this minimum were excluded from the analyses, which constituted 20 observations (1.06 % of the 1,880 total observations). The maximum acceptable viewing time, 20 seconds, was based on examining the viewing time distribution across all conditions and experiments. In Experiment 1, no viewing times exceeded this maximum (however, some were found in later experiments). After that, viewing times >3 standard deviations above the means for each of the bridging-event-present and absent conditions were removed, which eliminate a total of 37 observations (1.9 %). Additionally, any participant for whom >25 % of their data was removed based on the above criteria had their data removed from the analyses in total. However, no participants exceeded this threshold in Experiment 1.

Data were analyzed using a 2 (image sequence: end state vs. end state + 1) × 2 (bridging-event presence: present vs. absent) within-subjects factorial ANOVA. In all experiments reported here, all effect sizes for simple main effects are reported using Cohen's *d*. Means and standard errors are shown in Fig. 3. Critically importantly, there were longer viewing times in the bridging-event-absent condition (M = 2,738, SE = 136) than the present condition (M = 2,556; SE = 115), $F(1, 39) = 6.28$, $MSE = 211{,}935.82$, $p = .01$, $\eta^2 = .14$. Just as importantly, there was a significant image sequence × bridging-event-presence interaction, $F(1, 39) = 13.39$, $MSE = 155{,}647.37$, $p < .001$, $\eta^2 = .22$. Specifically, as shown in Fig. 3, the longer viewing times for the bridging-event missing condition relative to the present condition was restricted to the end-state panel—end-state + 1 image viewing times did not differ between bridging-event-image-present versus absent conditions.

Experiment 1 and the pilot study exemplify the three-pronged method (Magliano & Graesser, 1991). That is, convergence between the verbal protocols (pilot study) and the picture viewing times strongly support the conclusion that viewing time differences between conditions reflect bridging inferences being generated in the bridging-event-absent condition. Furthermore, this experiment makes a methodological contribution by demonstrating that picture viewing times are sensitive to spontaneously occurring back-end mental model construction processes, beyond the assumed front-end perceptual processes necessary to process each picture (c.f., Cohn & Wittenberg, 2015).

Experiment 2

The purpose of Experiment 2 was to assess the extent to which linguistic and/or visuospatial WM systems support computing bridging inferences for sequential narratives, and we used a dual task paradigm to do so (c.f., Fincher-Kiefer & D'Agostino, 2004). We assumed that to the extent that a specific WM load used resources needed to generate sequential narrative inferences, it would preferentially attenuate the

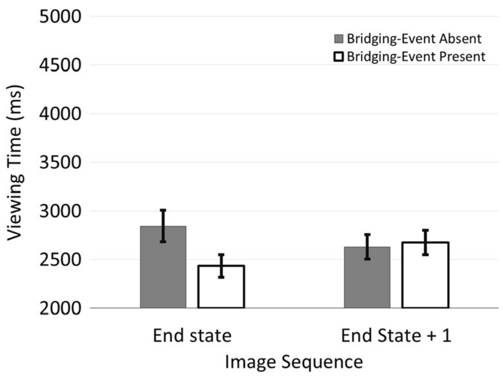

Fig. 3 Experiment 1: Mean viewing time as a function of Image (end-state image vs. end-state image + 1, i.e., the image after the end-state image) and bridging-event presence. Error bars represent the standard error of the mean

inference effect on viewing times shown in Experiment 1. Participants saw the same picture stories used in Experiment 1, but we varied the presence/absence of a verbal or visuospatial WM load task. The visuospatial primacy hypothesis predicted that only the visuospatial load would produce such attenuation of the inference effect, while the shared systems hypothesis predicted that both load tasks would do so.

Method

Participants One hundred fifty-eight subjects (female = 83) at Kansas State University participated in the study for course credit.

Materials The story materials were the same as in Experiment 1. We were concerned that the type of load tasks used by Fincher-Kiefer and D'Agostino (2004) did not consume sufficient WM resources to impede bridging inferences. Thus, in Experiment 2, we used potentially more demanding load tasks, requiring participants to remember seven-item random sequences of words or spatially presented dots. Participants had to remember the sequence over a short retention interval (the viewing time for the end-state image), and then recreate the sequence at test. Thus, WM loads were presented immediately prior to each target episode end-state image, and the WM recall was tested immediately following the end-state image. In this way, the WM load tasks were designed to selectively interfere with any WM demands associated with generating a bridging inference while viewing the end-state image.

The verbal WM seven-item word lists were generated by pseudo-randomly selecting from four color words: *red*, *blue*, *green*, and *ray*, with the constraint that no color word be used more than twice per sequence. The color words were presented at the white screen center, in black, 40-point, Times New Roman font. The verbal WM task began by presenting a fixation dot at the center of the screen for 1,250 ms, then a color word for 1,250 ms, and alternating fixation dot and color words until the seventh color word. After the retention interval, participants were prompted to make a memory response by the appearance of a 2 × 2 color word matrix at the screen center, and participants were asked to click the color words in the same order as the presentation sequence using a mouse.

The visuospatial WM 7-dot spatial sequences were generated similarly, by pseudo-randomly selecting dot locations from the four corners of the computer screen, with the constraint that no corner be used more than twice per sequence. As with the color word WM task, the visuospatial WM task began by presenting a (15 × 15 pixel) fixation dot at the center of the white screen for 1,250 ms, then the black dot reappeared in one of the four screen corners for 1,250 ms, and central and corner dots alternated until the seventh corner dot appeared. After the retention interval, participants were prompted to make a memory response by the appearance of 4 (100 × 100 pixel) gray squares, one in each screen corner, and participants were asked to click the gray squares in the same order as the dot sequence using a mouse.

Participants in the verbal and visuospatial no-load control conditions passively saw the same load task sequences (either color words or corner dots) but were instructed to ignore them.

Design Experiment 2 employed a 2 (bridging-event presence: present vs. absent) × 2 (Load presence: load present vs. no-load control) × 2 (load modality: verbal vs. visuospatial) mixed factorial design. As in Experiment 1, bridging-event presence was a within-participants variable, whereas load presence and load modality were between-participants variables. Viewing times for the end-state pictures was the dependent variable. The same bridging-event-presence condition and story-order counterbalancing scheme used in Experiment 1 was used in Experiment 2.

Procedure The procedure for Experiment 2 was the same as Experiment 1, with the following exceptions. Participants were randomly assigned to one of four different WM conditions. Those in the load conditions were instructed that they would be occasionally presented with the appropriate sequence (color words or dot locations) and that they would later be asked to recall them, using the appropriate test (as described above). Conversely, those in the no-load control conditions were explicitly told to ignore the color words or dots.

Prior to the experiment, participants were given practice with their assigned WM task. During the practice for both the verbal and visuospatial load tasks, the WM load increased sequentially from a four-item load to a seven-item load, with two practice trials at each load level. The final seven-item WM practice trials were at the same load level as the experiment. Pilot testing had shown that the seven-item verbal and visuospatial loads were equally taxing. After the WM practice, participants were told they would view six picture stories that they were to comprehend, as shown by writing a short summary of each story at its end, and that at various points in the story they would be asked to do their respective WM task. All other procedures were the same as in Experiment 1.

Results and discussion

Data trimming Prior to analysis, the data were cleaned using the same criterion-based and normative-based trimming rules used in Experiment 1. The criterion-based trimming removed a total of 59 observations (1.5 % of the total). The normative-based trimming removed 68 observations (1.8 %). No participants were removed based on trimming >25 % of their observations. However, one participant's data were removed due to task noncompliance.

WM performance Chance performance in the WM tasks was one in four, thus, WM performance for each condition was compared to chance performance (25 %) using one-sample *t* tests. Performance for the verbal WM group (*M* = 66.6, *SE* = 2.1) and the visuospatial WM group (*M* = 54.6, *SE* = 1.5) were both well above chance—verbal WM: *t*(39) = 19.78, *p* < .001, Cohen's *d* = 3.13; visuospatial WM: *t*(38) = 19.43, *p* < .001, Cohen's *d* = 3.07—indicating that each group performed its respective WM task.

Inference effect on viewing times A 2 (bridging-event presence: present vs. absent) × 2 (load presence: load-present vs. no-load control) × 2 (load modality: verbal vs. visuospatial) mixed factorial ANOVA was conducted on the end-state picture viewing times (see Fig. 4). As in Experiment 1, there was a main effect of bridging-event presence, with end-state image viewing times longer when the bridging-event images were absent (*M* = 3,796 ms, *SE* = 110) than when present (*M* = 3, 292 ms, *SE* = 91), *F*(1, 153) = 77.27, *MSE* = 258,374.17, *p* < .001, η^2 =.34 . There was also a main effect of load presence, such that viewing times were shorter with a load (*M* = 3, 108 ms, *SE* = 114 ms) than with no-load (*M* = 3,830 ms, *SE* = 114 ms), *F*(1, 153) = 10.26, *MSE* = 2,929,432.70, *p* = .002, η^2 = .12. Critically importantly, however, these two main effects were qualified by a significant load presence × bridging-event presence interaction, *F*(1, 153) = 4.85, *MSE* = 258, 374.17, *p* = .029, η^2 = .031. Neither the main effect of load type, nor any interactions involving load type were significant (both *p*s < .20). Thus, in further analyses, we collapsed across load type, as shown in Fig. 5. The interaction between bridging event presence and load presence appears to be a magnitude interaction. Thus, effect sizes for the simple effects were computed to explore this possibility. Figure 5 shows that the inference effect (i.e., effect of absent bridging event) was reduced for the load conditions (Δ = 370 ms, Cohen's *d* = .29) relative to the no-load control conditions (Δ = 638 ms, Cohen's *d* = .49). Together, the results of Experiment 2 are consistent with the shared systems hypothesis.

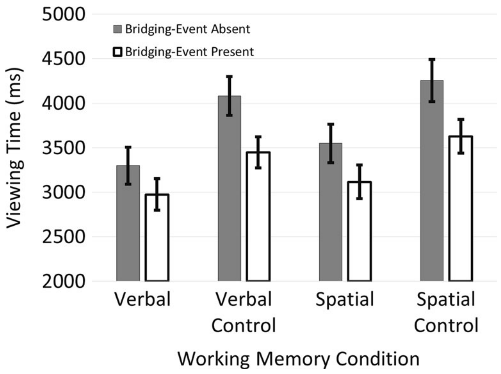

Fig. 4 Experiment 2: Mean viewing time for the end-state image as a function of the WM task and bridging-event presence. The error bars represent the standard error of the mean

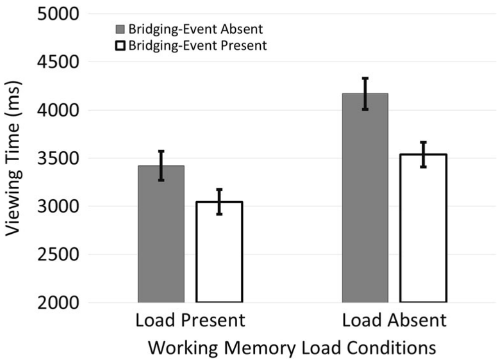

Fig. 5 Experiment 2: Mean viewing time for the end-state image as a function of WM load presence and bridging-event presence (averaged across load type)

Experiment 3

Experiment 2 suggested that both visuospatial and linguistic WM resources support bridging inference generation. Regarding linguistic WM resources, this support could be limited to the activation, storage, and manipulation of linguistic knowledge, or it could also involve subvocalization of the inferences. For example, viewers sometimes spontaneously activate the names of visually presented items (e.g., Meyer et al., 2007), and subvocalization supports constructing abstract memory representations for complex images (Brandimonte et al., 1992a, b). Thus, subvocalization could also support generating bridging inferences for sequential narratives. We conducted Experiment 3 to test this hypothesis using an articulatory suppression (AS) task (e.g., Baddeley, Thomson, & Buchanan, 1975) in which participants repeated a monosyllabic word (e.g., *the*) while viewing the picture stories. As a standard control, we also included a simple repetitive motor response condition that lacked any linguistic component (mouse clicking). The *subvocalization support hypothesis* predicts AS will attenuate the inference effect. Conversely, if subvocalization plays no role in generating such inferences, AS should not attenuate the inference effect.

Method

Participants One hundred fifty-four undergraduate students at Northern Illinois University (female = 76) participated for course credit.

Materials We used the same sequential narratives used in Experiments 1 and 2.

Design A 2 (bridging-event presence: present vs. absent) × 3 (suppression task: AS, clicking, no-concurrent task) mixed factorial design was used with bridging-event presence as a within-subjects factor and suppression condition as a between-subjects factor.

Procedure The procedures for Experiment 3 were identical to those of Experiment 2, with the following exceptions. Participants were randomly assigned to one of three suppression task conditions: AS, mouse clicking, or no-concurrent task. The no-concurrent task condition was therefore identical to the viewing conditions of Experiment 1. Prior to the experiment, participants in the clicking and AS conditions were given additional instructions. AS condition participants were instructed to repeat the word *the* out loud continuously, at a fast and consistent pace of articulation, while looking at the picture stories but not when writing their summaries after each story. Then, participants would see a reminder to begin AS again before the next story, and begin AS immediately on seeing the reminder, before beginning the next story. The pace of AS articulation was demonstrated by the experimenter, and participants briefly practiced the AS task and received feedback.

Verbal instructions for the clicking condition were the same as for the AS condition with the following exceptions. Participants were told they would be clicking their right mouse button while looking at the story pictures (with accommodations offered for left-handed participants). Participants in the no-concurrent task control condition were given no special instructions, practice, or secondary task reminders.

We assessed participants' suppression task compliance by recording their behavior, using audio-recording for the AS condition and mouse-clicking rate captured by Experiment Builder for the clicking condition. Compliance with secondary task instructions was assessed on an ongoing basis throughout data collection, and data from participants who did not maintain a regular and sufficient rate of articulation or clicking throughout the stories were removed from the experiment, with replacement participants obtained in the next data collection session. Six participants in the AS condition were excluded for noncompliance in one or more stories, and another participant was removed for failing to maintain a consistent rate of articulation (i.e., frequent pauses between articulations >2 seconds). Nine participants in the clicking condition were excluded for noncompliance in one or more stories.

Results and discussion

Data from 11 participants were lost due to computer error. The data were cleaned using the same procedures as in Experiments 1 and 2 (an initial criterion-based trimming procedure followed by normative trimming). A total of 66 observations (2.2 % of the original 3,024) were excluded due to insufficient viewing times (i.e., <480 ms). As in the previous experiments, a 20,000 millisecond (20 s) maximum viewing time was used, and this resulted in excluding five observations (0.17 %). Normative trimming eliminated a total of 64 (2.1 %) of the observations from the analysis. Additionally, four participants were removed because >25 % of their observations were excluded based on the criteria described above. This exclusion removed an additional 49 observations (1.6 %). Thus, in total, data from 122 of the 154 participants were used in the analyses, and 6.08 % of observations were removed.

A 2 (bridging-event presence: present vs. absent) × 3 (suppression task: AS, clicking, no-concurrent task) mixed ANOVA was conducted on the end-state image viewing times. As can be seen in Fig. 6, we replicated the inference effect found in Experiments 1 and 2, with longer viewing items on end-state images when the bridging-event images were absent (M = 2,710 ms, SE = 104) than when present (M = 2,353 ms, SE = 88), $F(1, 119) = 59.84$, MSE = 129,705.42, $p < .001$, η2 = .33 . There was no significant main effect of suppression condition, $F(1, 119) = 1.34$, MSE = 2,144,597.88, $p = .266$, nor was the interaction between suppression and bridging-event presence significant, $F(1, 119) = 1.88$, MSE= 129,705.42, $p = .158$. Thus, the results of Experiment 3 were inconsistent with the sub-vocalization support hypothesis, suggesting that sub-vocalization plays little or no role in generating the bridging inferences that were constructed in the bridging-event-absent condition.

General discussion

The primary goal of this study was to assess the extent to which visuospatial and linguistic WM systems support the construction of bridging inferences for sequential narratives. A secondary goal was to verify that picture viewing times

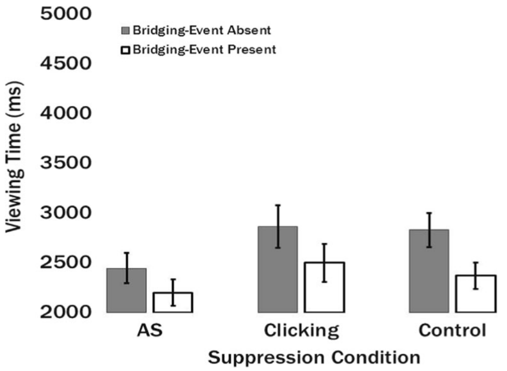

Fig. 6 Experiment 3: Mean viewing time for the end-state image as a function of the suppression task and bridging-event presence (AS = articulatory suppression). The error bars represent the standard error of the mean

were sensitive to spontaneously generated bridging inferences. While there have been studies demonstrating that viewers construct mental models while viewing films (e.g., Magliano, Dijkstra, & Zwaan, 1996; Magliano, Miller, & Zwaan, 2001; Magliano, Taylor, & Kim, 2005; Zacks, Speer, & Reynolds, 2009), relatively little research has investigated mental model construction while comprehending static image sequential narratives (cf. Cohn, 2013b; Cohn et al., 2012, 2014; Cohn & Paczynski, 2013; Gernsbacher, 1985; Magliano, Kopp, McNerny, Radvansky, & Zacks, 2011; Nakazawa, 2005; Saraceni, 2001, 2003; West & Holcomb, 2002). Importantly, Experiment 1 demonstrated that picture viewing times, like reading times (e.g., Clark, 1977), are sensitive to processes involved in mental model construction, consistent with the findings of Cohn and Wittenberg (2015), but here involving spontaneously generated inferences. These findings suggest that picture viewing times provide a simple, robust, unintrusive, natural, and implicit measure of moment-to-moment processing of sequential narratives, analogous to sentence reading times and eye movements in the context of comprehending narrative text (e.g., Rayner, Raney, & Pollatsek, 1995).

The most significant contribution of the present study was its exploration of roles of visuospatial and linguistic WM in generating bridging inferences during sequential narrative comprehension. Experiment 2 provided clear support for the shared systems hypothesis in that both visuospatial and verbal loads attenuated the inference effect (i.e., longer viewing times in the bridging-event-*absent* rather than *present* condition). This is consistent with a general assumption of SPECT (Loschky et al., 2014, 2015), namely that front-end perceptual processes involved in encoding, here visuospatial relations, are important for back-end processes in memory involved in mental model construction, here bridging inference generation. Exactly how visuospatial WM processes supported bridging inference generation cannot be determined from this study. However, we speculate that the visuospatial mapping of entities across pictures using WM may signal when a bridging inference is necessary. Either a high or a low degree of overlap in entities' spatial arrangement across pictures would suggest no bridging inference is needed. High overlap would indicate no gap needing an inference to fill it (e.g., "moment-to-moment" panel-to-panel transitions in comics; McCloud, 1993). Repetition of agents and objects helps perceiving continuity of events (Saraceni, 2001). Low overlap would often indicate a transition to a new story episode (e.g., "scene-to-scene" transitions; McCloud, 1993), leading to shifting to build a new event model (Gernsbacher, 1990; Zacks & Tversky, 2001; Zacks, Tversky, & Iyer, 2001), which is beyond the scope of bridging inferences. Thus, we propose that bridging inference generation is triggered when visuospatial processes detect moderate perceptual feature overlap (e.g., entities, spatial–temporal locations) and the depiction of different events (e.g., Boy running down the hill to catch the frog, Boy in the water, similar to McCloud's 1993, "action-to-action" transitions; see also Saraceni, 2001). This testable hypothesis warrants further study.

The fact that both load conditions attenuated bridging inferences is consistent with VLT's claim that language systems support sequential narrative comprehension (Cohn, 2013a, 2014), but how? Again, the current study was not designed to answer this question. However, we propose that an important aspect of both the visuospatial and verbal load tasks in Experiment 2 was that participants had to maintain sequences in WM. Our results are consistent with the claim that general systems that process hierarchically structured sequencing in language also support processing of hierarchically structured sequences of images in sequential narratives (Cohn, 2013a, 2014).

We should mention an alternative methodologically based explanation for the effect of WM loads on the inference effect on viewing times in Experiment 2. Specifically, one might try to explain the decreased inference effect as simply due to speeding up viewing times on the outcome image, allowing viewers to more quickly respond to the WM task, to improve their performance. However, there is a key problem for this alternative explanation. Specifically, this alternative explanation would predict that outcome image viewing times should be faster with a WM load than without one, yet a comparison of Fig. 3 (Experiment 1, no load) and Fig. 5 (Experiment 2, load) shows that this is clearly *not* the case. More specifically, the average of both WM loads of Experiment 2 *increased* viewing times by roughly 600 ms compared to those in Experiment 1, which lacked any secondary load (neither due to performing the WM task itself nor due to passively viewing the load stimuli). Thus, this alternative explanation of the effect of WM load on the inference effect is clearly *not* supported by our data.

Nevertheless, it is surprising that passively viewing the load stimuli (load controls in Experiment 2) increased viewing times on end-state images relative to when no-load stimuli were presented (Experiment 1). A simple explanation for this increase in viewing time is that, in Experiment 1, the target sequence pictures were presented consecutively, whereas, in the Experiment 2 no-load conditions, they were separated by the passively viewed WM load stimuli. Apparently, increasing the temporal distance between pictures increased overall processing time but, importantly, without dramatically attenuating the inference effect in the Experiment 2 no-load control conditions compared to Experiment 1. Specifically, the size of the inference effect in Experiment 1 (Cohen's d = .60) is slightly larger than in the no-load conditions in Experiment 2 (Cohen's d = .49). Conversely, the Experiment 2 load conditions substantially lowered the inference effect size (Cohen's d = .29) compared to both of the above-mentioned conditions, which is consistent with the shared systems hypothesis.

The current results are in contradiction to Fincher-Kiefer and D'Agostino (2004), who provided evidence that WM resources (visuospatial or verbal) are *not* needed to support generating bridging inferences using concurrent load tasks somewhat similar to those of Experiment 2. However, as noted earlier, Fincher-Kiefer and D'Agostino's (2004) load tasks did not require remembering sequences. If sequence processing supports bridging inference generation, then a load task involving sequencing should be more disruptive than one that does not. This possibility could be tested in further research.

The results of Experiment 3 suggest linguistic support for bridging inferences in seqeuntial narratives does not involve subvocalization, at least wordless seqeuntial narratives (i.e., no dialogue). At first glance, the current results seem to contradict those of Brandimonte et al. (1992a, b), which suggested subvocalization was involved in processing static images. However, their materials were ambiguous geometrical figures, which participants were asked to learn, thus the task constraints were very different from the current study. Furthermore, the current results are consistent with ERP studies, suggesting that seqeuntial narrative processing does not require subvocal mediation (West & Holcomb, 2002; Cohn et al., 2012). Subvocalization may occur during certain sequential narrative comprehension processes, such as viewers' overt or subvocal responses to suspenseful movies (e.g., "No! Don't open that door!"; Bezdek, Foy, & Gerrig, 2013). However, generating such (predictive) inferences would logically seem to occur prior to producing (sub)vocal viewer responses based on those inferences, as shown in studies of speech production based on comprehending pictorial stimuli (Griffin & Bock, 2000). Thus, eliminating any such (sub)vocal viewer responses may have no effect on the prior processes involved in inference generation, as shown by Experiment 3.

While it has been argued that mental model construction is similar across input modalities (Gernsbacher, 1990; Gernsbacher et al., 1990; Kintsch, 1998; Magliano et al., 2007, 2013), this has rarely been directly tested (e.g., Bagget, 1979; Magliano et al., 2011). But, models of text comprehension may help explain the results of the current study. The construction-integration (CI) model (Kintsch, 1988, 1998) provides an account of inference generation based on an iterative, two-stage process. During an *activation stage*, input associated with a story constituent (here, some part of a picture, e.g., an agent, its location, or the agent's action) provides retrieval cues leading to the unconstrained activation of knowledge from both the discourse (or pictorial) representation and world knowledge (see also Myers & O'Brien, 1998; McKoon & Ratcliff, 1998). During a subsequent *integration phase*, relevant knowledge remains highly activated while irrelevant knowledge is deactivated via constraint satisfaction guided by spreading activation. According to the CI model, bridging inferences are based on knowledge that remains activated after the integration phase is completed.

The CI model (Kintsch, 1988, 1998) suggests an explanation for the finding in the current study that viewing times were longer when the bridging events were missing than when they were present. Specifically, a longer time would be required for the constraint satisfaction cycles to settle during the integration phase when the bridging-event actions were missing, due to a greater conceptual distance between the content of the beginning-state and end-state pictures than when they were present. If subsequent research to test this hypothesis provides evidence consistent with it, then specialized models of seqeuntial narrative comprehension, such as VLT (Cohn, 2013a, 2014) and SPECT (Loschky et al., 2014, 2015) should also account for comprehension processes delineated by theories of text comprehension. This is an underlying assumption of SPECT.

In sum, the current study is part of a growing body of research on the comprehension of seqeuntial narratives in the context of film (Bagget, 1979; Magliano et al., 1996, 2001, 2005; Schwan & Ildirar, 2010; Zacks et al., 2009) and static seqeuntial narratives (Cohn, 2013a, 2014; Cohn et al., 2012, 2014; Gernsbacher et al., 1990; Magliano et al., 2011). The current study contributes to this literature by showing that both visuospatial and verbal WM resources support the comprehension of wordless seqeuntial narratives, specifically bridging inference generation, and confirms that picture viewing times are sensitive to mental model construction processes. Exactly how these visuospatial and verbal systems support comprehension of seqeuntial narratives demands further investigation. Such research will be important both for advancing theories of seqeuntial narrative comprehension (Cohn, 2013a, 2014; Loschky et al., 2014, 2015) and exploring the generalizability of theories of text comprehension to non-text-based narratives (Gernsbacher, 1990; Loschky et al., 2014, 2015; Zacks et al. 2001, 2009). Such theories are of fundamental importance in the study of cognition, because they aim to explain comprehension across multiple sensory and representational modalities, in an attempt to capture the essential mechanisms of human understanding.

Author Notes Conceived of experiments: J. M. L., L. Designed experiments: J. M., L. L., A. L., K. H. Performed experiments: A.L., K. H. Analyzed data: A. L., K. H. Wrote the paper: J. M., L. L., A. L., K. H. This research was partially supported by the Center for the Interdisciplinary Study of Language and Literacy at Northern Illinois University and by a grant awarded to the first author by the National Science Foundation (Award no. 1319974). All opinions are those of the authors and not the funding agencies.

References

Allbritton, D. (2004). Strategic production of predictive inferences during comprehension. *Discourse Processes, 38*(3), 309–322.

Baddeley, A. D. (2001). Is working memory still working? *The American Psychologist, 56*(11), 851–864.

Baddeley, A. D., Thomson, N., & Buchanan, M. (1975). Word length and the structure of short-term memory. *Journal of Verbal Learning and Verbal Behavior, 14*(6), 575–589.

Baggett, P. (1975). Memory for explicit and implicit information in picture stories. *Journal of Verbal Learning and Verbal Behavior, 14*(5), 538–548. doi:10.1016/s0022-5371(75)80031-4

Baggett, P. (1979). Structurally equivalent stories in movie and text and the effect of the medium on recall. *Journal of Verbal Learning and Verbal Behavior, 18*(3), 333–356.

Bezdek, M. A., Foy, J. E., & Gerrig, R. J. (2013). "Run for it!": Viewers' participatory responses to film narratives. *Psychology of Aesthetics, Creativity, and the Arts, 7*(4), 409–416.

Brandimonte, M. A., Hitch, G. J., & Bishop, D. V. M. (1992a). Influence of short-term memory codes on visual image processing: Evidence from image transformation tasks. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 18*(1), 157–165.

Brandimonte, M. A., Hitch, G. J., & Bishop, D. V. M. (1992b). Verbal recoding of visual stimuli impairs mental image transformations. *Memory & Cognition, 20*(4), 449–455.

Calvo, M. G. (2001). Working memory and inferences: Evidence from eye fixations during reading. *Memory, 9*(4/6), 365–381.

Calvo, M. G. (2005). Relative contribution of vocabulary knowledge and working memory span to elaborative inferences in reading. *Learning and Individual Differences, 15*(1), 53–65.

Clark, H. H. (1977). Inferences in comprehension. In D. LaBerge & S. J. Samuels (Eds.), *Basic processes in reading: Perception and comprehension* (pp. 243–263). Hillsdale, NJ: Erlbaum.

Cocchini, G., Logie, R. H., Della Sala, S., MacPherson, S. E., & Baddeley, A. D. (2002). Concurrent performance of two memory tasks: Evidence for domain-specific working memory systems. *Memory & Cognition, 30*(7), 1086–1095.

Cohn, N. (2013a). *The visual language of comics: Introduction to the structure and cognition of sequential images*. London, England: Bloomsbury.

Cohn, N. (2013b). Visual narrative structure. *Cognitive Science, 37*(3), 413–452.

Cohn, N. (2014). You're a good structure Charlie Brown: The distribution of narrative categories in comic strips. *Cognitive Science, 38*(7), 1317–1359. doi:10.1111/cogs.1216

Cohn, N., Jackendoff, R., Holcomb, P. J., & Kuperberg, G. R. (2014). The grammar of visual narrative: Neural evidence for constituent structure in sequential image comprehension. *Neuropsychologia, 64,* 63–70. doi:10.1016/j.nueropsychologia.2014.018

Cohn, N., & Paczynski, M. (2013). Prediction, events, and the advantage of agents: The processing of semantic roles in visual narrative. *Cognitive Psychology, 67*(3), 73–97.

Cohn, N., Paczynski, M., Jackendoff, R., Holcomb, P. J., & Kuperberg, G. R. (2012). (Pea) nuts and bolts of visual narrative: Structure and meaning in sequential image comprehension. *Cognitive Psychology, 65*(1), 1–38.

Cohn, N., & Wittenberg, E. (2015). Action starring narratives and events: Structure and inference in visual narrative comprehension. *Journal of Cognitive Psychology.* doi:10.1080/20445911.2015.1051535

Darke, S. (1988). Effects of anxiety on inferential reasoning task performance. *Journal of Personality and Social Psychology, 55*(3), 499–505.

Estevez, A., & Calvo, M. G. (2000). Working memory capacity and time course of predictive inferences. *Memory, 8*(1), 51–61.

Fincher-Kiefer, R. (2001). Perceptual components of situation models. *Memory & Cognition, 29*(2), 336–343.

Fincher-Kiefer, R., & D'Agostino, P. R. (2004). The role of visuospatial resources in generating predictive and bridging inferences. *Discourse Processes, 37*(3), 205–224.

Gernsbacher, M. A. (1985). Surface information loss in comprehension. *Cognitive Psychology, 17*(3), 324–363. doi:10.1016/0010-0285(85)90012-x

Gernsbacher, M. A. (1990). *Language comprehension as structure building* (Vol. 11). Hillsdale, NJ: Erlbaum.

Gernsbacher, M. A., Varner, K. R., & Faust, M. E. (1990). Investigating differences in general comprehension skill. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 16*(3), 430–445.

Graesser, A. C., Singer, M., & Trabasso, T. (1994). Constructing inferences during narrative text comprehension. *Psychological Review, 101*(3), 371–395.

Griffin, Z. M., & Bock, K. (2000). What the eyes say about speaking. *Psychological Science, 11*(4), 274–279.

Gustavson, P. (1939). *The fantom of the fair.* New York, NY: Centaur Publications. Retrieved from http://goldenagecomics.co.uk/

Herman, L. M., & Austad, S. N. (1996). Knowledge acquisition and asymmetry between language comprehension and production: Dolphins and apes as general models for animals. In C. Allen & D. Jamison (Eds.), *Readings in animal cognition* (pp. 289–306). Cambridge, MA: MIT Press.

Hitch, G. J., Brandimonte, M. A., & Walker, P. (1995). Two types of representation in visual memory: Evidence from the effects of stimulus contrast on image combination. *Memory & Cognition, 23*(2), 147–154.

Just, M. A., & Carpenter, P. A. (1992). A capacity theory of comprehension: Individual differences in working memory. *Psychological Review, 99*(1), 122–149.

Kintsch, W. (1988). The role of knowledge in discourse comprehension: A construction–integration model. *Psychological Review, 95*(2), 163–182.

Kintsch, W. (1998). *Comprehension: A paradigm for cognition*. New York, NY: Cambridge University Press.

Larson, A. M., & Loschky, L. C. (2015). *From scene perception to event conception: How scene gist informs event perception.* Manuscript submitted for publication.

Linderholm, T., & van den Broek, P. (2002). The effects of reading purpose and working memory capacity on the processing of expository text. *Journal of Educational Psychology, 94*(4), 778–784.

Loschky, L. C., Hutson, J., Magliano, J. P., Larson, A. M., & Smith, T. (2014, June). *Explaining the film comprehension/attention relationship with the scene perception and event comprehension theory (SPECT)*. Paper presented at the annual conference of the Society for Cognitive Studies of the Moving Image, Lancaster, PA.

Louwerse, M. M. (2008). Embodied relations are encoded in language. *Psychonomic Bulletin and Review, 15*(4), 838–844.

Louwerse, M. M., & Zwaan, R. A. (2009). Language encodes geographical information. *Cognitive Science, 33*(1), 51–73.

Magliano, J. P., Baggett, W. B., Johnson, B. K., & Graesser, A. C. (1993). The time course in which causal antecedent and causal consequence inferences are generated. *Discourse Processes, 16*, 35–53.

Magliano, J. P., Dijkstra, K., & Zwaan, R. A. (1996). Generating predictive inferences while viewing a movie. *Discourse Processes, 22*(3), 199–224.

Magliano, J. P., & Graesser, A. C. (1991). A three-pronged method for studying inference generation in literary text. *Poetics, 20*(3), 193–232.

Magliano, J. P., Kopp, K., McNerney, M. W., Radvansky, G. A., & Zacks, J. M. (2011). Aging and perceived event structure as a function of modality. *Aging, Neuropsychology, and Cognition, 19*(1/2), 264–282.

Magliano, J. P., Loschky, L., Clinton, J., & Larson, A. (2013). Differences and similarities in processing narratives across textual and visual media. In B. Miller, L. Cutting, & P. McCardle (Eds.), *Unraveling the behavioral, neurobiological, and genetic components of reading comprehension* (pp. 78–90). Baltimore, MD: Brookes.

Magliano, J. P., Miller, J., & Zwaan, R. A. (2001). Indexing space and time in film understanding. *Applied Cognitive Psychology, 15*(5), 533–545.

Magliano, J. P., Radvansky, G. A., & Copeland, D. E. (2007). Beyond language comprehension: Situation models as a form or autobiographical memory. In F. Schmalhofer & C. Perfetti (Eds.), *Higher level language processes in the brain: Inference and comprehension processes* (pp. 379–391). Mahwah, NJ: Erlbaum.

Magliano, J. P., Taylor, H. A., & Kim, H. J. J. (2005). When goals collide: Monitoring the goals of multiple characters. *Memory & Cognition, 33*(8), 1357–1367.

Mayer, M. (1967). *A boy, a dog and a frog*. New York, NY: Dial Press.

Mayer, M. (1969). *Frog, where are you?* New York, NY: Dial Press.

Mayer, M. (1973). *Frog on his own*. New York, NY: Dial Press.

Mayer, M. (1974). *Frog goes to dinner*. New York, NY: Dial Press.

Mayer, M. (1975). *One frog too many*. New York, NY: Dial Press.

Mayer, M., & Mayer, M. (1971). *A boy, a dog, a frog and a friend*. New York, NY: Dial Press.

McCloud, S. (1993). *Understanding comics: The invisible art*. New York, NY: Harper Perennial.

McKoon, G., & Ratcliff, R. (1998). Memory-based language processing: Psycholinguistic research in the 1990s. *Annual Review of Psychology, 49*(1), 25–42.

McNamara, D. S., & Magliano, J. P. (2009). Towards a comprehensive model of comprehension. In B. Ross (Ed.), *The psychology of learning and motivation* (Vol. 51, pp. 297–384). New York, NY: Elsevier Science.

Meyer, A. S., Belke, E., Telling, A. L., & Humphreys, G. W. (2007). Early activation of object names in visual search. *Psychonomics Bulletin & Review, 14*(4), 710–716.

Myers, J. L., & O'Brien, E. J. (1998). Accessing the discourse representation during reading. *Discourse Processes, 26*(2/3), 131–157.

Nakazawa, J. (2005). Development of manga (comic book) literacy in children. In D. W. Shwalb, J. Nakazawa, & B. J. Shwalb (Eds.), *Applied developmental psychology: Theory, practice, and research from Japan* (pp. 23–42). Greenwich, CT: Information Age.

Paivio, A. (1990). *Mental representations: A dual coding approach*. New York, NY: Oxford University Press.

Rai, M. K., Loschky, L. C., Harris, R. J., Peck, N. R., & Cook, L. G. (2011). Effects of stress and working memory capacity on foreign language readers' inferential processing during comprehension. *Language Learning, 61*(1), 187–218. doi:10.1111/j.1467-9922.2010.00592.x

Rai, M. K., Loschky, L. C., & Harris, R. J. (2014). The effects of stress on reading: A comparison of first language versus intermediate second-language reading comprehension. *Journal of Educational Psychology, 107*(2), 348–363. doi:10.1037/a0037591

Rayner, K. (1998). Eye movements in reading and information processing: 20 years of research. *Psychological Bulletin, 124*(3), 372–422.

Rayner, K., Raney, G. E., & Pollatsek, A. (1995). Eye movements and discourse processing. In R. F. Lorch & E. J. O'Brien (Eds.), *Sources of coherence in reading* (Vol. 14, pp. 9–35). Hillsdale, NJ: Erlbaum.

Reichle, E. D., & Mason, R. A. (2007). The neural signatures of causal inferences: A preliminary computational account of brain-imaging and behavioral data. In F. Schmalhofer & C. A. Perfetti (Eds.), *Higher level language processes in the brain: Inference and comprehension processes* (pp. 211–232). Mahwah, NJ: Erlbaum.

Saraceni, M. (2001). Relatedness: Aspects of textual connectivity in comics. In J. Baetens (Ed.), *The graphic novel* (pp. 167–179). Leuven, Belgium: Leuven University Press.

Saraceni, M. (2003). *The language of comics*. London, England: Routledge.

Schnotz, W. (2002). Towards an integrated view of learning from text and visual displays. *Educational Psychology Review, 14*(2), 101–120.

Schnotz, W., & Bannert, M. (2003). Construction and interference in learning from multiple representations. *Learning and Instruction, 13*(2), 141–156. doi:10.1016/S0959-4752(02)00017-8

Schooler, J. W., & Engstler-Schooler, T. Y. (1990). Verbal overshadowing of visual memories: Some things are better left unsaid. *Cognitive Psychology, 22*(1), 36–71.

Schwan, S., & Ildirar, S. (2010). Watching film for the first time: How adult viewers interpret perceptual discontinuities in film. *Psychological Science, 21*(7), 970–976.

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*(1), 4–27.

Singer, M., & Halldorson, M. (1996). Constructing and validating motive bridging inferences. *Cognitive Psychology, 30*(1), 1–38.

Smith, E. E., & Jonides, J. (1997). Working memory: A view from neuroimaging. *Cognitive Psychology, 33*(1), 5–42.

Smith, E. E., Jonides, J., & Koeppe, R. A. (1996). Dissociating verbal and spatial working memory using PET. *Cerebral Cortex, 6*(1), 11–20.

Speer, N. K., Reynolds, J. R., Swallow, K. M., & Zacks, J. M. (2009). Reading stories activates neural representations of visual and motor experiences. *Psychological Science, 20*(8), 989–999.

St. George, M. S., Mannes, S., & Hoffman, J. E. (1997). Individual differences in inference generation: An ERP analysis. *Journal of Cognitive Neuroscience, 9*(6), 776–787.

Teichner, W. H., & Krebs, M. J. (1972). Laws of the simple visual reaction time. *Psychological Review, 79*(4), 344–358.

Trabasso, T., & Nickels, M. (1992). The development of goal plans of action in the narration of a picture story. *Discourse Processes, 15*(3), 249–275.

West, W. C., & Holcomb, P. J. (2002). Event related potential during discourse-level semantic integration of complex picture. *Cognitive Brain Research, 13*, 363–375.

Zacks, J. M., Speer, N. K., & Reynolds, J. R. (2009). Segmentation in reading and film comprehension. *Journal of Experimental Psychology: General, 138*(2), 307–327. doi:10.1037/a0015305

Zacks, J. M., & Tversky, B. (2001). Event structure in perception and conception. *Psychological Bulletin, 127*(1), 3–21.

Zacks, J. M., Tversky, B., & Iyer, G. (2001). Perceiving, remembering, and communicating structure in events. *Journal of Experimental Psychology: General, 130*(1), 29–58.

Zwaan, R. A. (2014). Embodiment and language comprehension: Reframing the discussion. *Trends in Cognitive Sciences, 18*(5), 229–234.

Zwaan, R. A., & Radvansky, G. A. (1998). Situation models in language comprehension and memory. *Psychological Bulletin, 123*(2), 162–185.