Mem Cogn (2012) 40:812–826
DOI 10.3758/s13421-011-0179-8

Starting from scratch and building brick by brick in comprehension

Christopher A. Kurby · Jeffrey M. Zacks

Published online: 27 January 2012

Abstract During narrative comprehension, readers construct representations of the situation described by a text, called *situation models*. Theories of situation model construction and event comprehension posit two distinct types of situation model updating: incremental updating of individual situational dimensions, and global updates in which an old model is abandoned and a new one created. No research to date has directly tested whether readers update their situation models incrementally, globally, or both. We investigated whether both incremental and global updating occur during narrative comprehension. Participants typed what they were thinking while reading an extended narrative, and then segmented the narrative into meaningful events. Each typed think-aloud response was coded for whether it mentioned characters, objects, space, time, goals, or causes. There was evidence for both incremental and global updating: Readers mentioned situation dimensions more when those dimensions changed, controlling for the onset of a new event. Readers also mentioned situation dimensions more at points when a new event began than during event middles, controlling for the presence of situational change. These results support theories that claim that readers engage in both incremental and global updating during extended narrative comprehension.

Keywords Discourse processing · Situation models · Event segmentation · Updating

Stories are composed of a series of events, and readers perceive them as such. For example, a reader may perceive a story about a girl going to school in the morning as made up of waking up, talking to parents during breakfast, and meeting up with friends at the bus stop. Each event presents a new set of circumstances: Waking up might involve the character of the child, the location of the bedroom, and the goal of getting dressed, whereas eating breakfast might introduce new characters, change the location, and call up new goals. When the child walks into the kitchen and greets her parents, the reader may well activate representations of the new spatial location, of the new characters, and of why the characters are doing what they are doing in order to construct a representation of the new event. More broadly, readers may engage more actively with the situation described by a story at those points when new events in the story begin.

When new events begin, readers may engage in a *global updating* process, in which they abandon their old event representations and create new ones (Kurby & Zacks, 2008). The new event representation sets up the reader to comprehend the new activities in the story. Once a new event has begun, however, readers may then engage in *incremental updating* of their event representations as changes to the situation are described by the text (Zwaan, Langston, & Graesser, 1995). For example, the girl in our story has engaged in a breakfasting event, but now carries out certain actions to complete the activity. She will likely interact with new objects and pursue new lower-level goals, such as putting her cereal bowl on the counter and grabbing some milk. She is still breakfasting, but is now further along

C. A. Kurby (✉)
Department of Psychology, Grand Valley State University,
2224 Au Sable Hall,
Allendale, MI 49401, USA
e-mail: kurbyc@gvsu.edu

J. M. Zacks
Washington University,
St. Louis, MO, USA

in the episode. Readers may track these incremental changes and update their event representations accordingly in order to build their story representations (e.g., Gernsbacher, 1990; Zwaan et al., 1995). Global and incremental updating of event representations are likely critical components of comprehension. They may well play quite different roles in comprehension. However, thus far when they have been studied, each has been studied in isolation. This has left a critically important knowledge gap: It is currently unknown whether global updating, incremental updating, or both occur during language comprehension. Here we aimed to investigate the joint contributions of these processes to understanding.

Most current theories of event comprehension argue that readers track and update representations of situational dimensions (Gernsbacher, 1990, 1997; Zacks, Speer, Swallow, Braver, & Reynolds, 2007; Zwaan et al., 1995; Zwaan & Radvansky, 1998). However, these theories differ in their proposals for how event representations are constructed and updated. The event-indexing model proposes an incremental updating process, in which changed information is incorporated into one's representation as it changes (Zwaan et al., 1995; Zwaan, Magliano, & Graesser, 1995; Zwaan & Radvansky, 1998). This model argues that readers develop situation models by tracking values along multiple situational dimensions, including characters, objects, and goals. For each dimension, when a change in that dimension is described by the text, the model is updated on that dimension to reflect the new information, building up a representation as it were "brick by brick." As a result, according to this model, the reader maintains an "integrated model" that represents what has happened so far in the story and updates this model as new changes occur (Zwaan & Radvansky, 1998). For example, when the girl in our story puts her cereal bowl on the counter and grabs the milk, readers would update their representations to reflect this new character–object interaction. This requires updating the object dimension, but not the dimensions of character or spatial location.

Another theory of event comprehension, event segmentation theory (EST; Zacks et al., 2007), proposes a global updating process in which something causes the reader to abandon one representation and create a new one from scratch. EST claims that readers routinely segment narratives into discrete units or events. The theory posits a prediction-error-monitoring and gating system whereby mental models of the current event, called *event models*, guide predictions about what will happen next. Event models are working memory representations that are distinctive in several regards: First, they are multimodal. Second, they have a larger capacity than is typically measured for simple verbal or spatial materials. In these regards, they are similar to the *episodic buffer* proposed by Baddeley (2000). Finally, event models most of the time are protected against updating. This allows them to maintain event representations in the face of occlusion, distraction, and missing information, which is valuable for driving predictions. Those predictions are checked against what actually happens in the story. Most of the time, prediction error is low and the current event model is maintained. However, when the situation changes, prediction error will transiently increase. When prediction error transiently increases, the old event model is released and a new one is constructed on the basis of the previous model's contents and the currently available perceptual (or textual) information. In computational modeling, this has been implemented as a gated recurrent network, with the inputs to the event models gated open only at increases in prediction error (Reynolds, Zacks, & Braver, 2007). Once the new model is constructed, prediction error typically declines, and the reader continues with the new event model (Kurby & Zacks, 2008; Zacks et al., 2007). A side effect of this updating process is the perception of an event boundary (Zacks et al., 2007). Recent behavioral and neurophysiological research has shown that prediction error is highest at event boundaries, and making predictions across event boundaries activates neural systems important for signaling prediction error during learning (Zacks, Kurby, Eisenberg, & Haroutunian, 2011).

The incremental mechanism proposed by the event-indexing model and the global updating mechanism proposed by EST are two theoretically distinct ways that a comprehender could update memory representations in response to changes in a story's situation. However, the two mechanisms need not be mutually exclusive: Event representations could be updated globally at event boundaries *and* incrementally within events. This suggests that the current forms of the event-indexing model and EST may be incomplete: The event-indexing model could be extended to include a global updating mechanism, and EST could be extended to include an incremental mechanism. A third comprehension theory, the *structure-building framework* (Gernsbacher, 1990, 1997), explicitly includes mechanisms for both types of updating. The structure-building framework states that readers build an initial mental structure, termed "laying a foundation," onto which they incrementally map new story information, termed "mapping." When incremental mapping of story information becomes difficult, however, readers segment the story and shift to build a new structure, termed "shifting." This global shift suppresses old story information and begins mapping new story information onto a new structure. Although some evidence does support the separate components of the structure-building framework (foundation laying, shifting, etc.; see Gernsbacher, 1997, for a review), the components have consistently been studied in isolation. Furthermore, the study of individual components (e.g., foundation laying vs. shifting) without a systematic comparison of the components to each other does not allow one to assess whether they reflect different types of updating. Indeed, McNamara and Magliano (2009) argued that the current evidence does not conclusively support the claim that the components posited by the structure-building framework are distinct functional mechanisms. Thus, the important theoretical question of the present study is whether each of these models, and future models, should be developed so as to explicitly include incremental updating, global updating, or both.

To date, laboratory studies have provided strong evidence that updating takes place but have not teased apart incremental from global updating. This becomes clear when comparing the results that have been used as evidence for incremental and global updating of situation models. Incremental updating is based on situational dimensions being updated when they change. Studies of reading time, memory access, and brain activity have found evidence for updating at situation changes (for reviews, see Kurby & Zacks, 2008, and Zwaan & Radvansky, 1998). Readers slow down at changes in the situation (Rinck & Weber, 2003; Zacks, Speer, & Reynolds, 2009; Zwaan, 1996; Zwaan et al., 1995). After a situation change, the accessibility of recently presented information changes (Radvansky & Copeland, 2010; Rinck & Bower, 2000; Rinck, Hähnel, Bower, & Glowalla, 1997; Speer & Zacks, 2005; Zwaan, 1996). Situation changes correspond with later long-term memory for texts (Zwaan et al., 1995). Finally, a collection of brain regions have been found to increase in activity at points of situational change in the story (Speer, Reynolds, Swallow, & Zacks, 2009; Whitney et al., 2009). Similar results have been found for the processing of films (Swallow et al., 2011; Swallow, Zacks, & Abrams, 2009; Zacks, Speer, Swallow, & Maley, 2010).

Such results are consistent with incremental updating. However, they do not establish whether the updating is in fact incremental rather than global. According to EST, situation changes will tend to lead to event boundaries, and there is direct evidence that readers identify event boundaries at situation changes (Magliano, Miller, & Zwaan, 2001; Speer & Zacks, 2005; Zacks et al., 2009). Thus, it is possible that readers slow down and update their memory at situation changes because some situation changes are perceived as event boundaries.

Global updating says that event representations are updated as a whole unit when a reader encounters a significant boundary in the text. Consistent with this proposal, reading time slows at event boundaries (Zacks et al., 2009), and memory for recently presented words is reduced following an event boundary (Gernsbacher, 1985; Speer & Zacks, 2005). Information presented at the beginning of a story, episode, paragraph, or sentence tends to be processed more slowly (Gernsbacher, 1985; Glanzer, Fischer, & Dorfman, 1984; Haberlandt, Berian, & Sandson, 1980) and is more accessible during ongoing comprehension (Gernsbacher, Hargreaves, & Beeman, 1989). Event structure affects long-term memory for texts (Bower, Black, & Turner, 1979; Mandler & Goodman, 1982). At event boundaries, readers show transient changes in brain activity (Speer, Reynolds, & Zacks, 2007). Similar results have been found during movie viewing (Hanson & Hirst, 1989; Lichtenstein & Brewer, 1980; Newtson, Engquist, & Bois, 1977; Schwan & Garsoffky, 2004; Swallow et al., 2011; Swallow et al., 2009; Zacks et al., 2001).

Although these results are consistent with global updating, they also could be explained by incremental updating. Again, the key is that situation changes correspond with event boundaries. If readers incrementally update their memories at situation changes, and these co-occur with event boundaries, then updating would tend to occur at event boundaries. The extant data do not directly show that updating is global rather than incremental.

Thus, although previous data have provided strong evidence for memory updating during story reading and have related it to situation changes and event boundaries, the available data do not discriminate global from incremental updating. This is an important lacuna because, as we saw, current theories of discourse comprehension propose quite different roles for these two mechanisms. The key piece of missing information is this: During incremental updating, only the information that has changed is updated. During global updating, a new event model is built, updating both changed and unchanged information. That is, starting from scratch for global updates entails generating new event representations in which parts of the new model will reflect the newly changed information and parts will reflect the situational information that is still relevant from the previous event. In this study, we assessed both changed and unchanged information in order to dissociate the two kinds of updating.

We assessed the updating of concepts by measuring the likelihood of the mention of situational information in think-aloud responses.[1] Following previous arguments (Ericsson & Simon, 1993), we propose that information that is currently being updated, or added to the current event model, will have a high likelihood of being mentioned in a verbal report. Thus, when a situational dimension is being updated, it will likely be mentioned in the think-aloud response. Furthermore, we propose that variability in mention can be

[1] In the present study, participants typed their thoughts rather than reported them aloud. However, in order to stay consistent with the literature, we use the term "think-aloud" to refer to the verbal reports we collected.

accounted for by variables representing global or incremental updating.

Both theoretical and experimental work has suggested that a reader's sensitivity to situational changes is reportable in think-aloud responses; it can be available in working memory and accessible to conscious awareness. Think-aloud responses collected during comprehension have been argued to be sensitive to the activation strength of concepts in working memory (Ericsson & Simon, 1993; van den Broek, Risden, & Husebye-Hartmann, 1995) and to provide a window into the mental processes engaged during comprehension (Pressley & Afflerbach, 1995). The claim that think-aloud responses reveal online comprehension processes has been corroborated through converging measures such as reading times, probe response latencies, and eye movements. These comprehension processes have included causal and goal inferences (Langston & Trabasso, 1999; Trabasso & Magliano, 1996; Trabasso & Wiley, 2005); reader goals (Kaakinen & Hyönä, 2005; Narvaez, van den Broek, & Ruiz, 1999; van den Broek, Lorch, Linderholm, & Gustafson, 2001); the activation of story information (Langston & Trabasso, 1999); and predictions, explanations, and knowledge-based inferences (Magliano, Trabasso, & Graesser, 1999). Studies using think-aloud methodologies have shown that the content of reported thoughts can be affected by the situational structure of stories. Magliano, Zwaan, and Graesser (1999) found that readers change the types of inferences they generate depending on whether there are situational changes in the story. For example, readers were more likely to elaborate on current information and less likely to bridge back to previous story information when there was a break in the event structure. In Magliano, Zwaan, and Graesser's study, the content of think-aloud responses was also related to a reader's perceived fit between the current situation and the story as a whole. Think-aloud responses have been shown to reveal how readers track goals and causation during comprehension, demonstrating that think-aloud responses are sensitive to a reader's moment-by-moment event processing of narratives (Lutz & Radvansky, 1997; Magliano et al., 1999; Suh & Trabasso, 1993; Trabasso & Magliano, 1996). The comprehension processes revealed in think-aloud responses have also been used to distinguish skilled from less-skilled readers, and shown to correlate with online and offline measures of comprehension (Magliano & Millis, 2003; Magliano, Millis, The RSAT Development Team, Levinstein, & Boonthum, 2011; Millis, Magliano, & Todaro, 2006). Think-aloud methodologies have also been implemented in studies of event cognition more broadly, revealing, for example, people's perception of hierarchical structure in continuous activity (Zacks, Tversky, & Iyer, 2001). Thus, think-aloud responses are sensitive to online reading processes and can provide a window into the activation and updating of information in working memory.

In the present experiment, participants read a long narrative text a clause at a time and typed their thoughts after each clause.[2] We had participants report their thoughts after each clause because situational changes, and segmentation, can and do occur on a clause-by-clause basis, and we wanted to capture any corresponding moment-by-moment changes in processing. After the think-aloud task, participants segmented the text into fine (short-timescale) and coarse (long-timescale) events. These points of segmentation indicate, when a reader perceives an event boundary, that a new event in the story has begun. If global updating occurs during comprehension, it should occur at these moments. We coded the texts themselves for changes during each clause on the six situational dimensions proposed by the event-indexing model (Zwaan et al., 1995; Zwaan & Radvansky, 1998): character, object, space, time, goal, and cause. This provided a measure of the moment-by-moment changes in the situation. As a measure of situational activation during comprehension, we assessed whether each individual think-aloud response produced by the reader mentioned each of those six dimensions. Using the event segmentation data and the situation change coding, we assessed the contributions of incremental and global updating on mention of situational information.

First, we can make a general prediction on the basis of features common to the event-indexing model, EST, and the structure-building framework: Readers will be most likely to mention situational dimensions when those dimensions change and at event boundaries. Such a finding would be expected regardless of whether event model updating is incremental or global—readers tend to perceive event boundaries at situation changes (Zacks et al., 2009). This prediction is also supported by findings showing that readers focus attention on the "here and now" when constructing situation models, giving preference to what is new in the situation and currently relevant (Zwaan & Madden, 2004). Second, we can make different predictions depending on whether event model updating is incremental-only, global-only, or both. If updating is incremental-only, variance in the mention of situational dimensions will be explained by moment-by-moment changes in the situation, not by segmentation. That is, the probability of mentioning a dimension should increase when that dimension changes, independent of whether the change forms an event boundary. The opposite would be expected if updating is global-only: Variance in mention would be explained by segmentation, not by moment-by-moment changes in the situation.

[2] Work in verbal protocol analysis in comprehension has shown that typing has minimal effects on thought quality, relative to spoken-aloud thoughts (Muñoz, Magliano, Sheridan, & McNamara, 2006).

That is, if a change on a dimension is not considered an event boundary, the likelihood that that dimension will be mentioned will not change. If both incremental and global updating occur simultaneously during event comprehension, then when modeled simultaneously, both situation changes and segmentation will be uniquely associated with increased likelihood of mention.

In addition to testing the above possibilities, the present study also allows us to test predictions derived from the so-called "additivity hypothesis," which states that each additional change leads to an additional increase in situation model updating (Magliano et al., 2001; Zwaan et al., 1995; Zwaan, Radvansky, Hilliard, & Curiel, 1998). Previous research found support for such a possibility. For example, Zwaan et al. (1998) found a linear increase on reading times with the number of changes encountered at each clause (see also Rinck & Weber, 2003). Magliano et al. (2001) and Zacks et al. (2009) found that the probability of segmentation increased with an increasing number of situational changes. (However, evidence in support of the additivity hypothesis has not always emerged cleanly: see Zwaan et al., 1995). In this paradigm, the additivity hypothesis predicts that the more situation changes readers encounters in a clause, the more dimensions they will mention in their think-aloud responses.

Method

Participants

A group of 41 undergraduates participated in this experiment for course credit or cash payment. Seven of the participants were excluded due to lack of a full data set (n = 5) or to computer malfunctions (n = 2). One participant was replaced because of identifying more coarse than fine event boundaries (i.e., the participant was likely not following directions). The final sample included 34 participants.

Materials

The narrative texts were four scenes from the book *One Boy's Day* (Barker & Wright, 1951), which had been used previously by Zacks et al. (2009). *One Boy's Day* is an observational record of the activities of a seven-year-old boy (pseudonym Raymond Birch) throughout a 12-hour period on a day in the 1940s. The texts described Raymond getting up and eating breakfast ("Waking up"), playing in the schoolyard ("Play before school"), having a music lesson ("Music lesson"), and having an English lesson ("Class work"). The texts ranged from 1,107 to 1,404 words ("Waking up," 1,364 words, 192 clauses; "Play before school," 1,107 words, 178 clauses; "Music lesson," 1,404 words, 215 clauses; "Class work," 1,182 words, 172 clauses). As can be seen in the example in Fig. 1, the narratives described a natural flow of everyday activities. They differed from typical literary narratives in that there was no story arc or conflict that drove character motives. However, they did reflect the structure of everyday life, including shifts in causes, characters, goals, objects, and space. These materials were selected to afford generalization between texts and everyday events (Zacks et al., 2009).

Procedure

The experiment proceeded in two phases: a think-aloud phase and a segmentation phase. In the think-aloud phase, participants read a text one clause at a time. Each clause was presented left justified and centered vertically. After reading each clause, participants were instructed to type their thoughts regarding their understanding of the unit of text they had just read into a box in the bottom of the screen. They were told to do this on the basis of their understanding of the text so far. Participants were told that there were no correct or incorrect thoughts and that they should simply type in what thoughts came to mind as they read and understood the text. There was no time limit. When finished, the participants pressed the F1 key, which removed their response and replaced the previous clause with the next clause. Participants initially practiced the think-aloud task on a 28-clause scene that described Raymond receiving fishing lessons from his father. Participants were given an opportunity to ask questions about the task after the practice and then were randomly assigned to one of the four experimental texts, with the constraint of maintaining as close to equal ns across texts as possible ("Play before school," n = 9; "Waking up," n = 9; "Music lesson," n = 8; "Class work," n = 8).

Upon completion of the think-aloud phase, participants began the segmentation phase. They read the same text as in the think-aloud task, but this time on paper, single-spaced and without paragraph breaks. They segmented the text twice, once at a fine grain and once at a coarse grain. The participants were instructed to read the text again and to mark off the text into meaningful units of activity by placing a line between two words when, in their judgment, one unit of activity ended and another began. For coarse segmentation, participants marked off the largest units that seemed natural and meaningful to them, and for fine segmentation, they marked off the smallest natural and meaningful units. The participants segmented the entire text at one grain size before segmenting the text again at the other grain size. Before segmenting the experimental text for both grains, the participants practiced the task on the fishing text. The

Fig. 1 Examples of situation change coding. A dot indicates a change on that dimension from the previous clause

Clause	Cause	Character	Goal	Object	Space	Time
As soon as [Mrs. Logan] made a check mark on his paper,		●				●
[Raymond] hurried back to his desk.		●	●		●	
Nearing his desk,						
he crumpled the paper, seemingly without any disappointment or anxiety.	●		●	●		
His expression was one of "Well, that's that and I'm through."						
The teacher called to him pleasantly, "Did I grade your book?"	●	●	●			
Raymond answered with a negative shake of his head.		●	●			
He picked up his English workbook				●		
and returned to her desk.					●	
He walked briskly.						●
He laid his workbook on the desk.				●		

order of segmentation was counterbalanced across participants.

Scoring

Situation change coding of the narratives We used the situation change coding from Zacks et al. (2009). They coded the narratives, clause by clause, for changes on the character, object, space, goal, and cause dimensions. Given that the narratives described naturalistic activities, there were no temporal breaks. As such, the time dimension was coded regarding whether or not a clause contained a temporal reference, such as "immediately" or "slowly" (for details, see Zacks et al., 2009). As reported by Zacks et al. (2009), the average interrater reliability for the situation change coding was .77, as measured by Cohen's kappa. See Fig. 1 for an example of the situation change coding scheme.

Scoring event segmentation locations The scoring of the event boundary locations produced by the participants was identical to the method presented in Zacks et al. (2009). For each participant, a clause was considered an event boundary if the participant placed an event boundary within it. When an event boundary was placed at a sentence boundary, the boundary was assigned to the clause that immediately followed; that is, the new event began with the first word of the new clause.

Selection and scoring of think-aloud responses We selected a subset of the verbal responses produced by each participant to assess the mention of the six situation dimensions. Responses were selected randomly, and separately, for each participant on the basis of the location of their individually placed event boundaries. For each participant, we randomly selected up to 20 think-aloud responses from clauses that were event boundaries. Of these, 10 were fine event boundaries and up to 10 were coarse boundaries. A clause was considered a fine boundary if the participant identified it as a fine boundary and not as both a fine and a coarse boundary. We considered a clause a coarse boundary *only* if it was identified as both a coarse and a fine boundary; in the present data set, the percentage of coarse boundaries identified by participants as coarse but not fine boundaries was small (4.98%). The fact that coarse event boundaries typically co-occurred with fine event boundaries is consistent with previous findings that people segment activity and text hierarchically (Kurby & Zacks, 2011; Zacks et al., 2009; Zacks, Tversky, & Iyer, 2001). We also randomly selected up to 20 clauses not identified as either fine or coarse event boundaries, which we will refer to as *event middles*. Given that there was variability in how often participants segmented the text, some participants had fewer than 10 coarse boundaries or fewer than 20 event-middle clauses that could be selected. In these cases, we selected all of the clauses available. On average, 9.62 coarse boundaries (mode = 10, SD = 1.21) and 16.71 middles (mode = 20, SD = 5.24) were selected per participant. Every participant had 10 fine boundaries.

For each think-aloud response selected (up to 40 for each participant), we coded whether it mentioned each of the six situation dimensions: character, object, space, time, goal, and cause. A response could potentially mention any combination of dimensions. See Table 1 for a description of the coding scheme and Table 2 for example think-aloud responses and coding. Two coders were trained on the rules presented in Table 1 and practiced on responses to clauses in the practice text from 6 randomly selected participants. Agreement between the coders across the six dimensions was high; the average Cohen's kappa was .91. The two coders then each coded half of the experimental think-aloud responses. The coding team met to discuss their scoring, and disagreements were resolved by discussion.

Results

Segmentation

First, we examined readers' perceptions of the average event length. For each participant, we assessed the average length, in clauses, of the fine and coarse units by dividing the number of clauses in the narrative by the number of identified event boundaries (means presented in Table 3). A 2 (grain) × 4 (text) mixed ANOVA revealed a significant main effect of grain, $F(1, 30) = 46.43$, $p < .001$, $\eta_p^2 = .61$. As expected, the lengths were longer for coarse ($M = 8.96$ clauses, $SD = 6.64$) than for fine ($M = 2.11$ clauses, $SD = 1.81$) units, indicating that participants were able to follow the instructions. There were no other significant effects (largest $F = 1.49$).

We fit logistic mixed-effect models to assess the relation between situation changes and segmentation (Jaeger, 2008). Mixed-effect models allow for simultaneous adjustments of the error term on the basis of multiple random effects, while modeling the fixed effects of variables of interest (Baayen, Davidson, & Bates, 2008; Jaeger, 2008). We fit two such models: one predicting coarse segmentation locations from the presence or absence of the six situation changes, and one predicting fine segmentation from the same predictors. In addition to the six situation change predictors, each model included predictors coding for the random effects of subject and text and for the fact that subject was embedded in text (Baayen et al., 2008). Figure 2 presents the odds ratios computed by the models, with 95% confidence intervals. As can be seen in the figure, situation changes significantly increased the odds of segmenting for every dimension and grain, except for object changes predicting coarse segmentation, $z(6420) = 1.34$, $p = .18$. In sum, readers tended to segment the texts when dimensions of the described situation changed.

Predicting mention of situation dimensions from event structure and situation changes

We performed two analyses to test directly for the presence of global and incremental updating in the think-aloud responses. We began by computing logistic mixed-effect models predicting the mention of each of the six dimensions as a function of situation changes and event structure. Random effects of subject and text, as well as the random effect of subject

Table 1 Verbal protocol coding scheme

Dimension	Definition
Character	Intentional agent. Coded if mentioned by name or anaphoric reference, such as "he," "she," "him," etc.
Object	Concrete concepts that could be manipulated or acted on and were not spaces or characters. Coded if mentioned by name or anaphoric reference, such as "it," "thing," etc.
Space	Settings in which events can occur. Coded if mentioned by name or reference, such as "there" or "here."
Time	Temporal settings in which events can occur. Coded if mentioned by name, such as "early" or "late"; if reference to a temporal interval was made, such as "begin" or "end"; if a temporal adverb was mentioned; or if other temporal references were mentioned, such as "now," "later," "day," "slow," "fast," "continuously," etc.
Goal	A character's future desired outcome, or a future event/action or plan. Coded if plan/outcome was stated directly (e.g., "going to go fishing") or if statement was marked as goal-related by words such as "want" or "intend."
Cause	Causal explanation. Coded if explanation was mentioned for an action described in the text (see the table note below).

We performed two tests to determine whether a statement was a causal explanation. First, we conducted a *counterfactual reasoning test* based on techniques used by Trabasso and colleagues (Langston & Trabasso, 1999; Trabasso, van den Broek, & Suh, 1989) to identify the presence of causal information in texts. The test had the form "if [explanatory action/condition mentioned in verbal protocol] did not occur, would [action mentioned in text] occur?" The coder's judgment was based on what was reasonable given the depicted story world. If the coder's answer to the counterfactual test was "no," then the statement in the protocol was considered to have causal content. The example statement in the table passes the counterfactual test because, on the basis of what was occurring in the story world, the answer to, "If Raymond was not in a hurry, would he walk briskly?" was "no." The second test for causal content was a *"because" test*. For this test, we inserted "because" between the action described in the clause and the candidate statement in the verbal protocol, and then made a sensibility judgment about this new statement. If the new hypothetical statement was judged sensible, then it was considered to have causal content. The above protocol also passes the "because" test. A participant could also signal that they were producing a causal explanation by explicitly using the word "because." A candidate statement in a verbal protocol needed to pass both the counterfactual and "because" tests to be considered a mention of a cause.

Table 2 Example think-aloud responses and coding

Text clause	Verbal Protocol	Character	Object	Space	Time	Goal	Cause
Mrs. Birch chuckled with slight embarrassment	Mrs. Birch was embarrassed too.	1	0	0	0	0	0
and put them on,	He wants to wear the shorts.	1	1	0	0	1	0
meaning that he was to take off his pajama pants	He has to start over and undress.	1	0	0	0	1	0
and put on his underpants.	Oh, he's not wearing any underwear.	1	1	0	0	0	0
She stood next to him	because he needed her to be there to do that.	1	0	1	0	0	1
as he made the change.	Or maybe she wanted to give him space.	1	0	1	0	1	0
Mrs. Birch returned to the kitchen.	She has an egg cooking.	1	1	0	0	0	1
Raymond put on his blue-jean pants	Raymond then needed to re-tie his shoes	1	1	0	1	0	0
as he stood by his bed.	and was going to call his mother again.	1	0	0	1	1	0

A 1 in a cell indicates that the protocol was given credit for mentioning that dimension, and a 0 indicates that credit was not given.

embedded in text, were also included in the models. (For the situation change predictors, the no-change condition served as the reference condition; for the event structure predictor, the event-middle condition served as the reference condition.) These models will be referred to as *full* models. To test for incremental updating, we asked whether situation changes predicted mention while controlling for the effect of event structure by examining the regression weights associated with the situation change predictor in each full model. Table 4 presents the odds ratios obtained for each predictor in the full models, for each dimension. To aid interpretation of the situation change effects on mention, we computed the likelihood that each participant mentioned each dimension as a function of whether a change occurred on that particular dimension, plotting them in Fig. 3. As can be seen in Table 4 and Fig. 3, changes in the character, object, space, and time dimensions were associated with increased frequency of mentioning each of those dimensions, and was marginally so for goals. These results provide evidence for incremental updating. Change in the cause dimension was associated with *reduced* rates of mention; this effect was small but significant.

To test for global updating, we first assessed whether in the full models event structure predicted unique variance in mention after controlling for situation changes. We did so by examining the regression weights associated with the event structure predictors. To aid interpretation of the event structure effects on mention, we computed the likelihood that each participant mentioned each dimension after reading event middles, fine boundaries, and coarse boundaries, and plotted them in Fig. 4. As can be seen in Table 4 and Fig. 4, relative to event middles, mention of the character, goal, and cause dimensions significantly increased for fine boundaries, and mention of the time dimension increased for coarse boundaries. Second, we compared each full model to a reduced model that included only the situation change variable as a predictor of mention on that particular dimension, and used log-likelihood ratio tests to determine whether including the event structure variable as a predictor improved model fit (Baayen, et al., 2008; Jaeger, 2008). Table 5 presents each model. The χ^2 and p values describe the results of the log-likelihood ratio tests. As can be seen in Table 5, event structure improved model fit for the mention of character, time, goal, and cause, but not for the mention of objects and space. These results provide evidence for the presence of global updating.

Table 3 Mean fine and coarse unit lengths, in number of clauses, for each text (*SD*s in parentheses)

Text	Fine	Coarse
Play before school	1.91 (1.26)	11.30 (7.73)
Waking up	1.60 (0.90)	6.19 (2.94)
Music lesson	2.90 (3.02)	8.44 (6.63)
Class work	2.13 (1.54)	9.98 (8.21)

Global updates at event boundaries should update unchanged information as well as changed information. Although in the above analyses we statistically controlled for variability associated with situation changes when assessing the effects of segmentation, we conducted follow-up analyses to further investigate global updating. These analyses took the same form as those reported above, but were restricted to the mention of dimensions that *did not change* in each clause. The results converged with the previous analyses, supporting the possibility that nonshift dimensions are updated at event boundaries. Figure 5 presents the mean likelihoods of mention for each dimension when no change occurred on that dimension. When we compared boundaries to event middles in this no-change data set, fine boundaries were associated with a significant increase in the mention of goals [odds ratio = 2.29; $z(926) = 2.87$, $p = .004$] and causes

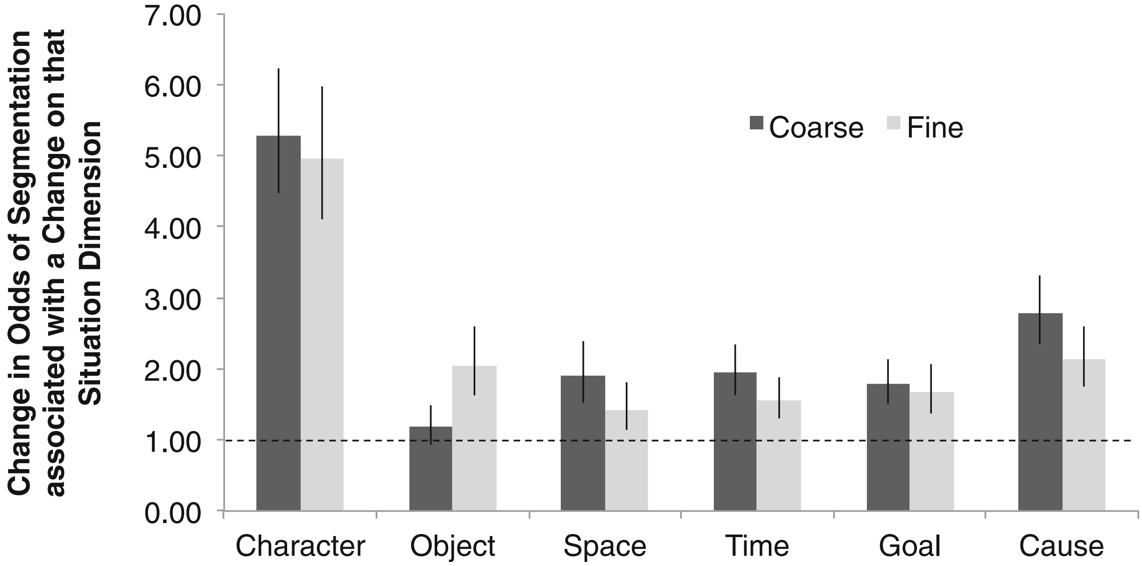

Fig. 2 Odds ratios representing the relation between situation changes and coarse and fine segmentation. Error bars indicate 95% confidence intervals. Odds ratios of 1 indicate no change

[odds ratio = 1.81; $z(927) = 2.95$, $p = .003$]. [The effect was marginally significant for characters: odds ratio = 1.39, $z(844) = 1.76$, $p = .079$.] Coarse boundaries were associated with a significant increase in the mention of time [odds ratio = 1.53, $z(1007) = 2.07$, $p = .039$]. [Consistent with the original mixed-effect analyses, there were no significant effects of segmentation for the mention of space, smallest $p = .454$. For objects, fine boundaries were marginally associated with a decrease in mention probability: odds ratio = 0.72, $z(1107) = –1.81$, $p = .071$.] Together with the previous analyses, these data support the possibility that readers update both changed and unchanged information at event boundaries.

Finally, we computed two linear mixed-effects models to assess the additivity hypothesis. For these models, the

Table 4 Odds ratios and 95% confidence intervals of the predictors in the full logistic mixed-effects models

Model	Predictor	Odds Ratio	95% CI	$z(1232)$	p
Character	Character change[a]	1.60	1.15–2.22	2.79	.005**
	Coarse boundary[b]	1.35	0.94–1.93	1.64	.101
	Fine boundary[b]	1.75	1.26–2.43	3.32	<.001***
Object	Object change[a]	3.39	2.27–5.05	5.99	<.001***
	Coarse boundary[b]	0.87	0.62–1.20	–0.85	.394
	Fine boundary[b]	0.81	0.58–1.12	–1.28	.201
Space	Space change[a]	5.99	3.76–9.54	7.54	<.001***
	Coarse boundary[b]	0.66	0.39–1.14	–1.49	.136
	Fine boundary[b]	0.93	0.56–1.54	–0.29	.769
Time	Time reference[a]	1.51	1.05–2.17	2.24	.025*
	Coarse boundary[b]	1.57	1.10–2.23	2.49	.013*
	Fine boundary[b]	1.11	0.77–1.61	0.55	.585
Goal	Goal change[a]	1.54	0.95–2.50	1.76	.079
	Coarse boundary[b]	1.10	0.61–1.97	0.31	.759
	Fine boundary[b]	2.11	1.29–3.46	2.96	.003**
Cause	Cause change[a]	0.64	0.43–0.96	–2.16	.031*
	Coarse boundary[b]	1.39	0.91–2.11	1.51	.131
	Fine boundary[b]	1.88	1.30–2.71	3.36	<.001***

[a] For the change predictors, the reference condition is the no-change condition (or no temporal reference for the time reference predictor). [b] For the event structure predictors, the reference condition is the event-middle condition. *** $p < .001$, ** $p < .01$, * $p < .05$.

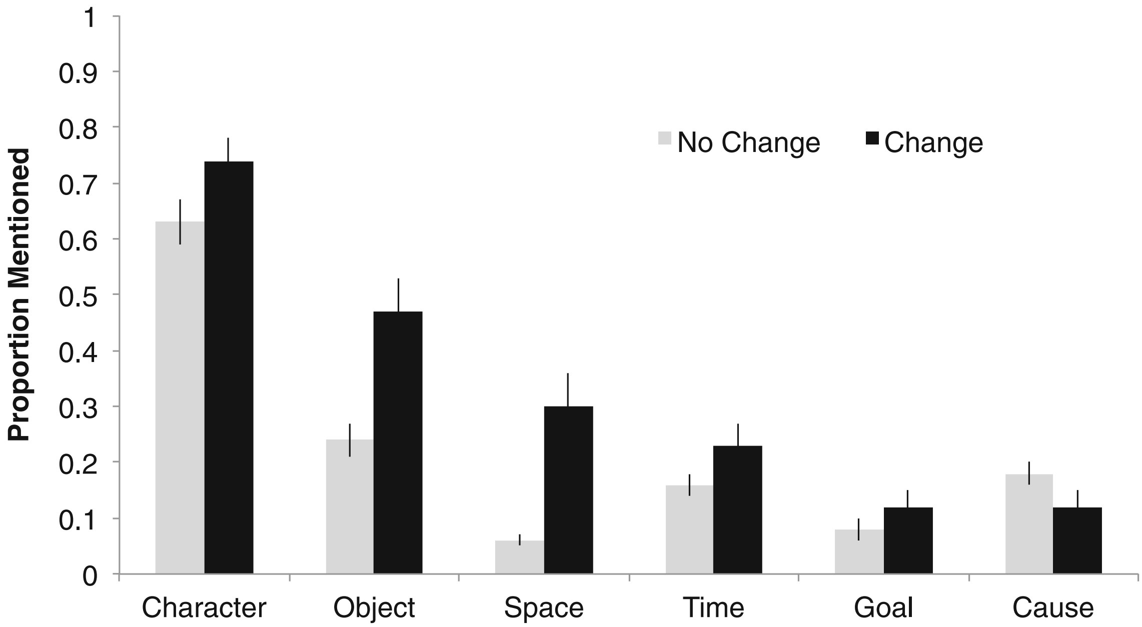

Fig. 3 Proportion mention scores for each dimension by change condition. The bars indicate mean performance collapsed across participants. The error bars depict standard errors

dependent measure was the number of dimensions mentioned in each think-aloud response. The predictor variables were (1) the number of situation changes in each clause, (2) the presence of a coarse boundary, and (3) the presence of a fine boundary. We then computed the two models: a full model, using the same random effects as the previous analyses, which predicted the total number of mentioned dimensions from the total number of situation changes and the presence of event boundaries (with the event-middle condition as the reference condition); and a reduced model that included only the number-of-changes variable as a predictor. Table 6 presents the coefficients from each model. As can be seen in the table, the number-of-changes predictor was a significant predictor of the number of dimensions mentioned in the reduced model, but not in the full model. In the full model, the coefficient for fine boundaries was a significant predictor (with coarse boundaries a marginal predictor). To test whether event structure explained additional variance in mention

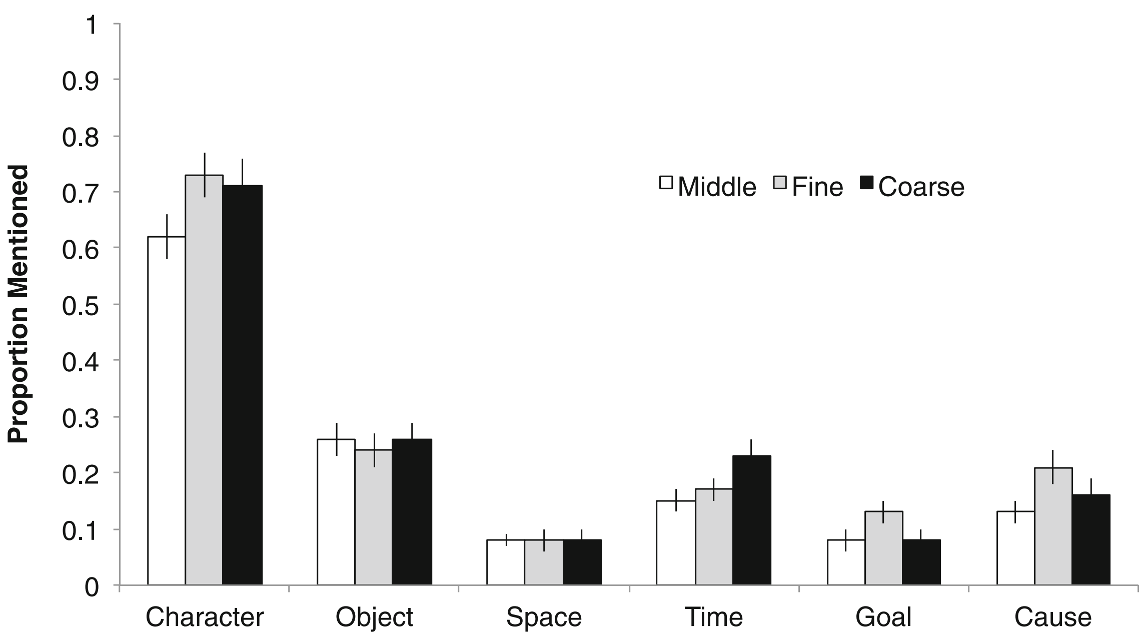

Fig. 4 Proportion mention scores for each dimension by event structure condition. The bars indicate mean performance collapsed across participants. The error bars depict standard errors

Table 5 Logistic mixed-effect models for each dimension and log-likelihood tests between the reduced and full models for the event structure analyses

Model	Equation	Log Likelihood	$\chi^2(2)$	*p*
$\text{Character}_{\text{reduced}}$	Character Mention = Character Change	–708.96	–	–
$\text{Character}_{\text{full}}$	Character Mention = Character Change + Event Structure	–703.42	11.09	.004**
$\text{Object}_{\text{reduced}}$	Object Mention = Object Change	–667.62	–	–
$\text{Object}_{\text{full}}$	Object Mention = Object Change + Event Structure	–666.73	1.78	.412
$\text{Space}_{\text{reduced}}$	Space Mention = Space Change	–328.90	–	–
$\text{Space}_{\text{full}}$	Space Mention = Space Change + Event Structure	–327.68	2.45	.294
$\text{Time}_{\text{reduced}}$	Time Mention = Time Reference	–567.75	–	–
$\text{Time}_{\text{full}}$	Time Mention = Time Reference + Event Structure	–564.61	6.29	.043*
$\text{Goal}_{\text{reduced}}$	Goal Mention = Goal Change	–349.30	–	–
$\text{Goal}_{\text{full}}$	Goal Mention = Goal Change + Event Structure	–344.07	10.45	.005**
$\text{Cause}_{\text{reduced}}$	Cause Mention = Cause Change	–531.46	–	–
$\text{Cause}_{\text{full}}$	Cause Mention = Cause Change + Event Structure	–525.91	11.09	.004**

** $p < .01$, * $p < .05$

above the number of changes, we compared the full model to the reduced model using log-likelihood ratio tests. Table 7 shows the results of this analysis. As can be seen in the table, event structure significantly improved model fit. The fact that the number of dimensions mentioned increased with the number of dimensions changing supports the additivity hypothesis. The fact that this was partially explained by event structure suggests a particular mechanism: It may be that each additional situation change has an additional effect because it increases the probability of segmentation.

Discussion

In this study, we tested predictions from theories of situation model construction and event comprehension regarding

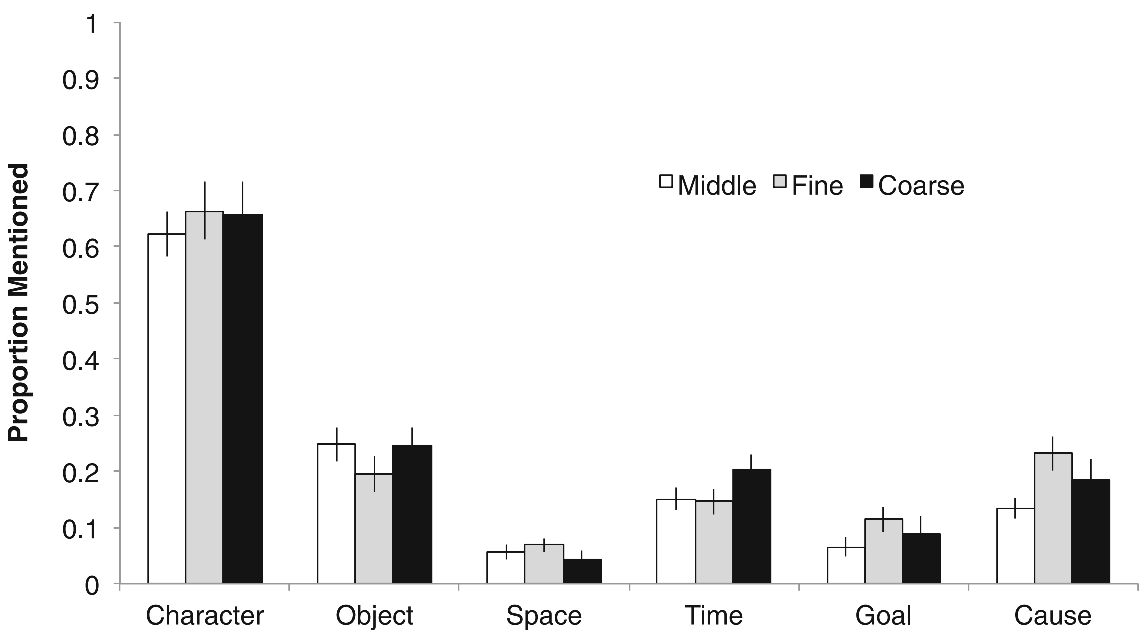

Fig. 5 Proportion mention scores for each dimension by event structure condition, when there was no change on that dimension. The bars indicate mean performance collapsed across participants. The error bars depict standard errors. Because event boundaries tend to occur when there is a change on a dimension, excluding clauses for which there was a change resulted in the lack of observations for some participants in the event boundary conditions. For the character dimension, 1 participant had no coarse observations, and 2 participants lacked coarse observations for the cause dimension. These situations were treated as missing data, and the data for those cases were imputed by predictions from a within-participant logistic regression predicting mention from the event structure variables and the individual situation change variable of interest

Table 6 Results from the mixed-effect models testing the effects of the total number of changes on mention

Model	Predictor	Coefficient	*SE*	*t*(1232)	*p*
Total $\text{Mention}_{reduced}$	Total changes	.06	.02	2.76	.006**
Total Mention_{full}	Total changes	.04	.03	1.41	.183
	Coarse boundary[a]	.14	.08	1.76	.079†
	Fine boundary[a]	.24	.07	3.32	<.001***

[a] For the event structure predictors, the reference condition is the event-middle condition. *** $p < .001$, ** $p < .01$, † $p < .10$.

whether both incremental and global forms of updating occur during narrative comprehension. We found evidence for both. Readers were more likely to mention the dimensions of character, object, space, and time when those dimensions changed, supporting the claim that they updated those dimensions incrementally when each dimension changed. Finding incremental updating is consistent with previous results showing incremental updating during narrative comprehension (for a review, see Zwaan & Radvansky, 1998) and with the predictions of the event-indexing model and the structure-building framework. As predicted by these theories, readers mentioned situational information more at situation changes. (Readers were slightly but significantly *less* likely to mention the cause dimension when there were causal changes. This result was surprising, and does not support incremental updating. If it is found to replicate in future studies, an account of this pattern will be needed.)

Over and above the effects of situational changes on mention of particular dimensions, the presence of an event boundary was associated with increased mention of the character, cause, goal, and time dimensions. The results also show that the updating at event boundaries was independent from the updating associated with the accumulation of incremental changes. These novel findings support the claim that situation models are updated globally, "from scratch," when a new event begins. It is consistent with global-updating predictions from the structure-building framework and EST. Readers mentioned more situational information at event boundaries than in event middles, and this was true even after controlling for the effects of situation changes on mention. Also consistent with EST, we found evidence that unchanged information is also updated at event boundaries, supporting the proposal that new event models will also contain relevant previous situational information. This is the first study to show such an effect. In addition to the global-updating results, most situation change variables were significant even when simultaneously modeling the effect of event structure on mention. This suggests that both incremental and global updating processes independently contribute to situational activation during comprehension.

It is important to point out that the relation observed here between segmentation and mention is not likely to be due to task demands: Participants did not engage in the explicit segmentation task until after they were finished with the think-aloud phase. This supports the proposal that segmentation is an ongoing feature of event comprehension (Zacks et al., 2007).

The present data fit very nicely with the structure-building framework. Readers appear to initiate a new structure when a new story episode begins and then to map information incrementally throughout the episode (Gernsbacher, 1990, 1997). EST, in contrast, has no mechanism for incremental updating: Each time an update is required, the system resets the previous event model and opens the event model's input gate to build an entirely new structure (Zacks et al., 2007). The gating mechanism in EST is thus too inflexible to account for both incremental and global updating. One possibility is that gating is less complete than EST proposes, allowing information to "leak" into the event models over the course of an event. Another possibility is that event model updating is indeed an all-or-none process, but that other working memory representations are updated incrementally throughout the duration of an event. More research will be needed to distinguish between these possibilities. Conversely, the event-indexing model can easily accommodate the incremental-updating effects found here, but it does not provide an explicit mechanism for global updating in situation model construction. Although the event-indexing model does not rule out such a mechanism, it does not make explicit any

Table 7 Log-likelihood tests between the reduced and full models for the total-number-of-changes analyses

Model	Equation	Log-Likelihood	$\chi^2(2)$	*p*
Total $\text{Mention}_{reduced}$	Total Mention = Total Changes	–1,789.7	–	–
Total Mention_{full}	Total Mention = Total Changes + Event Structure	–1,784.1	11.14	.004**

** $p < .01$.

mechanisms that might explain it. Thus, a weakness of both the event-indexing model and EST is that they are not clear enough to explicitly exclude the type of updating for which they do not strongly advocate, leaving room for post-hoc inclusions of them in the models. Computational models will be needed to fully flesh out how incremental and global updating would work in situation model construction.

In addition to its main findings, the present study also provides evidence that readers track spatial information during narrative comprehension. However, whether readers track space during comprehension has been a matter of debate (see Zwaan & Radvansky, 1998). Reading time studies have typically reported no effects of spatial changes on processing time, unless extensive knowledge or strategic instructions are given (Therriault, Rinck, & Zwaan, 2006; Zwaan et al., 1995; Zwaan et al., 1998). However, in their *rapid-spatial-updating* view, Radvansky and Copeland (2010) argued that spatial updating in comprehension is easy and efficient because such updating is critical to adaptive functioning in the everyday world. They argued that reading times likely only slow down when updating is difficult, which may explain the lack of spatial effects in reading times. Using explicit memory probes for spatial information, Radvansky and Copeland found evidence of spatial updating in the absence of reading time differences. Given that our think-aloud task explicitly probed the contents of working memory, the present results are consistent with this rapid-spatial-updating account.

It is surprising, however, that we did not observe evidence for global updating of objects and space. Previous work has shown that spatial information features prominently in situation models, and in event models more broadly (Morrow, Bower, & Greenspan, 1989; Zacks et al., 2007; Zwaan et al., 1995). Additionally, evidence has suggested that object information is modulated by segmentation; in both textual and filmed narratives, object information is retrieved more slowly, and generally less accurately, across event boundaries than from within the current event (Speer & Zacks, 2005; Swallow et al., 2011; Swallow et al., 2009). The present data also show that changes in objects and spaces were associated with the perception of event boundaries. Thus, according to the structure-building framework and EST, global updating should have occurred. Given this evidence, we believe it most likely that the lack of global-updating effects for objects and spaces was a case of Type II error. Another somewhat surprising finding was the reverse effect of causal changes on the mention of causality. Such a finding is opposite to what would be expected according to the event-indexing model. It would be premature to draw strong conclusions on such a finding, which needs replication. Given that causes were more likely to be mentioned at event boundaries than event middles, it appears that causality was tracked appropriately. Events are structured hierarchically, with smaller events clustered by larger events (Zacks et al., 2001). It is possible that the causal shifts within higher-order episodes are processed differently than those signaling the end of one of these episodes. However, the present study was not designed to assess such a possibility, and more work will be required in order to understand our observed reverse causal effect and the circumstances that may have produced it.

The present study demonstrates that a reader's sensitivity to situational changes and event structure is reportable in think-aloud responses. This suggests that situational processing during comprehension is, at least partially, available to conscious awareness. This finding is consistent with both theoretical and empirical work on the utility of think-aloud methodologies in revealing online comprehension processes. However, these methodologies do have some limitations (Ericsson & Simon, 1993; Pressley & Afflerbach, 1995). First, it is possible that some of the reported thoughts were generated through strategic and reflective processes rather than though purely online, on-the-fly processes. Second, it is unlikely that think-aloud methodologies can tap comprehension processes not available to conscious awareness. Third, not all comprehension processes available to conscious awareness can be easily translated into a verbal report, such as integration difficulties. As with most behavioral and physiological measures, it is important to use converging measures in order to maintain high confidence in the validity of think-aloud data. Much research in the past has used converging measures to validate such data (e.g., Magliano et al., 1999; Suh & Trabasso, 1993). Following the logic of those studies, it will be important for future work on the present topic to use chronometric designs to test the validity of think-aloud responses in the assessment of situational updating. However, given these concerns, there is good reason to believe that think-aloud data are important to understanding the moment-by-moment processing of events. Studies of narrative comprehension that have used both offline measures, such as think-aloud data, and more online measures, such as reading time, have shown a large correspondence between the two types of data (Lutz & Radvansky, 1997; Magliano et al., 1999; Rinck & Weber, 2003; Scott Rich & Taylor, 2000). On occasion, the results of the two do diverge, with the offline measures showing less sensitivity to subtle changes in event structure than do reading times (e.g., Rinck & Weber, 2003). However, such a finding increases our confidence in the findings of the present study, because theoretically meaningful updating effects were found in the face of such possible reduced sensitivity.

Readers perceive stories as constructed of discrete events. The present results suggest that this subjective experience corresponds with the updating of working memory. Features of events change continuously over the course of a story. When features within an event change from moment to moment, they are updated incrementally, but when a new event is perceived to occur, the reader constructs a new mental model.

Author note Preparation of the manuscript was partially supported by Grants T32 AG000030-31 and RO1-MH70674 from the National Institutes of Health.

References

Baayen, R. H., Davidson, D. J., & Bates, D. M. (2008). Mixed-effects modeling with crossed random effects for subjects and items. *Journal of Memory and Language, 59,* 390–412. doi:10.1016/j.jml.2007.12.005

Baddeley, A. (2000). The episodic buffer: A new component of working memory? *Trends in Cognitive Sciences, 4,* 417–423. doi:10.1016/S1364-6613(00)01538-2

Barker, R. G., & Wright, H. F. (1951). *One boy's day: A specimen record of behavior*. New York: Harper & Brothers.

Bower, G. H., Black, J. B., & Turner, T. J. (1979). Scripts in memory for text. *Cognitive Psychology, 11,* 177–220.

Ericsson, K. A., & Simon, H. A. (1993). Verbal reports as data. *Psychological Review, 87,* 215–251. doi:10.1037/0033-295X.87.3.215

Gernsbacher, M. A. (1985). Surface information loss in comprehension. *Cognitive Psychology, 17,* 324–363.

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

Gernsbacher, M. A. (1997). Two decades of structure building. *Discourse Processes, 23,* 265–304.

Gernsbacher, M. A., Hargreaves, D., & Beeman, M. (1989). Building and accessing clausal representations: The advantage of first mention versus the advantage of clause recency. *Journal of Memory and Language, 28,* 735–755.

Glanzer, M., Fischer, B., & Dorfman, D. (1984). Short-term storage in reading. *Journal of Verbal Learning and Verbal Behavior, 23,* 467–486.

Haberlandt, K., Berian, C., & Sandson, J. (1980). The episode schema in story processing. *Journal of Verbal Learning and Verbal Behavior, 19,* 635–650.

Hanson, C., & Hirst, W. (1989). On the representation of events: A study of orientation, recall, and recognition. *Journal of Experimental Psychology: General, 118,* 136–147.

Jaeger, T. F. (2008). Categorical data analysis: Away from ANOVAs (transformation or not) and towards logit mixed models. *Journal of Memory and Language, 59,* 434–446. doi:10.1016/j.jml.2007.11.007

Kaakinen, J. K., & Hyönä, J. (2005). Perspective effects on expository text comprehension: Evidence from think-aloud protocols, eyetracking and recall. *Discourse Processes, 40,* 239–257.

Kurby, C. A., & Zacks, J. M. (2008). Segmentation in the perception and memory of events. *Trends in Cognitive Sciences, 12,* 72–79.

Kurby, C. A., & Zacks, J. M. (2011). Age differences in the perception of hierarchical structure in events. *Memory & Cognition, 39,* 75–91.

Langston, M. C., & Trabasso, T. (1999). Modeling causal integration and availability of information during comprehension of narrative texts. In S. R. Goldman & H. van Oostendorp (Eds.), *The construction of mental representations during reading* (pp. 29–69). Mahwah: Erlbaum.

Lichtenstein, E. H., & Brewer, W. F. (1980). Memory for goal directed events. *Cognitive Psychology, 12,* 412–445.

Lutz, M. F., & Radvansky, G. A. (1997). The fate of completed goal information in narrative comprehension. *Journal of Memory and Language, 36,* 293–310.

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

Magliano, J. P., & Millis, K. K. (2003). Assessing reading skill with a think-aloud procedure and latent semantic analysis. *Cognition and Instruction, 21,* 251–284.

Magliano, J. P., Millis, K. K., The RSAT Development Team, Levinstein, I., & Boonthum, C. (2011). Assessing comprehension during reading with the Reading Strategy Assessment Tool (RSAT). *Metacognition and Learning, 6*, 131–354.

Magliano, J. P., Trabasso, T., & Graesser, A. C. (1999). Strategic processing during comprehension. *Journal of Educational Psychology, 91,* 615–629.

Magliano, J. P., Zwaan, R. A., & Graesser, A. C. (1999). The role of situational continuity in narrative understanding. In S. R. Goldman & H. van Oostendorp (Eds.), *The construction of mental representations during reading* (pp. 219–245). Mahwah: Erlbaum.

Mandler, J., & Goodman, M. (1982). On the psychological validity of story structure. *Journal of Verbal Learning and Verbal Behavior, 21,* 507–523.

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

Millis, K. K., Magliano, J. P., & Todaro, S. (2006). Measuring discourse-level processes with verbal protocols and latent semantic analysis. *Scientific Studies of Reading, 10,* 251–283.

Morrow, D. G., Bower, G. H., & Greenspan, S. L. (1989). Updating situation models during comprehension. *Journal of Memory and Language, 13,* 441–469.

Muñoz, B., Magliano, J. P., Sheridan, R., & McNamara, D. S. (2006). Typing versus thinking aloud when reading: Implications for computer-based assessment and training tools. *Behavior Research Methods, 38,* 211–217. doi:10.3758/BF03192771

Narvaez, D., van den Broek, P., & Ruiz, A. B. (1999). Reading purpose, type of text and their influence on think-alouds and comprehension measures. *Journal of Educational Psychology, 91,* 488–496.

Newtson, D., Engquist, G., & Bois, J. (1977). The objective basis of behavior units. *Journal of Personality and Social Psychology, 35,* 847–862.

Pressley, M., & Afflerbach, P. (1995). *Verbal protocols of reading: The nature of constructively responsive reading*. Hillsdale: Erlbaum.

Radvansky, G. A., & Copeland, D. E. (2010). Reading times and the detection of event shift processing. *Journal of Experimental Psychology: Learning, Memory, & Cognition, 36,* 210–216.

Reynolds, J. R., Zacks, J. M., & Braver, T. S. (2007). A computational model of event segmentation from perceptual prediction. *Cognitive Science, 31,* 613–643.

Rinck, M., & Bower, G. H. (2000). Temporal and spatial distance in situation models. *Memory & Cognition, 28,* 1310–1320.

Rinck, M., Hähnel, A., Bower, G. H., & Glowalla, U. (1997). The metrics of spatial situation models. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 23,* 622–637.

Rinck, M., & Weber, U. (2003). Who when where: An experimental test of the event-indexing model. *Memory & Cognition, 31,* 1284–1292.

Schwan, S., & Garsoffky, B. (2004). The cognitive representation of filmic event summaries. *Applied Cognitive Psychology, 18,* 37–55.

Scott Rich, S., & Taylor, H. A. (2000). Not all narrative shifts function equally. *Memory & Cognition, 28,* 1257–1266.

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

Speer, N. K., Reynolds, J. R., & Zacks, J. M. (2007). Human brain activity time-locked to narrative event boundaries. *Psychological Science, 18,* 449–455.

Speer, N. K., & Zacks, J. M. (2005). Temporal changes as event boundaries: Processing and memory consequences of narrative time shifts. *Journal of Memory and Language, 53,* 125–140.

Suh, S., & Trabasso, T. (1993). Inferences during reading: Converging evidence from discourse analysis, talk-aloud protocols, and recognition priming. *Journal of Memory and Language, 32,* 279–301.

Swallow, K. M., Barch, D. M., Head, D., Maley, C. M., Holder, D., & Zacks, J. M. (2011). Changes in events alter how people remember recent information. *Journal of Cognitive Neuroscience, 23,* 1052–1064.

Swallow, K. M., Zacks, J. M., & Abrams, R. A. (2009). Event boundaries in perception affect memory encoding and updating. *Journal of Experimental Psychology: General, 138,* 236–257.

Therriault, D. J., Rinck, M., & Zwaan, R. A. (2006). Assessing the influence of dimensional focus during situation model construction. *Memory & Cognition, 34,* 78–89.

Trabasso, T., & Magliano, J. P. (1996). Conscious understanding during text comprehension. *Discourse Processes, 21,* 255–288.

Trabasso, T., van den Broek, P., & Suh, S. (1989). Logical necessity and transitivity of causal relations in stories. *Discourse Processes, 12,* 1–25.

Trabasso, T., & Wiley, J. (2005). Goal plans of action and inferences during the comprehension of narratives. *Discourse Processes, 39,* 129–164.

van den Broek, P., Lorch, R. F., Jr., 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., Risden, K., & Husebye-Hartmann, E. (1995). The role of readers' standards for coherence in the generation of inferences during reading. In R. F. Lorch & E. J. O'Brien (Eds.), *Sources of coherence in text comprehension* (pp. 353–373). Mahwah: Erlbaum.

Whitney, C., Huber, W., Klann, J., Weis, S., Krach, S., & Kircher, T. (2009). Neural correlates of narrative shifts during auditory story comprehension. *NeuroImage, 47,* 360–366.

Zacks, J. M., Braver, T. S., Sheridan, M. A., Donaldson, D. I., Snyder, A. Z., Ollinger, J. M., & Raichle, M. E. (2001). Human brain activity time-locked to perceptual event boundaries. *Nature Neuroscience, 4,* 651–655. doi:10.1038/88486

Zacks, J. M., Kurby, C. A., Eisenberg, M. L., & Haroutunian, N. (2011). Prediction error associated with the perceptual segmentation of naturalistic events. *Journal of Cognitive Neuroscience, 23,* 4057–4066.

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

Zacks, J. M., Speer, N. K., Swallow, K. M., Braver, T. S., & Reynolds, J. R. (2007). Event perception: A mind/brain perspective. *Psychological Bulletin, 133,* 273–293.

Zacks, J. M., Speer, N. K., Swallow, K. M., & Maley, C. J. (2010). The brain's cutting-room floor: segmentation of narrative cinema. *Frontiers in Human Neuroscience, 4,* 168.

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

Zwaan, R. A. (1996). Processing narrative time shifts. *Journal of Experimental Psychology: Learning, Memory, & Cognition, 22,* 1196–1207.

Zwaan, R. A., Langston, M. C., & Graesser, A. C. (1995). The construction of situation models in narrative comprehension: An event-indexing model. *Psychological Science, 6,* 292–297.

Zwaan, R. A., & Madden, C. J. (2004). Updating situation models. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 30,* 283–288.

Zwaan, R. A., Magliano, J. P., & Graesser, A. C. (1995). Dimensions of situation-model construction in narrative comprehension. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 21,* 386–397.

Zwaan, R. A., & Radvansky, G. A. (1998). Situation models in language comprehension and memory. *Psychological Bulletin, 123,* 162–185. doi:10.1037/0033-2909.123.2.162

Zwaan, R. A., Radvansky, G. A., Hilliard, A., & Curiel, J. M. (1998). Constructing multidimensional situation models during reading. *Scientific Studies of Reading, 2,* 199–220.