(Pea)nuts and bolts of visual narrative: Structure and meaning in sequential image comprehension

Neil Cohn [a],[*], Martin Paczynski [a], Ray Jackendoff [b], Phillip J. Holcomb [a], Gina R. Kuperberg [a],[c]

[a] Department of Psychology, Tufts University, Medford, MA 02155, United States
[b] Center for Cognitive Studies, Miner Hall, Tufts University, Medford, MA 02155, United States
[c] Department of Psychiatry, Athinoula A. Martinos Center for Biomedical Imaging, Massachusetts General Hospital, Bldg 149, 13th Street, Charlestown, MA 02129, United States

A R T I C L E   I N F O

A B S T R A C T

Just as syntax differentiates coherent sentences from scrambled word strings, the comprehension of sequential images must also use a cognitive system to distinguish coherent narrative sequences from random strings of images. We conducted experiments analogous to two classic studies of language processing to examine the contributions of narrative structure and semantic relatedness to processing sequential images. We compared four types of comic strips: (1) Normal sequences with both structure and meaning, (2) Semantic Only sequences (in which the panels were related to a common semantic theme, but had no narrative structure), (3) Structural Only sequences (narrative structure but no semantic relatedness), and (4) Scrambled sequences of randomly-ordered panels. In Experiment 1, participants monitored for target panels in sequences presented panel-by-panel. Reaction times were slowest to panels in Scrambled sequences, intermediate in both Structural Only and Semantic Only sequences, and fastest in Normal sequences. This suggests that both semantic relatedness and narrative structure offer advantages to processing. Experiment 2 measured ERPs to all panels across the whole sequence. The N300/N400 was largest to panels in both the Scrambled and Structural Only sequences, intermediate in Semantic Only sequences and smallest in the Normal sequences. This implies that a combination of narrative structure and semantic relatedness can facilitate semantic processing of upcoming panels (as reflected by the N300/N400). Also, panels in

* Corresponding author. Address: Department of Psychology, Tufts University, 490 Boston Ave., Medford, MA 02155, United States. Fax: +1 617 812 4799.
E-mail address: neilcohn@emaki.net (N. Cohn).

the Scrambled sequences evoked a larger left-lateralized anterior negativity than panels in the Structural Only sequences. This localized effect was distinct from the N300/N400, and appeared despite the fact that these two sequence types were matched on local semantic relatedness between individual panels. These findings suggest that sequential image comprehension uses a narrative structure that may be independent of semantic relatedness. Altogether, we argue that the comprehension of visual narrative is guided by an interaction between structure and meaning.

1. General introduction

Drawings have been conveying narratives through sequences of images for millennia, whether painted on cave walls, carved into reliefs, hung on medieval tapestries, or, in their modern context, appearing in comic books (Kunzle, 1973; McCloud, 1993). Compared to research on the comprehension of verbal narrative, however, few studies have examined the driving forces behind our understanding of visual narrative: what are the representations and mechanisms engaged during sequential image comprehension? And, how is meaning integrated with structure across a sequence of images?

This research addresses some of these questions by examining online processing of sequential images using both reaction time (RT) measures and event-related potentials (ERPs). In this section, we first provide an overview of experimental and theoretical approaches to the comprehension of text/discourse and sequential images, showing how principles that inform research on text and discourse can also be applied to the comprehension of sequential images. We begin by discussing coherence relationships across individual sentences and individual images. We then progress to studies that extend beyond such linear relationships, examining how global narrative structure is built across text/discourse and sequential images. After this, we consider attempts to describe such a narrative structure at a theoretical level. Finally, we highlight important gaps in the existing literature, setting the stage for the current studies.

1.1. Linear coherence relationships in language and sequential images

1.1.1. Text and discourse

In early work on text and discourse, researchers emphasized how related concepts were often drawn together into common semantic fields (Bransford & Johnson, 1972; Brown & Yule, 1983; Halliday & Hasan, 1985; van Dijk, 1977) or scripts and schemas (Schank & Abelson, 1977). Structure was thought to be imposed on these general semantic fields primarily through coherence relationships between individual sentences (Halliday & Hasan, 1976; Hobbs, 1985; Kehler, 2002; Mann & Thompson, 1987).

Over the past 40 years, multiple types of coherence relationships have been described across dimensions of reference (Haviland & Clark, 1974), temporal and event structure (Mandler, 1986; Speer & Zacks, 2005; Zwaan, 1996), space (Clark, 1972; Linde & Labov, 1975; Morrow, Greenspan, & Bower, 1987), intentionality (Lichtenstein & Brewer, 1980; Schank & Abelson, 1977), and causation (Black & Bower, 1980; Keenan, Baillet, & Brown, 1984; Trabasso, Secco, & van den Broek, 1984). Indeed, a large body of psycholinguistic work now indicates that such relationships are constructed during online discourse and text comprehension to build up a "situation model"—a mental representation of the situation described in a discourse (see Zwaan & Radvansky, 1998 for a review).

1.1.2. Sequential images

Just as for verbal discourse, most work on sequential images has focused on linear relationships between individual "panels" or frames. Comic author and theorist Scott McCloud (1993) proposed a popular model of six types of linear "transitions" between panels based on changes between characters,

locations, aspects of the environment, or time shifts. McCloud's ideas have also been mapped onto concepts from the discourse literature. For example, Saraceni (2000, 2001, 2003) reconceptualized McCloud's transitions in terms of a balance between semantic fields and referential cohesion. Semantic fields describe how individual images are bound to a common meaning. For example, disparate panels of a horse's head and legs, a jockey, and spectators together convey the concept of a "horse-racing track" (Saraceni, 2000, 2001, 2003). Referential cohesion involves repeating the same characters across panels. For some transitions, semantic fields play a relatively large role in facilitating comprehension, while for others, referential cohesion plays a larger role.

There have been no experimental studies explicitly examining the psychological validity of McCloud's linear transitions as people read panels in comics. However, there have been attempts to describe the comprehension of film in terms of the types of causal, temporal or spatial coherence relationships and shifts discussed in discourse models (as in Zwaan & Radvansky, 1998). For example, Magliano, Miller, and Zwaan (2001) showed a movie to participants and asked them to record the points at which the narrative depicted changes in time, characters, or spatial location. Viewers' intuitions for these changes between film shots aligned with the experimenters' a priori coding of these shifts. Viewers were most sensitive to temporal changes, then to the changes of characters, and finally to changes in spatial location. Because these results resembled those found in analyses of verbal discourse, the authors suggested that similar mechanisms might mediate the understanding of verbal and visual narrative.

1.2. Global narrative structure in language and sequential images

1.2.1. Text and discourse

Beyond local coherence relationships linking individual sentences, people also have a sense of a *global structure* to discourse and text. Several studies of both written and auditory language show that people are better at remembering stories when the sentences are presented in a coherent order as opposed to when they are scrambled, suggesting that a sense of global structure aids comprehension (Bower, Black, & Turner, 1979; Gernsbacher, Varner, & Faust, 1990; Mandler, 1978; Mandler & DeForest, 1979; Mandler & Johnson, 1977; Stein & Nezworski, 1978).

One approach to describing this global structure has extended local coherence relationships throughout an entire narrative. For example, causal networks describe local connections not just between juxtaposed sentences or clauses, but between all units in a discourse that may have causal relationships (Schank, 1975; Trabasso & Sperry, 1985; Trabasso & van den Broek, 1985; Trabasso et al., 1984). Similar approaches have described referential relationships extending throughout a discourse (Kintsch & van Dijk, 1978). These types of extended linear relationships, however, cannot fully explain how a progression of sentences can convey overall meaning.

An approach that reaches beyond linear relationships appeals to the idea of discourse being globally organized into sub-episodes that form a global narrative structure. An experimentally-driven understanding of global *narrative* structure in discourse comes from Gernsbacher's (1990) *structure-building framework*. This model conceived of the situation model as having an intrinsic global structure, which builds up as discourse is read or heard. At the outset, readers "lay the foundation" with new information. As the discourse unfolds, coherence is mapped between sentences until a boundary of the structure is reached, at which point the process resets and repeats for the next structure (Gernsbacher, 1990). Evidence for this process comes from several studies of textual discourse. For example, self-paced reading studies have shown that the first few sentences of a discourse are read more slowly than subsequent sentences, whether at the outset of the discourse or at the start of sub-episodes of the narrative (Glanzer, Fischer, & Dorfman, 1984; Greeno & Noreen, 1974; Haberlandt, 1984; Kieras, 1978; Mandler & Goodman, 1982). This has been taken as evidence that readers have an easier time assimilating information further on in a narrative because of the foundation established at its beginning.

1.2.2. Sequential images

People also seem to have a sense of global structure in understanding sequences of images. For example, they are highly proficient at being able to reconstruct an original comic sequence when

presented with randomly-ordered panels (Nakazawa, 2004; Nakazawa & Nakazawa, 1993). Indeed, reconstruction techniques like this are widely used as a measure of "logical/sequential reasoning" in the WAIS-III test of non-verbal IQ (Kaufman & Lichtenberger, 2006). Additionally, individual images are recalled better when originally presented as part of coherent picture stories than when presented as part of scrambled picture stories (Gernsbacher, 1983, 1985).

More direct evidence that global narrative structure is used easily and intuitively during the comprehension of sequential images again comes from Gernsbacher's (1990) structure-building framework which was, in fact, originally developed using "picture stories." Notably, in her dissertation, Gernsbacher (1983, 1985) showed that seemingly continuous graphic stories can be divided up into discrete parts. A fully pictorial story from a children's book was presented to participants, who were asked to mark where they thought the episode boundaries lay in the overall story. Participants showed high agreement on where to mark such boundaries, and these were also consistent with the experimenters' intuitions for boundary locations.

Gernsbacher (1983, 1985) also showed that narrative boundaries influenced peoples' memory of the individual images constituting that story. One experiment measured the accuracy of participants' recall for particular images in a picture story. These images were either flipped (in their left–right orientation) or not. Recall for an image's left–right orientation was worse when it had originally appeared after, as opposed to before, a boundary (as defined by the segmentation process described above). In other words, crossing a boundary appeared to interfere with recall. These findings were interpreted as evidence that the structure of a sub-episode of narrative builds until a boundary is encountered, at which point encoding resets with the beginning of the next sub-episode. Similar results have been described in filmic narrative: comprehenders' recognition of individual film shots, whose physical features were manipulated, was worse when such shots appeared after rather than before a segment boundary (Carroll & Bever, 1976).

Another approach to understanding the global structure of sequential images (although not explicitly narrative structure) comes from Zacks and colleagues in their study of real-world events depicted in film (Zacks & Tversky, 2001; Zacks, Tversky, & Iyer, 2001; Zacks et al., 2001). As in Gernsbacher's studies, participants were asked to identify boundaries. However, rather than being instructed to show where boundaries lay in the narrative, participants were explicitly asked to indicate where individual events ended and began. These studies consistently showed that individuals agreed where to segment events, both at coarse-grained ("putting a sheet on a bed") and fine-grained levels ("tucking the sheets under the corners of the mattress"). Because of this multilevel evidence, Zacks and colleagues argued that the comprehension of visual events relies on a hierarchical structure.

Zacks and colleagues have also applied their approach to understanding the comprehension of larger segments of films (Zacks & Magliano, 2011; Zacks, Speer, Swallow, & Maley, 2010). Here, they showed that the boundaries between individual events can be mapped onto the types of causal, temporal and spatial coherence relationships discussed in studies of verbal text (Zwaan & Radvansky, 1998). For example, the end of one event and the beginning of the next seem to align with the linear shifts in narrative along dimensions of goals, space, causes, and locations (Zacks, Speer, & Reynolds, 2009; Zacks et al., 2010).

1.3. Theoretical approaches to narrative structure

1.3.1. Text and discourse

Gernsbacher's work shows that people reliably identify boundaries that support the existence of a global narrative structure in text and discourse. However, it leaves open the question of what such a structure might be. The idea of an organized *narrative* goes back as far as Aristotle's description of a *Beginning–Middle–End* structure in *Poetics* to describe plotlines in theatre (see Butcher, 1902). In a modern context, theories of global narrative structure entail descriptions of large-scale boundaries and sub-narrative boundaries that may not necessarily rely on linear coherence relationships across individual sentences. Of course, breaks in linear coherence (in time, space, causation, reference, etc.) may sometimes co-occur with narrative boundaries, but this may not always be the case. In other words, there is not necessarily a one-to-one mapping between linear coherence relationships and global narrative structure.

In theoretical terms, the narrative structure of text has been most clearly formalized in "story grammars," which abounded in the late 1970s and early 1980s. These models based individual narrative categories around characters' goals, which were situated within a formalistic generative grammar that imposed a global structure for understanding how characters navigated through events (e.g., Mandler & Johnson, 1977; Rumelhart, 1975; Stein & Glenn, 1979; Thorndyke, 1977). For example, various models agree that stories begin with a description of the Setting of characters, places, and the backdrop of the narrative. An Initiating event sets the narrative in motion, which results in a protagonist experiencing an Internal Response in reaction to that event. The protagonist then establishes a Goal, Attempts to achieve it, and then deals with the Outcome of those attempts. Finally, Reactions to the Outcome close out the story. This narrative progression was formalized as a canonical schema that could be embedded in segments of Episodes within the full Story.

Experimental work provided some evidence for the psychological validity of story grammars. Using memory paradigms, researchers compared the patterns of individuals' recall of stories with specific predictions made by story grammar models. Narratives were recalled better when they followed the canonical story structure than when alterations were made to it, such as when the temporal order of events was changed (Mandler & Johnson, 1977), or when the order of sentences was reversed or scrambled (Mandler, 1978, 1984; Mandler & DeForest, 1979). Furthermore, the more a canonical story structure was altered, the worse participants' performance became when asked to recall or reconstruct that story (Bower et al., 1979; Stein & Nezworski, 1978).

1.3.2. Sequential images

Cohn (in preparation) has proposed a theoretical model that formalizes the global narrative structure of sequential images. Like story grammars' treatment of sentences, this model describes panels as playing narrative roles in relation to a global sequence of images (although there are several important differences from story grammars and indeed, the model could potentially extend to describe the structure of verbal discourse, as an alternative approach to story grammars[1]). Below we give an overview of this theoretical framework, as it was used to define the structural categories used in the present set of experiments. For a more detailed description, see Cohn (in preparation).

Cohn's narrative structure consists of core *narrative categories* that play different functional roles in the sequence of an "Arc," the narrative analog to syntax's "sentence" as a maximal level structure. A sequence may open with an Establisher, which sets up referential information (i.e. introduces the characters involved) and their interrelations, but without having them engage in actions. Next, Initials may initiate an interrelation or an event, for example with a preparatory action. The Arc climaxes in a Peak, which marks the fruition or culmination of an interrelation or event in the narrative sequence. Finally, a Release may wrap up or resolve the events at the end of the sequence. Importantly, these *narrative categories* are defined not only by the content of an individual image, but also by the constraints of the rest of the sequence. In other words, sometimes the same image can function as different narrative categories, depending on what other images surround it. This is important because, according to this theory, it is possible for a sequence of images to have a narrative structure, even though its individual panels may not be related to a particular semantic field (and, indeed, may not be locally related to one another along causal, temporal, spatial or referential dimensions). Within the broader Arc, categories become grouped into narrative *constituents*—coherent pieces of a structure—which can also play narrative roles. As in syntax, this is notated by having the parts of the constituent fall under a single node of a tree. However, in this case, the nodes are not syntactic categories such as noun phrases, but are narrative categories such as Initials. This is best illustrated with an example.

Consider Fig. 1, which illustrates the narrative structure of a comic strip. This sequence shows a baseball game in which Lucy hits the ball, which leads to Charlie Brown running home to score, escap-

---

[1] Importantly, Cohn separates narrative structure and semantic information into separate components that interface with each other. This contrasts with story grammars, which often left this relationship ambiguous or defined narrative categories through semantic qualities, which indeed was a criticism of such approaches (Black & Wilensky, 1979; de Beaugrande, 1982). Such a separation is not new in the study of narrative (see, for example Brewer & Lichtenstein, 1981; Chatman, 1978; Genette, 1980; Tomashevsky, 1965), though it has not received rigorous theoretical or experimental treatment using contemporary methodologies.
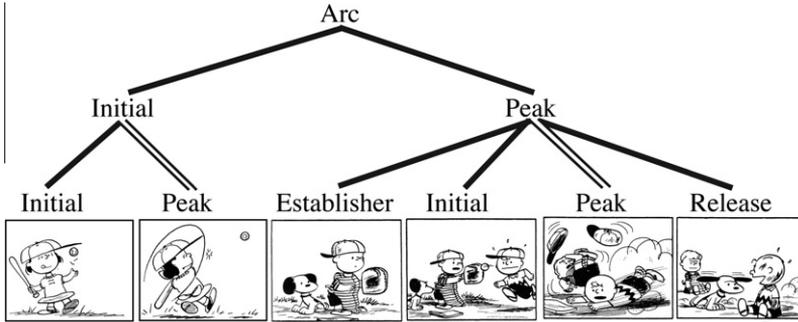
Fig. 1. Narrative structure for a sequence of images.

ing a tag by Schroeder. It begins with a narrative constituent of two panels, in which the Initial shows Lucy tossing a ball that she hits in the Peak. The second narrative constituent follows a canonical pattern, as a subordinate constituent to a larger Arc. It begins with a set-up of Schroeder waiting for the ball—nothing happens here except for the expectation that something may eventually occur (Establisher). The second panel of the subordinate constituent Initiates that event, but the penultimate Peak panel then interrupts the event of catching the ball with Charlie sliding into the base. This panel features the greatest narrative tension in the strip. Finally, the last panel features the Release of this tension, providing a resolution. The first narrative constituent (Lucy hitting the ball) facilitates the second (Charlie scoring) and thus is understood as an Initial at a higher level of structure. This Initial of the first constituent sets up the second constituent as a Peak.

This narrative structure does not appear in Fig. 2b, in which the individual panels are still related to the overall semantic field of baseball. In this sequence, cues related to baseball recur across panels
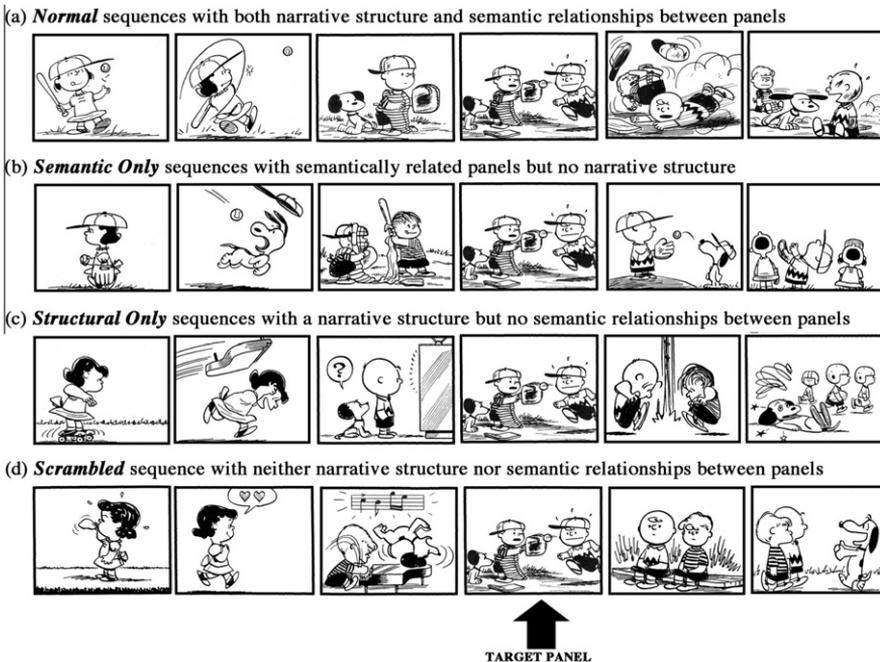

Fig. 2. Image sequences that manipulate the contributions of semantic relatedness and narrative structure.

(hats, balls, bats, gloves), just as they do for fully coherent narrative sequence in Fig. 2a. In Fig. 2d, however, no common semantic field or narrative structure binds the individual images. Importantly, however, according to this theory, Fig. 2c does have such a narrative structure, even though its individual panels do not belong to a common semantic field, and even though there are also no causal or temporal coherence relationships between individual pairs of panels. Fig. 2c has a sense of a narrative progression though, moving from a setup to a rising of tension, and then climaxing before presenting an aftermath. Indeed, if Fig. 2c is compared panel-by-panel with the Normal sequence in Fig. 2a, their panels share the same placement of narrative tension. The first panels both set up the sequence followed by second panels that set forth an initiating action. Panel 3 shifts to characters in relatively passive states. The fourth panels (the target panel) are identical, both showing the start of an interaction. The penultimate panels both show the culmination of some type of event, while the final panels feature a resolution. However, despite sharing these features, Fig. 2a makes sense and 2c does not—this sequence has the feel of a narrative structure, despite the fact that the individual panels do not semantically cohere.

Narrative structure, as discussed in Cohn's theory, is not considered a complete substitute for linear coherence relationships between individual panels. Rather, this structure interfaces with linear relationships in systematic ways, although not necessarily in a one-to-one fashion. For instance, narrative structural boundaries often coincide with major changes in linear causal, referential, or spatial domains, similar to the boundaries of events (e.g., Newtson & Enquist, 1976). In Fig. 2a, the first panel shows Lucy tossing a baseball in the air, about to hit it with the bat. This preparatory action provides the cue for this panel to play the role of an Initial in narrative structure. The second panel then maintains referential coherence with the first panel by showing all the same characters and objects. But in addition, there is a causative relationship between these panels: Lucy now hits the ball with the bat, completing the action prepared in the prior panel. This culminating action signals the Peak in narrative. This causative relation does not fall at a narrative structure boundary. Rather, together, these panels form a coherent whole that acts as an Initial in the broader Arc. At the third panel, a boundary of narrative structure is reached and a new narrative constituent begins, aligning with the major referential shift between Lucy in panel two and Schroeder with Snoopy in panel three (as depicted fully in Fig. 1). In this panel, the introduction of new characters in a passive state makes it recognized as an Establisher. Since Establishers often begin an Arc, it thereby signals a new narrative constituent along with the referential shift. The next panel now maintains the same characters and adds a new character: Charlie Brown. The expectations created by the characters' actions cue this as an Initial: Schroeder is *about to* catch a ball while Charlie is *about to* collide with him. However, these actions are not causally connected to the preceding Establisher—they just follow them temporally. The fifth panel is the Peak, indicated by the culmination of Charlie's collision with Schroeder, a clear causal connection to the preceding panel. This panel referentially changes only slightly from the previous one, in that Snoopy is no longer seen. Finally, the last panel shows all three characters again and returns to a state of relative passivity, where Snoopy's determination of "safe" cues that this is a Release. This panel also is causally related to the Peak as a resolution of its actions, a different causal relation than that between Peak and Initial, which was a culmination. Thus, overall, in this sequence the major referential shift between panels two and three does indicate a boundary between narrative constituents. However, minor referential shifts (the adding of Charlie to the sequence, the omission and then reemergence of Snoopy) and the temporal and causal shifts do not align with narrative boundaries—they fall within narrative constituents.

Cohn's theory also draws several broad analogies between how narrative categories combine to form a narrative structure of a sequence of images and how grammatical categories combine to form a syntactic structure of a sentence (see Cohn (in preparation) and Appendix A for a brief discussion of some of these analogies). It is important to emphasize, however, that the nature, rules, and constraints of narrative structure are quite different from those of syntactic structure. Also, the type and quantity of information encoded within a narrative category differs from that encoded in a syntactic category. Most individual panels do not encapsulate information that maps easily onto word-level lexical content. Rather, as Fig. 2 illustrates, panels mostly encode information closer to what is conveyed by a whole sentence or clause in a discourse. Nonetheless, these broad analogies between structure at the level of sentences and structure at the level of whole narrative inspired the design of the two stud-

ies presented here: we exploited methodology previously used to examine the build-up of syntactic structure in sentences to study the build-up of narrative structure across sequential images.

1.4. Outstanding questions

In sum, existing work suggests that sequential image comprehension draws upon a combination of relatedness to a semantic field, linear coherence relationships of various types across individual panels, and global narrative structure. There is also general agreement that structure builds up across a coherent sequence of images (Gernsbacher, 1983, 1985), and that this does not seem to occur in scrambled image sequences (Gernsbacher et al., 1990). However, much of this evidence comes from offline or memory tasks in which participants are asked to arrange scrambled sequences (Nakazawa, 2004; Nakazawa & Nakazawa, 1993), to define breaks in a seemingly continuous discourse (Gernsbacher, 1983, 1985), or to recall the content of a narrative that has been disrupted (Gernsbacher, 1985; Gernsbacher et al., 1990; Mandler & Johnson, 1977; Stein & Nezworski, 1978; Thorndyke, 1977). These methods, however, tell us little about the online processing of sequential images, as comprehenders build up overall meaning, image by image.

Moreover, as has been discussed in studies of verbal discourse (van Dijk & Kintsch, 1983), a drawback of using memory paradigms to explore comprehension is that the contents of people's recall tends to retain the semantics while losing memory for structural components. The use of memory tasks to validate story grammars has been criticized for this very reason (Black & Wilensky, 1979; de Beaugrande, 1982).

Finally, since the materials used in many previous studies compared only fully scrambled sequences with normal narrative sequences, it is unclear what aspects of comprehension might have motivated their experimental results: Semantic relatedness? Narrative structure? By analogy, contrasting normal sentences with scrambled words would tell us that the order of words in a sentence matters, but it would tell us little about the relative contributions of syntax and semantics to comprehension.

Therefore, fundamental questions remain about the comprehension of sequential images: Are the narrative categories discussed in Cohn's theory psychologically valid? Is it possible to detect the build-up of narrative structure and semantic relationships during online image-by-image comprehension of sequential images? Is it possible to detect the build-up of narrative structure independently of relatedness to a particular semantic field? Addressing these questions requires online measures that extend beyond memory paradigms and which control for the contributions of semantic relatedness and narrative structure.

This study presents two experiments to investigate the contributions of semantic relatedness and narrative structure during the panel-by-panel comprehension of sequential images. We identified semantically related panels that connected to a broader semantic field or topic, based on cues within their individual images. For example, a strip depicting the semantic field of "baseball" included individual panels showing characters carrying out actions related to baseball (e.g., throwing a ball, hitting a ball), or other related elements (e.g., hats, gloves, balls, bats). We used Cohn's theory to define narrative categories that built up the narrative structure. This enabled us to fully cross semantic relatedness and narrative structure to create four types of novel *Peanuts* comic strips: (1) Normal sequences with both semantic relatedness and narrative structure (Fig. 2a); (2) Semantic Only sequences with panels which were related through a common semantic field, but which had no narrative structure, (Fig. 2b); (3) Structural Only sequences which had a narrative structure but had no common semantic field (Fig. 2c), and (4) Scrambled sequences without any narrative structure or any semantic field (Fig. 2d).

This crossing between meaning and narrative can be conceived of as analogous to the crossing between semantic relatedness and syntactic structure at the sentence-level. A Normal sequence is broadly analogous to a normal sentence, such as "Lucy is tossing a baseball in order to hit it." A Semantic Only strip (with no structure) is analogous to a sentence with semantically related words but no syntax, e.g., "Out balls gloves throw pitcher bats running safe catch." A Structural Only strip is analogous to Chomsky's (1965) "Colorless green ideas sleep furiously," and a Scrambled sequence is analogous to "Ideas in tossing order colorless furiously balls." This analogy allowed us to design studies

that exploited two types of online methodologies that have previously been used to examine the relationships between semantics and structure within sentences, but in this case to examine the contribution of structure and meaning to the comprehension of sequential images.

Experiment 1 measured reaction times as participants monitored specific target panels within a sequence. Experiment 2 examined event-related potentials (ERPs), which gave insights into the neurocognitive processes engaged in comprehending sequential images. To anticipate our conclusions, our findings suggest that both semantic relatedness and narrative structure play independent and combined roles in panel-by-panel sequential image processing.

2. Experiment 1: Target monitoring

2.1. Introduction

As reviewed in Section 1, Cohn's theory claims that graphic narrative comprehension relies on a narrative structure that is distinct from, but functions alongside the relatedness of panels through a common semantic field. The psychological reality of this theoretical distinction, however, has not yet been explored during online processing. We examined this by drawing upon a classic online psycholinguistic paradigm originally used by Marslen-Wilson and Tyler (1975, 1980) to investigate syntax and semantics in sentence processing.

In a seminal study, Marslen-Wilson and Tyler (1975, 1980) asked participants to monitor for target words (e.g., "ideas") in normal sentences (e.g., "The boy's ideas formed silently"), syntactic-but-not-semantic sentences (e.g., "Colorless green ideas sleep furiously"), and randomly scrambled sentences (e.g., "Picnic strike ideas quiet launched"). They showed that reaction times to target words increased across these three sentence types. These findings were taken to support the view that syntactic structure influences online sentence comprehension, even in the absence of semantic information, with processing maximally facilitated by the presence of both semantics and syntax. This study further showed that, within sentences containing some syntactic structure (with or without semantics), reaction times became progressively faster to target words positioned further along in the sentence. Scrambled sentences, in contrast, showed no decreasing trend in reaction times across target word position. These findings suggested that the build-up of structure across a sentence progressively facilitates the processing of target words.

The present study analogously measured reaction times as comprehenders monitored for target panels in graphic sequences. Narrative structure and semantic relatedness were independently manipulated to yield four sequence types: (1) Normal sequences were like a standard comic strip, balancing both semantic relatedness and narrative structure (Fig. 2a); (2) Semantic Only sequences featured panels related through a common semantic field, but had no narrative structure, such as disparate images of characters playing baseball, which did not make sense as a whole (Fig. 2b); (3) Structural Only sequences used a narrative Arc, based on the model of narrative grammar described at the end of Section 1 (Cohn, in preparation), but displayed no coherent semantic relationships between individual panels, analogous to the syntactic "Colorless green ideas…" sentences (Fig. 2c); (4) Scrambled sequences used randomly-ordered panels without semantic relatedness or structural connections between panels (Fig. 2d). The Normal sequences, Structural Only sequences and Scrambled sequences were modeled on the analogous conditions used by Marslen-Wilson and Tyler (1975, 1980). The Semantic Only sequences were added as a further control.

If a narrative structure guides comprehension, this would predict that, as in Marslen-Wilson and Tyler (1975, 1980), participants would respond fastest to target panels in Normal sequences: the combination of both semantic relatedness and narrative structure in the preceding context would facilitate processing of the target. Participants should be slowest to respond to panels in Scrambled sequences because no expectations are built on the basis of either semantic relatedness or structure. Reaction times to target panels in the Semantic Only and Structural Only sequences should both fall between those of Normal and Scrambled conditions. Moreover, if a narrative structure builds up incrementally during comprehension, then, in those sequences containing structure (i.e. Normal and Structural Only sequences), target panels appearing late in the sequence should be read faster than those appearing at

the beginning. In contrast, in sequences where there is no structure (i.e. Scrambled and Semantic Only sequences), there should be no decreases in reaction time to monitor panels at later versus earlier ordinal positions.

2.2. Methods

2.2.1. Construction of stimuli

Graphic sequences were created using black and white panels scanned from the *Complete Peanuts* volumes 1 through 6 (1950–1962) by Charles Schulz (Fantagraphics Books, 2004–2006). *Peanuts* comics were chosen because (1) they have systematic panel sizes and content with repeated characters and situations; (2) their content is recognizable to most people; (3) there is a large corpus of sequences to draw from; and (4) they feature fairly consistent and recurrent themes (various sports, building snowmen, Lucy skipping rope, Linus and Snoopy fighting over a blanket, etc.).

Two hundred coherent six-panel long Normal sequences were initially created. Standard daily *Peanuts* strips are four panels long. We created novel six-panel experimental sequences by combining panels from existing comic strips, to ensure that any familiarity with specific *Peanuts* strips would not lead to episodic memory effects on the Normal sequences. In order to eliminate any influence of written language on comprehension, we used panels without text, or panels where text was deleted in Adobe Photoshop. All panels were adjusted to a single uniform size.

Normal sequences related to a common semantic theme: baseball, football, golf, piano playing, kite flying, weather (snow, rain, sunshine), seasonal activities (making snowmen, throwing snowballs, leaves falling off trees, etc.), and others. Semantic relatedness was based on cues within the individual panels. For example, a strip depicting the semantic field of "baseball" included individual panels showing characters carrying out actions related to baseball (e.g., throwing a ball, hitting a ball), or other related elements (e.g., hats, gloves, bats, pitching mound). Normal sequences also all had a coherent narrative structure, as defined by Cohn's theory. Across these Normal sequences, the pattern of narrative structure was varied so as not to bias the experimental results toward only one structural pattern (such as the basic canonical pattern of *Establisher-Initial-Peak-Release*).

To ensure that these novel six-panel Normal sequences were indeed fully coherent, they were presented, together with 119 longer filler sequences (Sunday *Peanuts* sequences which are longer than standard 4-panel strips) to seven Tufts University undergraduates (mean age of 19.14; 5 male, 2 female) who were familiar with *Peanuts* comic sequences and who were paid for their participation. All participants viewed all the sequences, which were randomized in a different order for each person. They rated each sequence for "how easy it is to understand" on a scale of 1 (hard to understand) to 7 (easy to understand). Nine of these Normal sequences were deemed too difficult to understand and were excluded. Each of the resulting 191 novel Normal 6-panel sequences was used to create the three additional experimental conditions.

In the Semantic Only sequences, each panel shared the same overall semantic field, but had no coherent narrative structure. Lacking this structure, the panels of these sequences could hypothetically be rearranged with no effect on the overall meaning. These sequences were constructed by assigning panels to semantic fields based on cues within their images, as described above. In each Semantic Only sequence, the semantic field matched that of its corresponding Normal sequence. In the example shown in Fig. 2b, the target panel relates to the overall semantic field of "baseball," and so its other panels repeat this theme (the characters appear in disparate facets of the game) but with no sense of structure or order across panels.

In the Structural Only sequences, panels retained the same global narrative structure as the Normal sequences, but they did not share a common semantic field. These sequences were created by two researchers trained in the criteria for the narrative structure outlined by Cohn's (in preparation) theory of narrative structure. The sequences used the same narrative categories in the same global narrative patterns as those used in corresponding Normal sequences. However, the panels in Structural Only sequences were drawn from numerous different strips that did *not* share a common semantic theme (no repetition of particular actions carried out by characters, and no elements within the images linking them to a particular semantic field). For example, an Establisher panel from one strip might start a sequence, while a Release from an unrelated strip may be used to end it, based on the

pattern of the corresponding Normal sequence. These Structural Only sequences did not show causal coherence between individual panels. Fig. 2c shows such a Structural Only sequence that matches the narrative structure of Fig. 2a. It starts with an Initial of Lucy skating along the street, followed by a Peak of Lucy dodging a thrown piano. An Establisher then shows a passive Charlie Brown and Snoopy watching TV, and then another Initial (the target) of Schroeder preparing for a baseball impact. The penultimate panel is again a Peak, here of Charlie Brown and Linus dodging an object crashing down from the sky, after which the sequence ends with a Release depicting Snoopy dizzy and various boys walking away. Thus, this Structural Only sequence matches the narrative structure of its corresponding Normal sequence, but without matching to the theme of "baseball" in the target, and without any semantic theme binding the panels.

Finally, the Scrambled sequences used neither narrative structure nor a common semantic field. These sequences combined disparate panels that did not share semantic fields with each other and also did not follow a coherent narrative progression. In the example shown in Fig. 2d, the panels feature different themes and no narrative Arc, leaving nothing at all to unite them.

Within each quadruplet, the same "target panel" appeared at the identical position (highlighted as panel 5 in Fig. 2). The target panels appeared in the second to the sixth panel positions, with equal numbers of targets at each position. Within each quadruplet, all four sequence types were matched on the number of characters per panel in each panel position, and on which characters appeared across panels to the best extent possible. That is, if the Normal sequence in a quadruplet featured Snoopy and Charlie Brown in the first two panels, and Snoopy and Lucy in the final four panels, the matching experimental sequence types attempted to retain Snoopy and Charlie Brown in the first two panels and Snoopy and Lucy in the final four. If the exact characters could not be matched (due to limitation of panel selection in counterbalancing), the same number of characters was matched. Thus, in all cases, the number of characters per panel was constant across positions between sequence types within each quadruplet.

In total, 191 quadruplets of six-panel sequences were generated (764 sequences in total). Further examples of these sets can be viewed in the Appendix.

Three further objective measures of coherence were used to further examine and constrain the stimulus set, as described below.

2.2.2. Objective measures of coherence

We first carried out two rating studies to examine global and local coherence of the strips. Participants in rating studies were recruited either from Tufts University or through the Internet via links on the first author's website (www.emaki.net). In these rating studies, participants' comic reading fluency was assessed using a pretest questionnaire that asked participants to self-rate the frequency with which they read various types of comics (comic books, comic strips, graphic novels, Japanese comics, etc.), read books for pleasure, watched movies, and drew comics, both currently and while growing up. These ratings were measured using a scale of 1–7 (1 = never, 7 = always), and the questionnaire also gauged their self-assessed "expertise" at reading and drawing comics along a five-point scale (1 = below average, 5 = above average). A "fluency rating" was then computed using the following formula:

$$(\text{Mean comic reading frequencies} \times \text{Comic reading expertise}) + \left(\frac{\text{Comic drawing frequency} \times \text{Drawing ability}}{2}\right)$$

This formula weighted fluency towards comic reading comprehension, giving an additional "bonus" for fluency in comic production. Most participants' fluency fell between the idealized average (a score of 12) to high (22), but, as noted below, the mean fluency across all ratings studies was high. Self-defined "comic readers" were chosen in order to reduce the heterogeneity in the population and to ensure that participants were familiar with the materials and this manner of assimilating sequential pictures. In addition to this assessment, all participants were asked first whether they knew *Peanuts* (yes/no) and how familiar they were with them (self-rated on a 1–5 scale).

The rating studies were hosted online using www.surveymonkey.com. Each sequence or pair of panels was presented as a whole and participants were able to progress through the sequences at their

own pace. Written consent was given by all participants in accordance with Tufts University guidelines. Participants were compensated for their participation.

2.2.2.1. Ratings of global coherence. Forty "comic-fluent" individuals with a mean age of 29.5 (29 male, 11 female) participated. Their mean comic-reading fluency was high, at 22.11 (SD = 11.35). All knew *Peanuts*, with a mean familiarity of 3.75. 191 quadruplets comprising the four sequence types were counterbalanced, using a Latin Square design, across four lists (10 participants per list), ensuring that each participant viewed only one sequence type of a quadruplet, but across all lists (and participants), the same target panel appeared in all four sequence types. The sequence types were randomized within lists. Participants were asked to rate the sequences on a scale of 1–7 for how much they made sense as a whole.

2.2.2.2. Ratings of local coherence. Another study examined the local coherence of immediately adjacent panel pairs within the sequences. One-hundred participants (mean age of 38.76, 76 male, 24 female) took part. Their comic reading fluency was again high at 24.91 (SD = 11.35), and again they had a high familiarity with *Peanuts* (M = 4.05). For each of the 191 quadruplets of six-panel sequences, a set of five pairs of panels composed of the immediately adjacent images from the sequences was constructed: 1–2, 2–3, 3–4, 4–5, 5–6. These panel pairs were then counterbalanced across 20 lists, such that only one pairing from each quadruplet was shown to an individual participant (i.e. 191 pairs per list). This ensured that no panel in a pairing was repeated in a list. However, across participants, all pairings for a given panel were viewed. Participants were asked to rate the panel pairs along a 1–7 scale for how "related in meaning" they were.

2.2.2.3. Assessment of referential coherence. As noted above, within each quadruplet, the four sequence types were matched for the number of characters per panel in each position, as well as which characters appeared across panels to the best extent possible. In addition, we quantified *changes* in characters between each adjacent pair of panels in all sequences. We coded a full change of characters between two panels as a "1", no change as "0", and a partial change (characters held constant but others added or omitted) as ".5."

2.2.3. Construction of final set of stimuli

The global and local coherence ratings studies described above were used to constrain the selection of the final set of 160 quadruplets. High global ratings were desired for the Normal sequences, while low global ratings were desired for the three other sequence types. High local ratings were desired for Normal and Semantic Only sequences to reflect their local semantic connections, while relatively low local ratings were required for the Structural Only and Scrambled sequences. A given quadruplet was included in the final stimulus set only if three of the following criteria were met: (1) Normal sequences: mean global and local ratings greater than 6; (2) Semantic Only sequences: mean global and local ratings greater than 4; (3) Structural Only sequences: mean global ratings less than 3.2 and mean local ratings less than 3; (4) Scrambled sequences: mean global ratings less than 1.8 and local ratings less than 1.7.

This process truncated the 191 stimuli in the ratings studies down to a final set of 160 quadruplets. Table 1 shows the mean global coherence ratings, the mean local coherence ratings (for all adjacent pairs of panels) and the mean local coherence ratings for target panels with their immediately preceding panel, for each sequence type in this final set of stimuli. An ANOVA revealed significant differences across the four sequence types in global ratings, $F(3,477) = 7492.1$, $p < .001$, with all pairwise comparisons between sequence types significantly different from one another (all $ts > 31$, all $ps < .001$). There were also significant differences across the four sequence types in local ratings across all pairs of panels, $F(3,477) = 8036.95$, $p < .001$, again with significant differences in all pairwise comparisons (all $ts < 26$, all $ps < .001$). Similar statistics were found when isolating ratings for target panels and their immediately preceding panels: there were overall differences across the four sequence types, $F(3,297) = 302.9$, $p < .001$, and all pairwise relationships were also significant (all $ts > 6.9$, all $ps < .001$).

In Table 2, we summarize our local referential coherence measure—the *changes* in characters between adjacent pairs as well between the target panel and its immediately preceding panel—for each

Table 1
Global and local ratings of semantic coherence across panels.

Global	Normal 6.13 (.72)		Semantic Only 3.56 (1.1)		Structural Only 2.66 (.73)		Scrambled 2.39 (.69)	
Positions	All	Targets	All	Targets	All	Targets	All	Targets
1–2	5.80 (.93)	6.09 (1.2)	4.44 (1)	3.72 (1.6)	3.05 (.99)	2.71 (1.5)	2.54 (1.1)	2.31 (1.3)
2–3	5.54 (1)	5.39 (1.7)	3.21 (1)	3.44 (1.6)	2.48 (.93)	2.86 (1.8)	4.94 (.86)	2.4 (1.3)
3–4	5.59 (.94)	5.69 (1.4)	2.56 (1)	3.89 (1.7)	4.70 (.78)	2.81 (1.4)	3.59 (.94)	2.55 (1.5)
4–5	5.66 (.94)	5.75 (1.3)	4.05 (.76)	3.47 (1.5)	3.78 (.91)	2.94 (1.5)	2.65 (.99)	2.43 (1.4)
5–6	5.40 (.96)	5.15 (1.8)	4.19 (1.0)	3.53 (1.5)	2.84 (.96)	2.64 (1.4)	2.46 (1.1)	2.36 (1.4)
Ave local	5.59 (.83)	5.61 (1.5)	3.57 (.83)	3.61 (1.6)	2.68 (.91)	2.79 (1.5)	2.46 (.95)	2.41 (1.4)

Mean ratings for semantic coherence for each sequence type (on a scale of 1–7). Global ratings are for each type of sequence as a whole. Local ratings are shown for all adjacent pairs of panels as well as more specifically for Targets and their immediately preceding panel.

Table 2
Rates of change of characters across paired panel positions.

Positions	Normal		Semantic Only		Structural Only		Scrambled	
	All	Targets	All	Targets	All	Targets	All	Targets
1–2	.15 (.27)	.13 (.22)	.40 (.41)	.34 (.43)	.21 (.29)	.19 (.28)	.26 (.36)	.31 (.42)
2–3	.23 (.32)	.38 (.36)	.43 (.42)	.58 (.42)	.28 (.33)	.41 (.35)	.33 (.37)	.44 (.35)
3–4	.23 (.34)	.20 (.28)	.4 (.40)	.31 (.35)	.26 (.35)	.25 (.31)	.30 (.39)	.25 (.34)
4–5	.14 (.29)	.13 (.28)	.39 (.40)	.33 (.41)	.18 (.28)	.17 (.27)	.26 (.38)	.19 (.30)
5–6	.23 (.31)	.28 (.33)	.41 (.39)	.44 (.38)	.26 (.32)	.30 (.31)	.31 (.36)	.28 (.31)
Ave local	.20 (.19)	.22 (.31)	.41 (.29)	.40 (.41)	.24 (.19)	.26 (.31)	.29 (.23)	.29 (.35)

Mean rates of change of characters across paired panel positions for all sequence types. Complete change of character was coded as "1", no change between panels as "0", and partial change (character added or subtracted while others maintained) as ".5". Mean rate of change is listed for all panel pairings as well as ratings for just Targets with their preceding panels. Standard deviations are shown in parentheses.

sequence type in the final stimulus set. An ANOVA showed significant differences across the four sequence types in local referential coherence between all pairs of panels, $F(3, 477) = 48.7$, $p < .001$, and for the target panel and its immediately preceding panel, $F(3, 477) = 17.04$, $p < .001$. Follow-ups comparing each sequence type with one another revealed significant differences for all panels (all $ts > 3.3$, all $ps < .005$), and for the target panel and its immediately preceding panel (all $ts > 2.2$, all $ps < .05$), except for between the Structural Only and Scrambled targets, $t(159) = -1.59$, $p = .114$.

Stimuli in the final set were then counterbalanced using a Latin Square design across four lists, each to be seen by an individual participant. This allowed participants to view only one sequence type of a quadruplet with a given target panel, but ensured that, across all lists (and participants), the same target panel would appear in all four sequence types. This resulted in 160 sequences (40 sequences of each sequence type, 8 at each position per type) per list. To each list, 80 additional filler sequences were added. These used longer sequences from 7 to 11 panels long in order to prevent participants from using a strategy of anticipating the target on panel 6—the final panel of all experimental stimuli. These fillers included existing coherent *Peanuts* Sunday sequences, which were already longer than average daily comic sequences, as well as expanded sequences of scenarios that had been rejected from the experimental stimuli after the ratings studies. Of the fillers, 30 were Normal, 10 Semantic Only, 10 Structural Only, and 30 were Scrambled. With both experimental and filler sequences, half the sequences contained semantic relationships (Normal, Semantic Only) and half did not contain semantic relationships (Structural Only, Scrambled), with the proportion of sequence types being 30% Normal, 20% Semantic Only, 20% Structural Only, and 30% Scrambled. Within each list, the order of experimental and filler sequences was randomized.

2.2.4. Participants in the panel monitoring experiment

Fifty-four experienced comic readers (30 male, 24 female) were recruited from the Tufts University student population. Their mean age was 20.4 (SD = 1.68), and they were paid for participation. All participants gave their informed written consent according to the Tufts University Human Subjects Review Board's guidelines. Based on the pretest questionnaire (see above), participants who were included in the study had a mean comic-reading fluency of 13.89 (SD = 6.81). All participants knew *Peanuts*, and had an average familiarity rating of 3. Data from two participants were discarded because they did not reach a threshold of 80% accuracy in the task.

2.2.5. Procedure

Participants sat in front of a computer screen where a target panel was presented first, followed by a sequence that contained the target. Strips were presented panel-by-panel using in-house software. Each trial began with a black screen reading READY in gray lettering. When the participant pressed the keypad, a fixation-cross appeared at the center of the screen for 1500 ms with a 300 ms interstimulus interval (ISI), followed by the target panel. The target panel remained on screen for 2500 ms, allowing participants to examine its features carefully. Following another 300 ms ISI, another fixation-cross appeared for 1500 ms and this was followed by the main sequence of panels. Each successive panel remained on the screen for 1500 ms with an ISI of 300 ms. The 1500 ms duration was used because a pilot self-paced comic reading study showed that this was the average time spent at each individual panel in normal 4 panel sequences; it was also the duration used by West and Holcomb (2002) in their previous ERP study on sequential images. The 300 ms ISI prevented the appearance of the sequences turning into a "flip-book" style animation. At the end of the sequence, a screen reading READY again appeared before the next trial.

Participants' task was to press a button as soon as they recognized the target in the sequence. Reaction time was measured to the target, time-locked to the onset of its presentation. In addition, after 25 sequences, randomly distributed across the experiment, a comprehension question asked about various properties of the sequence (e.g., "Was Snoopy scared?", "Did Snoopy swallow the ball?", etc.). These questions aimed to encourage participants to comprehend the sequences, rather than just looking for the physical features of the target panels.

Prior to the experiment itself, participants practiced with a list of 10 sequences. Throughout the main experiment, five breaks were given at designated intervals. After the experiment, a post-test questionnaire asked participants to reflect on the nature of the sequences to see if they were consciously aware of any specific patterns or characteristics of the sequences they had viewed.

2.2.6. Analysis of data

Accuracy for button presses was computed as the percentage of trials in which participants pressed the button at the appropriate target panels. For trials in which participants responded with multiple button presses, the first press was counted as valid. Incorrect responses were either omissions or false presses to panels other than targets. Analysis of the reaction times (RTs) used correctly-answered responses. In each participant, outlier reaction times—more than 2.5 standard deviations from the mean response within a given condition—were discarded.

Accuracy and RTs to correctly-answered responses were analyzed using a 4 (Sequence Type) × 5 (Position) repeated-measures ANOVA. Main effects of Position were followed up using polynomial contrasts to determine whether there were overall linear trends across ordinal position (collapsed across all Sequence Types). Main effects of Sequence Type were followed up using paired *t*-tests comparing each Sequence Type with one another (collapsed across Position). Interactions between Sequence Type and Position were followed up in two ways. First, we used polynomial contrasts to determine whether there were linear trends across each Position for each Sequence Type (we only report the trends that interacted with the linear component of the interaction between Sequence Type and Position). Second, at each Position, we carried out four-way ANOVAs to examine the effects of Sequence Type.

Finally, the effect of comic reading fluency was examined on (a) RTs to each sequence type and (b) differences in RTs between the specific pairs of sequences that probed the use of the narrative structure, i.e. RT differences between the Normal and Semantic Only sequences and between the Scrambled

and the Structural Only sequences. These analyses used a Pearson's correlation, set to an alpha level of .05.

2.3. Results

2.3.1. Accuracy

Participants' panel monitoring accuracy across all sequence types was 93%. Accuracy for the Normal sequences was worst (90% correct; mean: 36 out of 40 (SD = 2.90)), compared with 94% for Semantic Only (37.75, SD = 2.39), 94% for Structural Only (37.65, SD = 2.27), and 94% for Scrambled (37.6, SD = 2.45) sequences. A 4 (Sequence Type) × 5 (Position) ANOVA on accurate button presses showed significant main effects of Sequence Type, $F(3, 153) = 8.14$, $p < .001$. This main effect was followed up using paired t-tests which showed that, while accuracy to the Normal sequences was significantly worse than to all other sequence types (all $ts < -3.2$, all $ps < .005$), there were no other significant differences between any other pair of sequences (all $ts < .33$, all $ps > .190$).

The overall ANOVA showed no main effect of Position, $F(4, 204) = .845$, $p = .498$. However, there was a significant interaction between Sequence Type and Position, $F(12, 612) = 2.11$, $p < .05$, and polynomial contrasts to this interaction revealed significant linear trends, $F(1, 51) = 4.92$, $p < .05$, but not quadratic, cubic, or quartic trends (all $Fs < 1.4$, all $p > .240$). We followed up this interaction by examining the trends in accuracy across ordinal position for each sequence type individually. Normal sequences showed a significant linear trend to become less accurate as the sequence progressed, $F(1, 51) = 9.28$, $p < .005$ (the quadratic and cubic trends were not significant). The Semantic Only, Structural Only and Scrambled sequences showed no significant linear trends across sequence position (although the Structural Only sequences showed quadratic trends, $F(1, 51) = 10.81$, $p < .005$, and the Scrambled sequences showed cubic trends, $F(1, 51) = 4.64$, $p < .05$).

Greater differences in accuracy between Sequence Types were found as the sequence progressed. At Position 2 there were no significant differences in accuracy across Sequence Types, $F(3, 153) = .708$, $p = .549$, while differences at Position 3 only trended towards significance, $F(3, 153) = 2.49$, $p = .063$. Significant differences in accuracy across Sequence Types were, however, seen at Positions 4 through 6 (all $Fs > 3.8$, all $ps < .05$). At these positions, Normal panels differed significantly (or near significantly) from all other Sequence Types (all $ts < -1.96$, all $ps < .055$), but there were no differences between any other pair of Sequence Types (all $ts < 1.2$, all $ps > .196$).

Finally, the post-test questionnaires indicated that most participants noticed a distinction between the Scrambled and Normal strips. 46% of participants explicitly commented that the Semantic Only sequences featured "themes" of meaning, though no participant picked up on any difference between the Structural Only and the Scrambled strips. (Of note, 11 out of the 54 participants made no explicit comments indicating that they picked up on traits of the sequences.)

2.3.2. Reaction times for target monitoring

An overall 4 (Sequence Type) × 5 (Position) repeated-measures ANOVA showed a main effect of Sequence Type that reached significance, $F(3, 153) = 7.29$, $p < .001$, a main effect of Position, $F(4, 204) = 36.76$, $p < .001$, and a significant interaction between Sequence Type and Position, $F(12, 612) = 2.19$, $p < .05$.

The main effect of Sequence Type arose because targets in the Normal sequences were recognized fastest, while targets in the Scrambled sequences were recognized slowest, with RTs to targets in the Structural Only and Semantic Only sequences falling in between (see Fig. 3). Planned pairwise comparisons using paired t-tests confirmed this pattern of findings: faster RTs appeared to panels in the Normal strips than to panels in all other sequence types (all $ts < -3.2$, all $ps < .005$). RTs to targets in the Semantic Only sequences were shorter than in the Scrambled sequences (approaching significance, $t(51) = -1.99$, $p = .051$), and RTs in the Structural Only sequences were also shorter than in the Scrambled sequences, $t(51) = -2.14$, $p < .05$. However, RTs to targets in the Structural Only and Semantic Only sequences did not differ, $t(51) = .133$, $p = .895$. Because RTs were fastest in the Normal sentences where participants were least accurate, we examined the possibility of a speed-accuracy trade-off by correlating participants' accuracy and reaction times in these sentences: this correlation was negative,
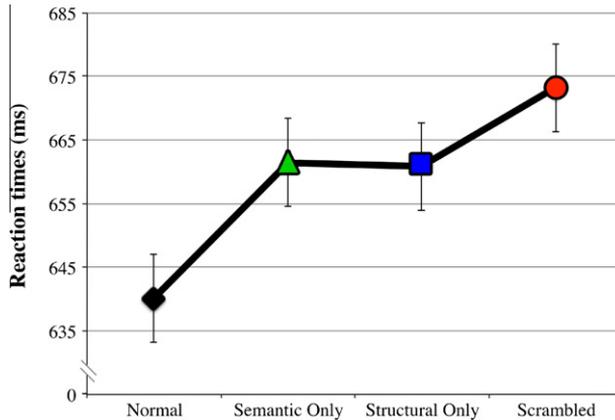
Fig. 3. Mean reaction times to target panels in the four sequence types, collapsed across linear positions. Error bars represent standard error.

$r(52) = -.295$, $p < .05$, i.e. the greater participants' accuracy, the shorter their reaction times. There was therefore no evidence for such a trade-off.

The main effect of Position reflected the fact that, across all Sequence Types, participants were slower to detect targets early in the sequence than those later in the sequence (see Fig. 4). Polynomial contrasts confirmed this, and revealed linear effects of Position, $F(1,51) = 138.26$, $p < .001$, but no quadratic, cubic, or quartic trends (all $Fs < .48$, all $p > .490$).

The Sequence Type × Position interaction in the overall ANOVA reflected differences between the four sequence types in the rate that RTs decreased across linear position, as illustrated in Fig. 4. This was confirmed by the fact that polynomial contrasts showed a significant interaction for linear trends across sequence position, $F(1,51) = 7.183$, $p < .05$, but no quadratic, cubic, or quartic effects (all $Fs < .38$, all $ps > .540$). We followed up this interaction with polynomial contrasts looking at the trends across linear position for each Sequence Type individually. All sequence types showed linear trends across position (all $Fs > 6.4$, all $ps < .05$), but no significant quadratic, cubic, or quartic trends (all $Fs < 2.31$, all $ps > .137$). RTs to Normal sequences decreased the least across position; the Semantic Only and Structural Only sequences had steeper slopes, and Scrambled sequences showed the most rapid decrease in RTs across Position.

We also followed up the main Sequence Type by Position interaction by carrying out four-way ANOVAs between Sequence Types at each Position. Significant main effects of Sequence Type were
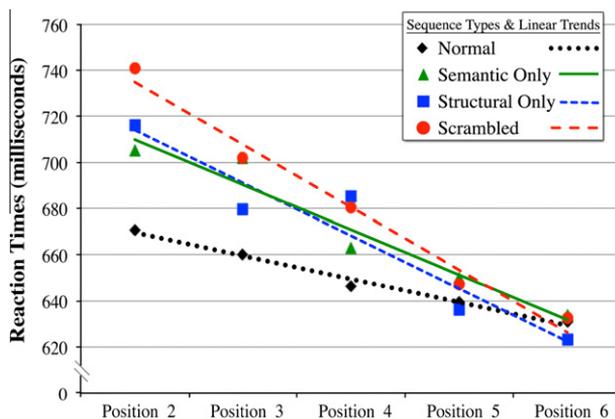

Fig. 4. Mean reaction times for panels in Sequence Types across ordinal positions.

found at Position 2, $F(3,153) = 7.86$, $p < .001$, Position 3, $F(3,153) = 4.21$, $p < .01$, and Position 4, $F(3,153) = 3.41$, $p < .05$. However, no significant effects were found at Positions 5, $F(3,153) = .394$, $p = .757$, or Position 6, $F(3,153) = .296$, $p = .828$. Paired t-tests at Positions 2, 3, and 4 compared each Sequence Type at each panel position. At Position 2, RTs differed between Normal sequences and all other types (all $ts < -2.2$, all $ps < .05$), and between Semantic Only and Scrambled sequences, $t(51) = -2.32$, $p < .05$. At Position 3, RTs differed between Normal sequences and Semantic Only, $t(51) = -2.69$, $p < .05$, and Scrambled sequences, $t(51) = -2.84$, $p < .05$. The differences between Structural Only and Semantic Only approached significance, $t(51) = 1.76$, $p = .085$, as did the difference between Structural Only and Scrambled sequences, $t(51) = -1.93$, $p = .059$. At Position 4, Normal sequences differed significantly from Structural and Scrambled sequences (all $ts < -2.6$, all $ps < .05$).

2.3.3. Relationships with comic reading fluency

Correlations were carried out between each individual's comic reading fluency and their mean RTs in each Sequence Type. A significant negative correlation was found between Fluency and RTs in all non-Normal sequence types (all $r < -.27$, all $p < .05$), but not for the Normal sequences, $r(52) = -.14$, $p = .336$. That is, RTs to Normal sequences were not affected by expertise, but for more fluent comprehenders, RTs were faster in the Semantic Only, Structural Only, and Scrambled sequences. No significant correlation was found between Fluency and the difference in RTs between Scrambled and Structural Only or Normal and Semantic Only sequence types.

2.4. Discussion

In this study, RTs were measured as viewers monitored for a target panel while comprehending four different types of image sequences, one panel at a time. As predicted, RTs on target panels in the Normal sequences were fastest while those in Scrambled sequences were the slowest. RTs to target panels in both Semantic Only and Structural Only sequences fell directly in between the RTs in the Normal and Scrambled sequences and did not significantly differ from one another. This gradation in reaction times for target panels across Normal, Structural Only, and Scrambled sequences mirrors the pattern of RTs observed by Marslen-Wilson and Tyler (1975, 1980) for target words in studies using spoken sentences. Additionally, we observed a decrease in RTs across ordinal sequence positions. Unlike in Marslen-Wilson and Tyler (1975, 1980), however, this decrease was seen in all sequence types—not only the ones with structure. Each of these findings will be addressed in more detail below.

2.4.1. Contrasts between sequence types

In the studies by Marslen-Wilson and Tyler (1975, 1980), reaction times were faster to target words in Normal sentences than in Syntactic Only sentences. This was taken to support the idea that syntactic structure acts together with semantics to construct the overall meaning of sentences. Experiment 1 shows an analogous result, with critical panels in the Normal sequences showing faster reaction times than those in Structural Only sequences. Since the latter sequence type features a narrative structure without a common semantic field across panels, it analogously suggests that narrative structure and semantic relatedness combine during panel-by-panel comprehension of Normal sequential images. There is, however, an alternative interpretation: since the Normal and Structural Only sequences were distinguished *only* by the presence/absence of an overall semantic theme, any facilitation of reaction time to targets in the Normal sequences might have reflected semantic priming between pairs of panels, rather than the combinatory effects of structure and meaning. Because of this, we added the control of the Semantic Only condition, which was not used in the Marslen-Wilson and Tyler studies. The Normal and Semantic Only sequences differed primarily in the presence (Normal) and absence (Semantic Only) of narrative structure. Yet, faster reaction times were still seen to target panels in the Normal than the Semantic Only sequences, suggesting that comprehenders did, indeed, use structural information over and above simple semantic relatedness during comprehension.

The most compelling evidence that participants were using a global narrative structure during comprehension comes from their faster reaction times to critical panels in the Structural Only sequences than in the Scrambled sequences. These once again differed primarily in the presence (Structural Only) and absence (Scrambled) of narrative structure. Neither type of sequence, however, contained many

local semantic relations between the individual panels. They also did not contain many causal or temporal relationships between individual panels. This finding therefore suggests that such a global narrative structure was being used during sequential image comprehension, regardless of semantic relatedness or other types of linear coherence relationships between panels. Interestingly, during feedback after the experiments, no participant reported that they noticed any difference between the Scrambled and Structural Only sequence types. This suggests that participants may have been using this narrative structure implicitly during processing, rather than through a conscious strategy.

Our aim in developing the stimuli was for the panels in both the Normal and Semantic Only sequences to share a common semantic field (e.g., a baseball theme) but for the Structural Only and Scrambled sequences not to share a common semantic field. In contrast, we constructed stimuli such that the Normal and Structural Only sequences both had a global narrative while the Semantic Only and Structural Only sequences had no such structure. We have interpreted our findings as reflecting the influences and interactions between the semantic field and global narrative structure. However, it is important to consider the possibility that our results were driven not by these global factors, but rather by local semantic relationships and/or local referential relationships across pairs of panels. We examined this possibility by coding these local relationships, as discussed below.

In a first set of rating studies, we asked participants to rate the semantic relatedness of individual pairs of adjacent panels for all sequence types (in addition to rating the semantic relatedness of the entire strips). Although we narrowed down our final stimulus set based on cut-offs of these local and global relatedness ratings, we were unable to fully match the Normal and Semantic Only panels, or the Scrambled and Structural Only panels on these ratings. It seems unlikely, however, that the remaining differences in relatedness ratings fully explain our findings as they did not fully mirror the pattern of monitoring times across the four conditions. First, the differences in both local and global semantic relatedness between the Normal and Structural Only sequences were much larger than between the Normal and Semantic Only sequences. Yet, the reaction time differences for each of these contrasts were the same. Similarly, the difference in relatedness ratings between the Scrambled and Structural Only sequences, although significant, was very small (0.27 for global ratings and 0.22 for local ratings, on a 1–7 scale), and much smaller than the difference between the Scrambled and Semantic Only sequences (1.26 for global ratings, 1.11 for local ratings). Yet, again, the reaction time differences for each of these contrasts were the same. In other words, if local semantic relationships were the primary factor driving the reaction times, then monitoring for target panels in the Semantic Only sequences should have been faster than in the Structural Only sequences, which was not the case.

It also seems unlikely that differences in the numbers of characters across panels, or differences in referential relationships across pairs of panels, can explain our findings. In constructing our stimuli, all four sequence types were matched for the number of characters per panel in each panel position, as well as for which characters appeared across panels. In addition, we quantified *changes* in characters between adjacent pairs of panels in all sequences. Here, there were some differences across the four sequence types, but, as with the relatedness ratings, they did not mirror the pattern of monitoring times. For example, Semantic Only sequences showed the most character changes between panels—twice as many changes as in the Normal sequences. Meanwhile, although significant, Structural Only sequences had only a few more character changes than the Normal sequences. However, reaction times did not differ between the Semantic Only and Structural Only sequences. If local referential relationships were driving differences in reaction times, we would have expected monitoring times to panels in the Semantic Only sequences to be the slowest, or to at least to be slower than in the Structural Only sequences. Furthermore, there was no significant difference in local referential relationships between targets and their preceding panels in the Structural Only and Scrambled sequences. As argued above, this was the critical contrast for probing the use of global narrative structure in the absence of semantic relationships. Our finding of a significant difference in monitoring times in this contrast, despite referential relationships being matched, supports this argument.

2.4.2. Effects of panel position

The main finding across panel position was a decrease in monitoring times along the course of the sequence. The slower reaction times early in the sequence are consistent with previous research using

both verbal discourse structure (Glanzer et al., 1984; Greeno & Noreen, 1974; Haberlandt, 1984; Kieras, 1978; Mandler & Goodman, 1982), and picture sequences (Gernsbacher, 1983), and has been hypothesized to reflect a laying of a foundation for the discourse (Gernsbacher, 1990). Notably, however, this decrease in reaction times across panel position was true for all sequence types. This contrasts with the sentence-level findings of Marslen-Wilson and Tyler (1975, 1980) who described a decrease in word monitoring times only for sentences with structure, but not for scrambled sentences. Marslen-Wilson and Tyler interpreted their findings as reflecting the facilitation of the processing of incoming words by the build-up of syntactic structure over the course of a sentence. On the basis of the present findings, we cannot make an analogous interpretation that a narrative structure, built up over the course of the image sequence, facilitated the processing of incoming images. Nonetheless, before coming to the opposite conclusion, it is important to consider differences between the comprehension of spoken language and sequential images, as well as differences in how the monitoring task itself might have influenced comprehension in each of these domains.

Monitoring a word within a sentence requires few resources: it takes relatively little effort to encode and retain a well-learned word in memory, and to recognize it after the delay of a few words. In addition, sentence comprehension is implicit and "automatic," making it easy to distinguish between grammatical sentences and scrambled words. Thus, in Marslen-Wilson and Tyler's experiment, it is likely that the monitoring task itself was just as easy at the beginning of a sentence as at the end of a sentence, and is unlikely to have interfered with sentence comprehension. This means that reaction times are likely to have directly reflected the influence of structure and/or semantic relatedness, and the influence of these factors on monitoring times would be maximal when more context was built up, i.e. towards the end of the sentence.

Individual images, on the other hand, are more complex than words and, for the most part, are not conventionalized (Cohn, 2007). They are therefore likely to have required more resources than words to encode, retain within working memory, and recognize during the monitoring task. This intuition is supported by the longer overall reaction times in the present experiment (more than 600 ms) than in Marslen-Wilson and Tyler's study (less than 300 ms). As the sequences progressed, however, monitoring is likely to have become easier in all sequence types, with participants becoming increasingly prepared to find a target image with the diminishing number of possibilities for its location, thereby leading to reaction times diminishing across linear position for all sequence types. While we attempted to reduce this anticipation by including filler sequences that were longer than the six-panel experimental sequences, this is unlikely to have fully prevented this anticipation strategy. In addition, comprehension itself is more difficult for sequential images than for sentences, meaning that reaction time measures are more likely to be susceptible to the effects of the task. Thus, at the beginning of a sequence, the effects of context had a clear impact. However, as the sequence progressed to later sequential positions, the effects of the task probably outweighed any effects of context on comprehension. Indeed, by the end of the sequences, participants may have abandoned attempts at comprehension altogether, explaining why reaction times converged across the four sequence types, perhaps to the fastest speed at which they could perform the task (a ceiling effect).

If aspects of the task were indeed confounding findings across sentence position, this could be overcome by using a more direct measure of online processing that does not rely either on a superimposed task or the measurement of a behavioral response. Overcoming this limitation was one aim of Experiment 2 in which we measured neural activity to all panels, without requiring participants to monitor for specific targets.

3. Experiment 2: Event related potentials

3.1. Introduction

Experiment 1 used a panel-monitoring paradigm to show differences between reaction times for target panels in Normal, Semantic Only, Structural Only, and Scrambled sequence types. This behavioral paradigm offered valuable insights. However, as discussed above, the task of monitoring for specific target panels may have interfered with the comprehension of the panel sequences, and reaction times are a somewhat indirect measure of processing. In order to overcome these limitations, we

carried out a second experiment using event related potentials (ERPs)—a technique with excellent temporal resolution that directly indexes underlying neural processes and which, in principle, is not dependent on a behavioral task.

3.1.1. ERPs and language processing

In language, the ERP component that has been most closely associated with semantic processing is the N400. In their seminal studies, Kutas and Hillyard (1980, 1984) identified this component as a negative deflection in the waveform, peaking at around 400 ms, which was smaller (less negative) in amplitude to words that were semantically congruous than those that were semantically incongruous or unexpected with their preceding contexts. The amplitude of the N400 is attenuated by featural overlap or semantic relatedness between an incoming word and its preceding context, whether this context be a single word prime (Bentin, McCarthy, & Wood, 1985; Rugg, 1984), a sentence context (Federmeier & Kutas, 1999; Kutas & Hillyard, 1984), or a global discourse context (Federmeier & Kutas, 1999; Kuperberg, Paczynski, & Ditman, 2011; St. George, Mannes, & Hoffman, 1997; van Berkum, Hagoort, & Brown, 1999; Yang, Perfetti, & Schmalhofer, 2007). The modulation of N400 amplitude to semantic congruity is termed the "N400 effect" and, for word stimuli, it typically localizes over centro-posterior sites (Kutas & Van Petten, 1994).

In an important study, Van Petten and Kutas (1991) showed that an N400 is evoked by *all* meaningful open-class words within sentences. However, its amplitude decreases to successive words, suggesting that, as context is built up throughout a sentence, semantic processing of each upcoming word is progressively facilitated. Critically, however, just as in the behavioral monitoring study by Marslen-Wilson and Tyler (1975, 1980), no such progressive attenuation of the N400 with increasing ordinal position was seen in sentences with structure but no semantics (syntactic only: "Colorless green ideas..."), or in random strings of words (scrambled sentences). Also, similar to Marslen-Wilson and Tyler (1975, 1980), Van Petten and Kutas (1991) showed that the N400 evoked by a given open-class word (collapsed across all word positions) was smaller in coherent sentences than in structural only or scrambled sentences. These findings were critical in establishing that the N400 is not simply a response to semantic anomalies or semantically unexpected stimuli, but rather that it reflects default semantic processing of all meaningful stimuli, which is reduced when lexico-semantic information and syntactic structure combine to build up a congruous context. Importantly, however, unlike Marslen-Wilson and Tyler (1975, 1980), Van Petten and Kutas (1991) observed no differences between the N400 amplitude evoked by open-class words in scrambled and structural only sentences. This suggested that the N400 to open-class words was not sensitive to structure alone, in the absence of semantics.

Taken together, all these findings suggest that the N400 to open-class words reflects a process of relating the meaning of an incoming word with its preceding context and with information stored within semantic memory (Kuperberg, 2007; Kutas & Federmeier, 2011). Although the amplitude of the N400 is influenced by the combination of structure and semantics, it does not itself *directly* reflect either structural processing or the process by which structure and meaning of an incoming item are combined with the context (for discussion, see Lau, Phillips, & Poeppel, 2008; van Berkum, 2009). Rather, it is thought to reflect the semantic processing of the incoming item itself, which can be facilitated by the combination of structure and meaning in the preceding context.

There are two ERP components thought to be sensitive to structural processing in language. First, a Late Positivity or P600 waveform—a centro-parietally distributed positive deflection, peaking from 600 to 800 ms—is evoked by syntactic ambiguities in garden path sentences (Osterhout & Holcomb, 1992), and by syntactic violations (Hagoort, Brown, & Groothusen, 1993) after a normal structure has been built up. The P600 also appears to be sensitive to semantic information. For example, the P600 effect to syntactic violations is larger when the context is more (versus less) semantically constraining (Gunter, Stowe, & Mulder, 1997). Additionally, Münte, Matzke, and Johannes (1997) reported that agreement violations between nouns and verbs in "Jabberwocky" sentences made up of nonsense words failed to evoke a P600 effect (although Hahne and Jescheniak (2001) did find a P600 when pseudoword sentences were presented auditorily). It has been proposed that the P600 evoked by syntactic violations reflects additional combinatorial analysis that is maximal in the presence of conflict

between the violation of structure and a possible alternative representation constrained by the semantic context (Kuperberg, 2007).

The second waveform that has been described in association with structure is a *left anterior negativity* (LAN), falling between 300 and 500 ms (although it can be prolonged), and distributed over frontal, often left-lateralized, electrode sites. The LAN has been tied to a number of syntactic operations, but was first described in association with phrase structure violations inverting prepositions and nouns such as "Max's of proof the theorem" (as opposed to "Max's proof of the theorem") (Neville, Nicol, Barss, Forster, & Garrett, 1991).

In contrast to the P600, the LAN has been observed to syntactic violations in Jabberwocky sentences (Münte et al., 1997). This suggests that, unlike the P600, it is sensitive to violations of syntactic structure, even when there is no build-up of semantic context (Friederici, 2002). It is unclear, however, whether the LAN effect is a response to syntactic violations following the build-up of a well-formed syntactic structure, or whether it can be evoked by words in scrambled sentences in the absence of any build-up of syntactic structure. Van Petten and Kutas (1991) did find that between 300 and 500 ms, closed-class function words within scrambled sentences evoked a larger negativity than in structural only sentences. Unlike the N400 to open-classed words, this negativity to closed-class words did not vary across sentence position. However, it is unclear whether this negativity effect had a left anterior distribution because only a single left anterior electrode site (F3) was included in the montage.

3.1.2. ERPs and image processing

The N400 is not only linked to processing the meaning of words; a similar waveform is also evoked by non-linguistic meaningful stimuli, including static images. The N400 evoked by images often has a more anterior distribution than that evoked by words (Ganis, Kutas, & Sereno, 1996; McPherson & Holcomb, 1999), but, like the N400 evoked by words, it is smaller to target pictures that are semantically congruous than incongruous with their preceding context. Context may constitute a preceding single picture in priming paradigms (McPherson & Holcomb, 1999), a surrounding scene (Ganis & Kutas, 2003) or a verbal sentence (Ganis et al., 1996).

Along with this N400, a preceding negativity, peaking at around 300 ms after the onset of the stimulus, and therefore termed the "N300," has also been observed to picture images (Barrett & Rugg, 1990; Ganis et al., 1996; McPherson & Holcomb, 1999). This N300 often overlaps with the N400 and once again has a more frontal distribution than the N400 observed in most language studies (McPherson & Holcomb, 1999). Its earlier onset is thought to reflect a more rapid access to the semantic features of objects than symbolic words (Sitnikova, West, Kuperberg, & Holcomb, 2006).

The first study to examine the comprehension of *sequential* images came from West and Holcomb (2002) who asked participants to distinguish congruous from incongruous final panels in image sequences taken from animations. The authors observed a smaller N300/N400 complex on final panels that were congruous versus incongruous with their preceding image sequence, indicating that semantic processing of the final image was facilitated when its preceding visual narrative was semantically consistent. In line with studies using static images, this N400 effect had a more anterior distribution than the linguistic N400 effect (peaking at right centro-frontal electrode sites), but it was still more widespread and lateralized than its preceding anteriorly distributed N300. It also had a longer duration than the N400 observed in language studies.

Another set of studies found a similar anteriorly-distributed N300/N400 to incongruous endings to short, silent movie clips depicting everyday events (e.g., a man ironing a shirt following a context in which he prepares to cut bread) (Sitnikova, Holcomb, & Kuperberg, 2008b; Sitnikova, Kuperberg, & Holcomb, 2003). Interestingly though, an additional posterior positivity between 500 ms and 800 ms appeared when a final scene depicted an event in which the central action predicted by the context was carried out with an object that did not possess the required semantic properties to carry out this action (such as a man attempting to cut bread with an iron) (Sitnikova et al., 2008b, 2003). This was interpreted as somewhat analogous to the P600 effect in language, once again reflecting additional processing in response to a conflict between the predicted action (constrained by the semantic context) and the violation between the semantic properties of the object and the depicted action (Sitnikova, Holcomb, & Kuperberg, 2008a).

Taken together, these studies indicate that, just as in language, comprehenders are able to use the context of sequential visual images to influence semantic processing of an upcoming image. There is also prolonged processing when the depicted image violates expectations about central actions and events. It remains unclear, however, what representations are used to build such expectations across sequential images. Does global narrative structure act in combination with semantic fields to influence semantic processing of upcoming images? And, as in language, is context built additively across sequence positions?

The present study aimed to address these questions. A different set of participants viewed the same stimuli as used in Experiment 1 but this time, rather than only examining target panels, we measured ERP activity to all panels in the sequence, mirroring the study design and logic of the language study reported by Van Petten and Kutas (1991), described above.

We first aimed to examine the pattern of N300/N400 modulation across the four sequence types (collapsed across panel position). The semantic priming studies described above show that the N300/N400 complex evoked to visual images is sensitive to semantic relatedness with the preceding context. Thus we expected that the amplitude of the N300/N400 would be smaller to panels in sequences sharing a common semantic field (the Normal and Semantic Only sequences) than in those without a common semantic field (the Structural Only or the Scrambled panels). Our main question was how the amplitude of the N300/N400 would be modulated in the Normal sequences relative to the Semantic Only sequences, and in the Structural Only sequences relative to the Scrambled sequences. If the build-up of visual semantic context depends on the combination of narrative structure and semantic relatedness across individual panels, this would predict that the N300/N400 would be smaller to panels in the Normal sequences than in the Semantic Only sequences. However, given that the N300/N400 is thought not to reflect structural/semantic combination per se, there should be no difference in the amplitude of the N300/N400 between panels in Structural Only and Scrambled sequences, just as in Van Petten and Kutas' study of language.

We also aimed to examine the pattern of ERPs across linear position in each of the four sequence types. As discussed, the panel-monitoring task in Experiment 1 may have interfered with comprehension, particularly towards the end of the sequence. This interference may have led to a convergence on similar reaction times at later panel positions that masked any reaction time differences across conditions. In the current ERP study, there were no dual task demands: participants were asked to view the panel sequences and, at the end of each sequence, to decide whether or not it made sense. Based on the sensitivity of the N300/N400 to semantic relatedness, we expected some decrease in N300/N400 amplitude across panels in the Semantic Only sequences, although it was unclear whether this decrease would be seen across all sequence positions. Again, our main question was whether narrative structure would further facilitate processing across the sequence, over and above semantic relatedness. If narrative structure and semantic relatedness combine to progressively build context across sequence position, this would predict a relatively greater decrease of the N300/N400 with increasing ordinal position in the Normal sequences than in the Semantic Only sequences. Given the insensitivity of the N300/N400 to structure alone, we predicted no such decrease across sequence position in the sequences without structure (the Structural Only or Scrambled sequences).

Despite the manipulation of structure, we did not expect to see modulation on the P600 component. As noted above, during language processing, the P600/Late Positivity is most likely to be produced when the context builds up a coherent syntactic and semantic context that is subsequently violated (Kuperberg, 2007; Münte et al., 1997). Analogously, during visual event comprehension, the P600 is triggered when the context sets up expectations of a goal-directed action whose structure is then violated (Sitnikova et al., 2008b, 2003; Sitnikova, Paczynski, & Kuperberg, 2007). In the present study, however, none of the visual sequences built a narrative structure that was subsequently violated. The Structural Only sequences built a narrative structure in the absence of semantic relationships, while the Semantic Only and Scrambled sequences did not build any structure at all.

We did not, however, rule out the possibility of finding a LAN effect that was sensitive to structure, in the absence of semantic relatedness, as probed by the contrast between the Scrambled and Structural Only sequences. As noted above, in language, a LAN is sensitive to structure without semantics: it is evoked by violations in otherwise grammatically felicitous sentences constructed with nonsense words (Münte et al., 1997). And Van Petten and Kutas (1991) did find a difference in a negative wave-

form to closed-class words between syntactic only and scrambled sentences which did not show the same characteristics of the N400 found to open-class words. Finally, a right anterior negativity has been reported to violations of musical structure (Koelsch, Gunter, Friederici, & Schröger, 2000; Patel, Gibson, Ratner, Besson, & Holcomb, 1998), providing a precedent for an anterior negativity in a domain outside of language. Thus, we did consider the possibility that a negativity, distinct from the N300/N400, might be seen in the comparison between the Structural Only and Scrambled sequences.

3.2. Methods

3.2.1. Participants

Twenty-four Tufts University undergraduates with a mean age of 19.4 (SD = 1.67)—12 males and 12 females—participated in the ERP study for compensation. Each participant gave informed written consent according to the guidelines of the Tufts University Institutional Review Board. Participants were pre-screened to be English-speaking comic readers with normal vision, no history of head trauma, and taking no neuropsychiatric drugs. All participants completed the comic fluency questionnaire (described under Experiment 1) and had a mean fluency rating of 16.99 (SD = 6.36). They all knew *Peanuts*, with an average familiarity overall (mean: 2.9 on a 1–5 scale).

3.2.2. Stimuli

The same four lists of counterbalanced sequences were used as in Experiment 1. However, fillers for the ERP experiment were changed to 80 Normal Sunday sequences. This resulted in 50% of the sequences being coherent and 50% being violated in some way. These fillers were also altered to be six panels long. The experimental and filler stimuli were randomized within lists.

3.2.3. Procedure

Participants sat in a comfortable chair across from a computer screen in a room separate from the experimenter and computers. Lights were kept on to avoid a "flashing" effect of the white panels appearing on the black screen (as this tended to induce blinks). Trials began with the word, READY, which remained on the screen until the participant pressed a button on a keypad. A fixation cross then appeared in the center of the screen for 1500 ms, followed by a 300 ms ISI, and then the first panel of the sequence appeared, centered on the screen. Each panel remained on the screen for 1500 ms with an ISI of 300 ms. An ISI of 500 ms followed the last frame, after which a question mark appeared. This cued participants to decide whether the sequence they just saw "made sense." This question was answered by pressing "yes" or "no" buttons on a keypad with either their left or right thumb, counterbalanced across lists. In 25 randomly interspersed sequences, after making the coherence judgment, participants were asked to answer additional questions about the meaning of the sequence (e.g., "Was Snoopy scared?") in order to ensure that they comprehended the sequences. Questions were designed to address events described in individual panels, meaning they applied equally to both felicitous and anomalous sequences.

A practice list of 10 sequences preceded the actual experimental trials to acclimate participants to the procedure and stimuli.

3.2.4. ERP recordings

ERPs were measured using an elastic cap with 29 tin electrodes distributed along the scalp according to the International 10–20 system plus additional sites over the left and right hemispheres, along with electrodes below the left eye and next to the right eye to record blinks and vertical and horizontal eye movements. Electrode sites were placed along the five midline sites (FPz, Fz, Cz, Pz, Oz), three pairs of medial sites (FC1/FC2, C3/C4, CP1/CP2), four pairs of lateral sites, (F3/F4, FC5/FC6, CP5/CP6, P3/P4), and five pairs of peripheral sites (FP1/FP2, F7/F8, T3/T4, T5/T6, O1/O2) on each hemisphere. All electrodes were referenced to an electrode placed on the left mastoid, while differential activity was monitored on the right mastoid.

A SA Bioamplifier amplified the electroencephalogram (EEG) using a bandpass of 0.01–40 Hz and continuously sampled at a rate of 200 Hz. Electrode impedances were kept below 10 kΩ for the eyes and below 5 kΩ at all other sites.

3.2.5. Behavioral data analysis

Accuracy for participants' judgments was computed as the percentage of correct responses for each Sequence Type. Correct responses were those in which participants responded that the Normal and Filler sequences "make sense" but all other Sequence Types "don't make sense." Participants were excluded if their accuracy was less than 80%.

3.2.6. ERP data analysis

ERPs evoked by all panels were examined. In an initial set of ANOVAs, we collapsed across linear position in order to simplify the analysis and its interpretation, i.e. linear position was not included as a within-subject variable. Following West and Holcomb (2002), analyses of mean voltages were conducted within the windows of 300–400 ms, 400–600 ms, and 600–900 ms to investigate the N300, N400, and sustained negativity effects, respectively.

ANOVAs were carried out along the midline column (FPz, Fz, Cz, Pz, Oz), and at lateral regions constituting five pairs of electrodes along both the left (F7/F3, FC5/FC1, T3/C3, CP5/CP1, T5/P3) and right (F8/F4, FC6/FC2, T4/C4, CP6/CP2, T6/P4) hemispheres, see Fig. 5 for montage. For the midline ANOVA, within-subject factors were the four levels of Sequence Type, and five levels of Anterior–Posterior (AP) Distribution corresponding to the five midline electrodes. For the lateral ANOVA, within-subject factors were the four levels of Sequence Type, two levels of Hemisphere, and five levels of AP Distribution corresponding to the five pairs of electrodes over each hemisphere, and Electrode Site, corresponding to individual electrode sites within each pair. These omnibus ANOVAs were followed up with midline and lateral ANOVAs that compared each Sequence Type with one another. In the midline ANOVAs, further interactions between Sequence Type and AP Distribution were followed up with paired *t*-tests examining the effects of Sequence Type at each electrode site. In the lateral ANOVAs, follow-ups at individual sites were carried out only when there was a higher-order interaction involving Sequence Type, Electrode Site, Hemisphere and/or AP Distribution.

As noted, the analyses described above did not include Sequence Position as a within-subject variable (to simplify presentation and interpretation). In order to determine how ERPs varied across sequence position, a more targeted analysis was carried out using a single dependent measure. For each of the N300, N400, and late negativity time-windows, ERPs to panels at each position in each sequence type were averaged across 14 anterior-central sites: frontal (F7, F8, Fz, F4, F8), centro-frontal (FC5, FC1, FC2, FC6) and central-temporal (T3, C3, Cz, C4, T4) regions, see Fig. 5. This is where effects were max-
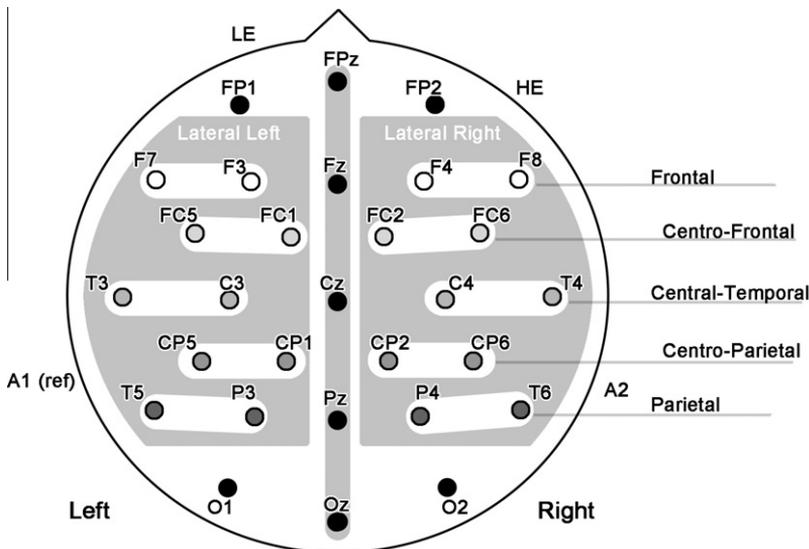

Fig. 5. Electrode montage, illustrating midline column and lateral regions for analysis of ERP data.

imal. These average values were the dependent measure in a 4 (Sequence Type) × 6 (Position) repeated-measures ANOVA. Follow ups to Sequence Type × Position interactions used polynomial contrasts to determine whether there were linear trends across ordinal positions in each sequence, as described in Experiment 1.

Finally, as in Experiment 1, participants' ratings of comic reading fluency were correlated with differences in ERPs between the Normal and Semantic Only sequence types and between the Scrambled and Structural Only sequence types. These ERP differences scores were averaged across the same centro-frontal 14 electrode sites described above, in each individual, and were then correlated with each individual's comic fluency rating using a Pearson's correlation set to an alpha level of .05.

3.3. Results

3.3.1. Behavioral data

Participants were most accurate in judging the acceptability of the Normal (M = .92, SD = .08) and Filler (M = .92, SD = .07) sequences, less accurate in judging the Scrambled (M = .89, SD = .24) and Structural Only (M = .87, SD = .25) sequences, and least accurate in judging the Semantic Only sequences M = .70, SD = .28). A four-way ANOVA found a main effect of Sequence Type, $F(3,69)$ = 8.80, $p < .001$, and follow-up pairwise comparisons confirmed that participants were least accurate in classifying the Semantic Only sequences as "not making sense" (significantly lower accuracy than each other sequence type, all ts > 3.29, all ps < .005). There were no significant differences in accuracy between the Normal and Structural Only, and between the Normal and Scrambled sequences (all ts < .390, all ps >.40). Overall, 75% of the comprehension questions were correctly answered across participants.

As in Experiment 1, the post-test questionnaires showed that most participants noticed the difference between the Scrambled and Normal strips, while 42% of participants explicitly commented that the Semantic Only sequences featured "themes" of meaning. However, no participant indicated any explicit awareness that the Structural Only sequences differed from the Scrambled strips.

3.3.2. ERP data

3.3.2.1. Effects of sequence type on ERPs, averaged across all panels.

3.3.2.1.1. 300–400 ms. Averaged across all panels, a significant negative deflection was observed starting between 250–300 ms and lasting until 400 ms, with a peak around 300 ms after the onset of the stimulus panel at some sites—the N300 component (Figs. 6 and 7). This was modulated across the sequence types, and a repeated measures ANOVA comparing the four sequence types (collapsed across ordinal position) showed significant main effects of Sequence Type, and/or interactions with AP Distribution, Hemisphere, and Electrode Site at the midline column and lateral regions (see Table 3). These analyses were followed up with ANOVAs comparing each sequence type with one another.

The N300 amplitude was smaller in the Normal sequences than in the Semantic Only sequences. The effect was widespread (main effects of Sequence Type in both midline and lateral regions analyses) but maximal at centro-frontal sites (Sequence Type × AP Distribution interactions in both analyses), see Figs. 6 and 7, and Table 3. The interaction between Sequence Type and AP Distribution in the midline analyses and a three-way interaction between Sequence Type, Hemisphere, and Electrode Site in the lateral analysis were followed up with paired *t*-tests, which revealed significant effects at all sites except Oz, T5, and T6.

The N300 was, in turn, smaller in the Semantic Only sequences than in the Structural Only sequences. This effect was again widespread (main effects of Sequence Type in both the midline and lateral analyses), but maximal at anterior sites (interactions between Sequence Type and AP Distribution at the midline column), especially on the right (interactions between Sequence Type, and Hemisphere, and between Sequence Type, Hemisphere, AP Distribution and Electrode Site in the lateral analysis), see Figs. 6 and 7, and Table 3. Follow-up *t*-tests showed significant effects at all sites (except for FPz where the effect approached significance).

The comparison between the Structural Only and Scrambled sequences revealed a smaller and less widespread effect. Main effects of Sequence Type appeared only in the lateral analysis, although there was a Sequence Type by AP Distribution interaction in both midline and lateral analyses. The effect
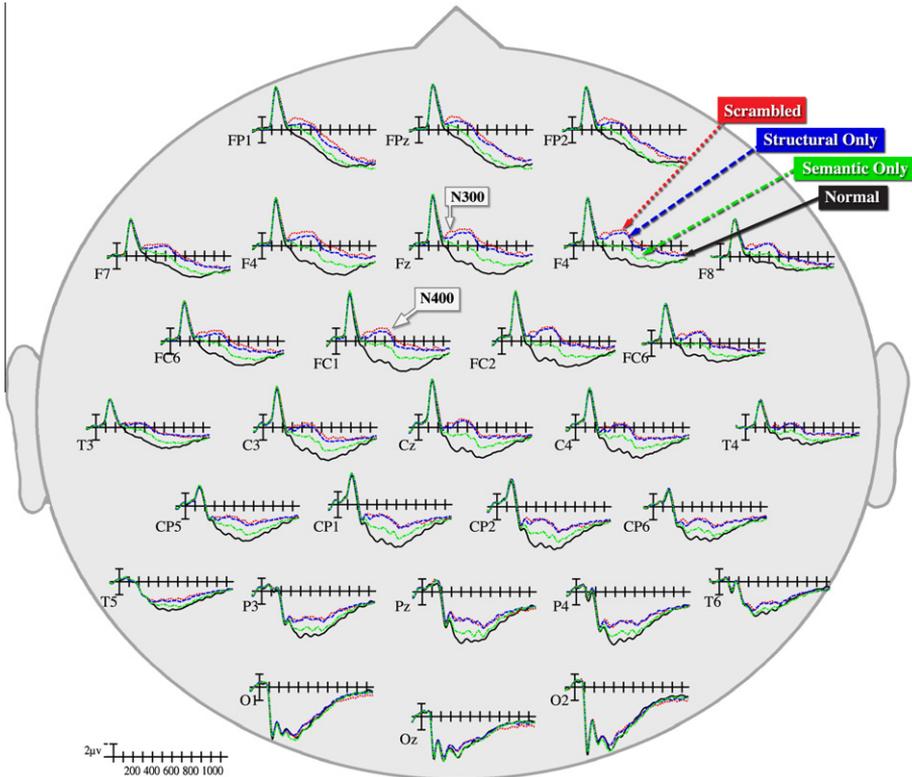
Fig. 6. ERPs to panels in all four sequence types.

appeared to be slightly somewhat left lateralized, with a four-way interaction between Sequence Type, Hemisphere, AP Distribution, and Electrode Site in the lateral regions analysis, see Table 3. Paired *t*-tests at individual sites showed significant differences only at midline frontal and left centro-frontal sites (FPz, Fz, F3, F7, FC1, FC5, C3). The scalp distribution of this effect is depicted in Fig. 8. *3.3.2.1.2. 400–600 ms.* Starting at around 400 ms and lasting until roughly 600 ms, a negative deflection peaked at 450 ms, consistent with an N400. Repeated measures ANOVAs comparing all four sequence types once again revealed significant main effects of Sequence Type at the midline column and lateral regions, interactions between AP Distribution and Sequence Type at the midline column, and an interaction between Sequence Type, AP Distribution, Electrode Site, and/or Hemisphere, in the lateral regions (see Table 3).

Just as for the N300, the amplitude of the N400 was smaller in the Semantic Only sequences than in the Structural Only sequences, and the N400 in the Normal sequences was smaller still. For both the Normal versus Semantic Only and the Semantic Only versus Structural Only contrasts, these effects were widely distributed, but maximal at anterior sites (significant main effects of Sequence Type and interactions between Sequence Type and AP Distribution in both the midline and lateral analyses), and with a slight rightward distribution (interactions between Sequence Type, Hemisphere, and Electrode Site in lateral analyses), see Figs. 6 and 7, Table 3. Follow-up *t*-tests showed effects at all sites (except for Oz in the Normal versus Semantic Only contrast).

Just as for the N300, however, the contrast between Structural Only and Scrambled sequences yielded a negativity effect which was smaller and less widespread, with a somewhat distinct scalp distribution. There were no main effects of Sequence Type, or interactions between Sequence Type and AP Distribution in either the midline or lateral analyses. Rather, there was only a four-way interaction between Sequence Type, Hemisphere, AP Distribution, and Electrode Site in the lateral analysis.
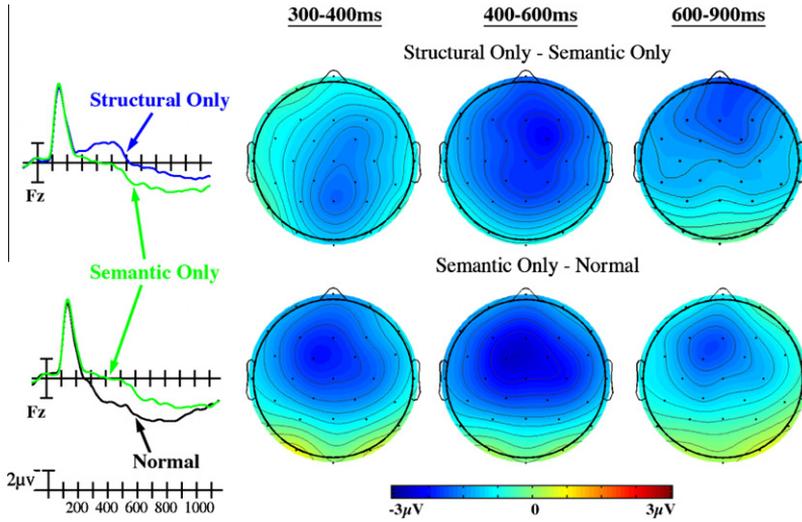
Fig. 7. Waveforms at Fz evoked by panels in Structural Only and Semantic Only sequences are shown on the left. Voltage maps illustrate the differences across the scalp surface of ERPs evoked by panels in Structural Only minus Semantic Only sequences (top row), and Semantic Only minus Structural Only sequences (bottom row) at the 300–400 ms (N300), 400–600 ms (N400) and 600–900 (late negativity) time windows.

Table 3
Results of ANOVAs comparing each sequence type.

	df	300–400 ms		400–600 ms		600–900 ms	
		Midline	Lateral	Midline	Lateral	Midline	Lateral
All Sequence Types							
ST	3,69	57.95***	73.28***	71.49***	96.76***	37.56***	53.55***
ST × AP	12,276	13.77***	17.58***	23.56***		19.80***	20.69***
ST × H	3,69						4.99**
ST × H × E	3,69		44.94***				19.85***
ST × AP × E	12,276				46.19***		
ST × AP × H × E	12,276		3.24***		3.92***		5.80***
Normal–Semantic							
ST	1,23	28.43***	39.58***	24.45***	42.93***	11.87**	17.03***
ST × AP	4,92	18.07***	23.70***	17.49***	18.69***	9.18***	15.30***
ST × H	1,23						4.90*
ST × H × E	1,23		30.37***		18.26***		12.47**
Semantic–Structural							
ST	1,23	36.58***	39.02***	45.78***	63.05***	32.19***	45.63***
ST × AP	4,92	3.42*		7.39***	6.36***	9.43***	6.50***
ST × H	1,23		5.98*				
ST × H × E	1,23				48.05***		9.66**
ST × AP × H × E	4,92		5.21***				
Structural–Scrambled							
ST	1,23		4.91*				
ST × AP	4,92	3.69**	4.77**			2.75*	2.64*
ST × H	1,23						3.01^
ST × AP × H × E	4,92		3.90**		4.75**		5.84***

Note. ST = Sequence Type, AP = AP Distribution, H = Hemisphere, E = Electrode Site. *F*-values are given.
  * $p < 0.05$.
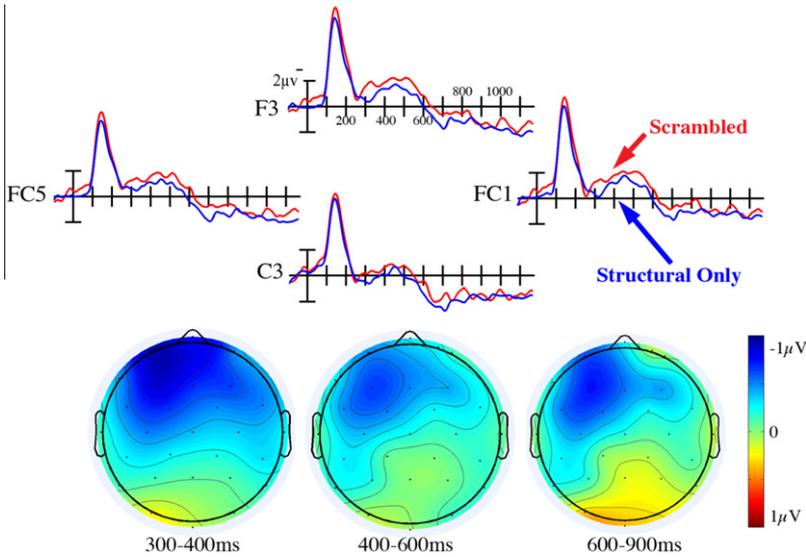 ** $p < 0.01$.
  *** $p < 0.001$.
  ^ $p < 0.1$.

Fig. 8. Waveforms at left anteriorly distributed electrode sites evoked by panels in Scrambled and Structural Only sequences (top). Voltage maps illustrate the scalp distribution of this effect at the 300–400 ms, 400–600 ms and 600–900 time windows. Note: this effect is smaller than the effects shown in Fig. 7. Therefore, a different voltage scale is used to illustrate its scalp distribution.

Follow-up of this four-way interaction using paired *t*-tests at individual electrode sites showed significant differences between Structural Only and Scrambled sequences only at left anterior sites (F3, FC5, FC1, C3), as depicted in Fig. 8.

3.3.2.1.3. 600–900 ms. The negativity effect continued throughout the 600–900 ms time window where again ANOVAs with all four levels of Sequence Type revealed significant main effects of Sequence Type and interactions between Sequence Type, AP Distribution, Hemisphere, and/or Electrode Site in both the midline column and lateral regions.

Again, panels in the Normal sequences evoked a smaller negativity than panels in the Semantic Only sequences, which, in turn, produced a smaller negativity than those in the Structural Only sequences. Both these effects were again widespread and maximal at anterior sites (main effects of Sequence Type and interactions between Sequence Type and AP Distribution in both midline and lateral analyses), and appeared to become slightly more left lateralized (interactions between Sequence Type, Hemisphere, and Electrode Site in the lateral analyses), see Table 3, Figs. 6 and 7. For the contrast between Normal and Semantic Only sequences, follow-up *t*-tests at individual sites showed effects at all sites except FPz, Oz, CP6, P4, T5 and T6. For the contrast between Semantic Only and Structural Only sequences, there were effects at frontal (FPz, F3, Fz, F4, F7, F8), centro-frontal (FC1 FC2, FC5, FC6, Cz) and two parietal sites (Pz and P3).

Finally, panels in the Structural Only sequences evoked a smaller negativity than those in the Scrambled sequences. This effect was once again smaller and less widespread than for the other contrasts, with no main effects but rather interactions between Sequence Type and AP Distribution in both midline and lateral analyses, and a four-way interaction between Sequence Type, Hemisphere, AP Distribution, and Electrode Site in the lateral analysis (Table 3). Follow-up of these interactions at individual sites showed that the effect localized to similar left anterior sites as in other time windows (Fz, F3, FC1, FC5, C3, CP5), as shown in Fig. 8.

3.3.2.2. Effects of sequence position. To simplify presentation and interpretation, the analyses described above did not include Sequence Position as a within-subject factor. To examine the effect of Sequence Position, a more focused analysis was carried out on ERPs evoked by panels for each sequence type, averaged across 14 electrode sites within frontal (F7, F8, Fz, F4, F8), centro-frontal (FC5, FC1, FC2,

FC6) and central-temporal (T3, C3, Cz, C4, T4) regions, see Fig. 5. This region of interest was selected because it covered the centro-anterior region where the effects described above were maximal. These values were used as the dependent measure in a 4 (Sequence Type) × 6 (Position) ANOVA, which, as in Experiment 1, was used to examine the effects of ordinal position across the different sequence types.

In all time windows, significant main effects were seen for Sequence Type (all $Fs > 46.1$, all $ps < .001$). Significant main effects of Position were found for the late negativity, $F(5, 115) = 4.36$, $p < .005$, but not for the N300 or N400 (all $Fs > 1.08$, all $ps > .266$). Of most interest, a significant interaction between Sequence Type and Position was found for all components (all $Fs > 2.6$, all $ps < .005$).

The Sequence Type × Position interaction in the overall ANOVA reflected differences in amplitudes across linear position between the four sequence types, as depicted in Fig. 9. This interaction showed linear trends in all time windows, as well as near-significant quadratic trends for the N300 (see Table 4). We followed up this interaction by analyzing the polynomial contrasts for linear trends across ordinal positions in each sequence type individually (statistics summarized in Table 4). For the Normal sequences, significant linear trends across ordinal Position for the N300/N400/late negativity components reflected a decrease in the amplitude across panel position. Additionally, Normal sequences showed quadratic trends on the N300 and late negativity (but not the N400). Semantic Only sequences showed no significant trends across positions in any time window. However, significant or near-significant linear trends on both the Structural Only and Scrambled sequences indicated an increase in amplitude across ordinal position on the N400 and late negativity, as well as on the N300 in the Scrambled sequences. There were no other significant effects, except for a quadratic trend in the Structural Only sequences on the N400, and a near-significant cubic trend in the Scrambled sequences on the N300.

3.3.2.3. Effects of comic reading fluency. Significant or trending correlations were observed between comic reading fluency and the magnitude of the N300 and N400 differences between Structural Only and Scrambled sequences averaged across frontal, centro-frontal, and central-temporal electrodes sites: N300 effect: $r(24) = -.429$, $p < .05$; N400 effect: $r(24) = -.381$, $p = .066$. These correlations indicated that greater fluency was associated with a larger negativity effect in these contrasts. No other contrasts yielded significant or near-significant correlations with fluency.

3.4. Discussion

Experiment 2 examined neural activity to all individual panels as participants viewed four different types of comic strips, one panel at a time. Collapsed across ordinal position, panels from Normal,
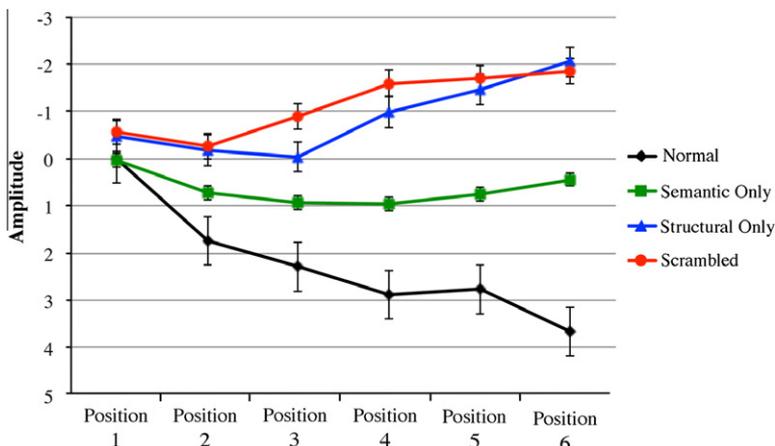

Fig. 9. Amplitudes of voltages evoked by panels in each sequence type at each sequence position are shown for the N400 (400–600 ms) collapsed across the frontal, centro-frontal, and central-temporal electrode regions (as with ERP data, positive is plotted down). Error bars represent standard error.

Table 4
Polynomial contrasts (*F*-values) for trends across ordinal sequence position in all four sequence types.

	300–400 ms			400–600 ms			600–900 ms		
	Linear	Quadratic	Cubic	Linear	Quadratic	Cubic	Linear	Quadratic	Cubic
All Sequence Types									
Position	0.26	1.46	4.96[*]	0	2.87	2.11	8.02[**]	3.51[^]	0.24
Pos × Type	27.86[***]	3.44[^]	0.143	57.33[***]	1.71	0.02	32.94[***]	0.73	0.06
Normal	9.97[**]	4.55[*]	0.77	17.36[***]	2.6	1.46	3.52[^]	7.43[*]	0.011
Semantic	0.26	0.18	1.29	0.18	2.63	0.08	2.11	2.13	0.15
Narrative	1.05	1.18	2.45	8.33[**]	5.60[*]	0.752	12.60[**]	0.133	0.023
Scrambled	3.51[^]	0.002	3.56[^]	11.37[**]	0.004	2.28	20.94[***]	2.38	0.48

df = 1,23. *F*-values are given.
 [*] *p* < 0.05.
 [**] *p* < 0.01.
 [***] *p* < 0.001.
 [^] *p* < 0.1.

Semantic Only, Structural Only and Scrambled sequences showed clear differences in a negative waveform. Consistent with previous ERP studies examining static images (see also Barrett & Rugg, 1990; Ganis et al., 1996; McPherson & Holcomb, 1999) and image and movie sequences (Sitnikova et al., 2008b, 2003; West & Holcomb, 2002), this negative deflection peaked at approximately 300 ms—the N300—followed by another negative deflection peaking at 500 ms—the N400. The negativity then continued past 900 ms, also consistent with previous research on image and movie sequences (Sitnikova et al., 2008b, 2003; West & Holcomb, 2002). Modulation of this negativity complex was widespread across the scalp surface but maximal at anterior electrode sites. The amplitudes of the N300/N400 complex and late negativity were largest to panels in sequences without any semantic relatedness (the Scrambled and Structural Only sequences), smaller to panels in the Semantic Only sequences, and smallest in the Normal sequences. This pattern is consistent with the modulation of the N400 across analogous conditions in Van Petten and Kutas's (1991) study of open-class words in sentences. In addition, we observed some divergence between the waveform evoked by panels in the Scrambled and Structural Only sequences in all epochs, but these effects were more localized than the N300/N400 effects described above, and, in the N400 and late negativity time windows, were significant only at left anterior sites.

We also observed effects of ordinal position on the modulation of the N300/N400 in the different sequence types. In the Normal strips, the amplitude of the N300/N400 showed a clear decrease along ordinal position of the panels within the sequence. In contrast, no decrease in N300/N400 amplitude appeared across ordinal position in either the Scrambled or Structural Only sequences. These findings also parallel those of Van Petten and Kutas (1991). Semantic Only sequences, which had no parallel in Van Petten and Kutas's experiment, also showed no decrease in the N300/N400 across sequence types. Below we discuss each finding in more detail.

The smaller N300/N400 to panels in strips without structure but with a common semantic theme (the Semantic Only strips), relative to those without a common semantic theme (the Structural Only and Scrambled strips) reflects, in part, the effects of semantic relatedness on processing of the pictures. This extends the findings of Barrett and Rugg (1990), McPherson and Holcomb (1999) and Federmeier and Kutas (2001) who observed similar effects of semantic relatedness in priming paradigms where a picture target was preceded by a single related prime picture. Of most interest, however, was the even smaller amplitude of the N300/N400 to panels in the Normal sequences, which contained both a common semantic theme and a narrative structure, than to panels in the Semantic Only strips. This suggests that, at least in the presence of semantic relatedness, a narrative structure conferred an advantage to the semantic processing of upcoming panels. This is analogous to the conclusions of Van Petten and Kutas who argued that the semantic processing of upcoming words in sentences was facilitated by context, which was built up through a combination of grammatical structure and meaning.

Importantly, there was much less separation between the waveforms evoked by panels in Structural Only sequences and Scrambled sequences. Indeed, at most electrode sites, there were no

differences between these two types of sequences. These results are once again analogous to those of Van Petten and Kutas (1991), who reported that open-class words within sentences with only syntax but no semantics evoked an N400 of the same amplitude as that evoked to words within random word strings. They took this to imply that syntax, in the absence of semantics, had no impact in reducing the amplitude of the semantically-sensitive N400 to such open-class words. Similarly, the present findings suggest that, in comprehending sequential images, the N300/N400 complex is relatively insensitive to narrative structure alone, in the absence of a common semantic theme linking panels.

Nonetheless, at some more localized left lateralized anterior electrode sites there was a small attenuation of the waveform evoked by panels in the Structural Only sequences relative to the Scrambled sequences. The left-lateralized scalp distribution of this effect is somewhat distinct from the N300/N400 effects described above and in previous studies of image processing. One possibility therefore is that it reflects a distinct ERP component, such as a LAN, which, as discussed, appears in the same time window as the N400 but has a left anterior distribution and has been associated with structural violations in language (Friederici, 2002; Neville et al., 1991), even in the absence of a semantically constraining context (Münte et al., 1997). Analogously, in the present study, the contrast between Structural Only and Scrambled sequences was designed to probe the build-up of a narrative structure across panels that is independent of semantic relatedness. Interestingly, Van Petten and Kutas (1991) did find a difference in this time window between scrambled sentences and structural only sentences to closed-class function words (which, in language, carry structural information), although it is unclear whether this effect had a left anterior distribution because activity was measured mostly at central and posterior electrode sites. If the ERP effect observed in the present study is indeed distinct from the N300/N400 effects seen in association with the other contrasts, this would suggest that a build-up of narrative structure across images is mediated by a neurocognitive process that is distinct from that which detects semantic relatedness. Such a process might have commonalities with that neurocognitive mechanisms that compute syntactic structure within sentences. We return to this idea in Section 4.

One observation that would be consistent with the idea of a neurocognitive mechanism for computing narrative structure is that the divergence between ERPs to panels in the Structural Only and Semantic Only sequences varied with the participants' level of fluency with comics. Fluent comprehenders of comics, who presumably had more exposure to this type of narrative structure, showed larger ERP differences between the Structural Only and Scrambled sequences than comprehenders with less fluency.

The second main finding of the current experiment was that ordinal position modulated the amplitudes of ERP effects in different ways, depending on sequence type. Analogous to the normal sentences in Van Petten and Kutas (1991), the amplitude of the negativity evoked by panels in the Normal sequences decreased across ordinal position. As for normal sentences, this finding suggests that full comprehension of sequential images relies on the build-up of *both* semantics and structure. We suggest that the combination of narrative structure and semantic theme built up a message-level context that facilitated the semantic processing of each successive panel. In contrast, also analogous to the findings of Van Petten and Kutas (1991), the amplitude of the N300/N400 did not decrease across ordinal position in the Scrambled or the Structural Only sequences. If anything, the amplitude of these components showed a slight increase across sequence position. This suggests that, in the absence of a semantic theme, the build-up of narrative structure alone across sequential position was not enough to facilitate the semantic analysis of upcoming images. Finally, the amplitude of the negativity evoked by panels in the Semantic Only sequences stayed relatively constant across sequence positions. This suggests that the presence of a semantic theme alone is insufficient to facilitate an incremental facilitation of semantic processing across sequence position.

4. General discussion

We carried out two experiments designed to test the hypothesis that sequential image comprehension involves an interaction between semantic relatedness across a common semantic theme, and a narrative structure. In Experiment 1, we measured reaction times while participants monitored for

target panels in sequences which featured coherent meaningful narratives (Normal), a semantic field without narrative structure (Semantic Only), a narrative structure without meaning (Structural Only), and totally random strings of images (Scrambled). In Experiment 2, we measured ERPs time-locked to all panels of these same sequence types in order to more directly examine neurocognitive processing without relying on behavioral performance.

In Experiment 1, reaction times were fastest to panels in the Normal sequences and slowest in the Scrambled ones, while in Experiment 2 an N300/N400 complex (with a widespread distribution but an anterior-central focus) was smallest to panels in the Normal sequences and largest to panels in the Scrambled ones. In both experiments, reaction times/N300/N400 amplitude to panels in the Semantic Only sequences fell in between—they were smaller than to panels in the Scrambled sequences, but larger than to panels in the Normal sequences. However, there were differences between Experiments 1 and 2 in the modulation of reaction times/ERPs to panels in the Structural Only sequences. In Experiment 1, reaction times to panels in these sequences were the same as in the Semantic Only sequences and faster than in the Scrambled sequences. In Experiment 2, however, the N300/N400 amplitude to panels in the Structural Only sequences was larger than to panels in the Semantic Only sequences, and, at most electrode sites, the same as in the Scrambled sequences (although, as discussed below, at more localized left anterior sites, there was a larger negativity in the Scrambled than the Structural Only sequences). Finally, in both Experiments 1 and 2, there was an effect of ordinal position on the behavioral/ERP response, but the pattern of this effect in each of the four Sequence types differed between the two experiments. In Experiment 1, reaction times decreased across ordinal position in all sequence types, converging to the same fast reaction times at the final position, possibly because participants abandoned comprehension in favor of carrying out the monitoring task, which became easier as sequences progressed. In Experiment 2, however, the N300/N400 amplitude decreased more through Normal sequences than in the other sequence types.

Below we consider these findings in more detail. We will suggest that sequential images use a narrative structure that combines with semantic theme to facilitate comprehension. We will then discuss the nature of such a narrative structure, considering open questions, before offering some final conclusions.

4.1. Effects of narrative structure in combination with semantic relatedness

The processing advantage of panels in Normal sequences over Structural Only sequences seen in both Experiments 1 and 2 aligns with previous studies showing that semantic relatedness facilitates the semantic processing of visual images (Ganis & Kutas, 2003; Ganis et al., 1996; McPherson & Holcomb, 1999), possibly through semantic priming mechanisms. The processing advantage of panels in Normal sequences over Semantic Only sequences shows that narrative structure adds more than simple semantic priming between individual panels: the structure works in combination with semantic relatedness to build up a situation model in the context which further facilitates semantic processing of each incoming panel. This is analogous to how structure and meaning are thought to combine during language processing to facilitate the semantic processing of incoming words. In language, there is mounting evidence that the amplitude of the N400 is not simply reducible to lexico-semantic associations or co-occurrence. Rather, it is influenced by the full meaning (propositional meaning) of the context—the meaning that results from the combination of semantics and structure (Kuperberg et al., 2011; Nieuwland & Kuperberg, 2008; Otten & Van Berkum, 2007).[2] Here we show that this is also true of the processing of sequential images.

Our finding in Experiment 2 that the amplitude of the N300/N400 decreased across ordinal position in the Normal sequences is consistent with this interpretation. We suggest that this decrease reflected

---

[2] While the N400 appears to be sensitive to the combination of structure and meaning during language processing, it does not in itself reflect the process of combining the structure and meaning. Rather it is thought to reflect the degree to which the semantic features of an incoming word match what is predicted by the preceding context (Kutas & Federmeier, 2011; Lau et al., 2008; van Berkum, 2009). Predictions may be generated before the incoming word is even encountered (for evidence for such truly predictive processing at the sentence level, see Delong, Urbach, & Kutas, 2005) or semantic matching processes may be initiated only after bottom-up features of the incoming word is encountered (see Marslen-Wilson, 1987).

the results of an incremental combination of semantic relatedness and narrative structure across the sequence: the more context built up, the greater the facilitation in semantic processing of each upcoming panel. This is analogous to the well-established decrement in the N400 amplitude across words within normal sentences. Importantly, there was no decrease in N300/N400 across ordinal position in the Semantic Only sequences. This suggests that it was the combination of narrative structure and semantic relatedness that drove semantic facilitation in the Normal sequences, rather than the build-up of semantic relatedness alone. Taken together, these findings offer converging evidence that sequential image comprehension involves the union of semantic relatedness and narrative structure.

4.2. Effects of narrative structure without semantic relatedness

The contrast between the Scrambled sequences and the Structural Only sequences directly probed comprehenders' use of narrative structure in the absence of semantic relatedness. The two experiments' dissociation in modulation of RTs/ERPs across these two sequences sheds light on the nature of this narrative structure. As explained above, in Experiment 1, RTs were significantly faster to target panels in the Structural Only than the Scrambled sequences. In Experiment 2, however, at most electrode sites, the N300/N400 amplitude was same to panels in the Structural Only and the Scrambled sequences. This mirrors a similar dissociation between behavioral and ERP findings seen in the analogous studies of sentence processing. Marslen-Wilson and Tyler (1975, 1980) reported reaction times which decreased from scrambled sentences, to syntactic only sentences. Van Petten and Kutas (1991), however, showed no difference in the amplitude of the N400 to open-class words in scrambled and syntactic only sentences. They suggested that while structure can facilitate processing of upcoming words, the N400 itself does not directly reflect structural integration costs. We offer an analogous interpretation for the current findings. We suggest that the N300/N400 did not distinguish between Structural Only and Scrambled sequences because the structure participants were using was relatively devoid of semantically related content.

Again, our finding that the N300/N400 did not show any incremental decrease across ordinal position in the scrambled sequences, which lacked both semantics and narrative structure, is consistent with this interpretation. Once again, it suggests that any structural system used to facilitate processing was not drawing on semantic relationships.

4.3. Neurocognitive processing of narrative structure in sequential images

We suggest that the Structural Only and Scrambled sequences were distinguished by a neurocognitive system that is sensitive to the build-up of narrative structure and is distinct from the system that computes semantic relatedness across panels. As discussed in Section 2.4, the shorter monitoring times to target panels in the Structural Only sequences relative to the Scrambled sequences cannot be attributed to local semantic relatedness or referential coherence between pairs of panels, and, as discussed above, the N300/N400 was relatively insensitive to the contrast between Structural Only and Scrambled sequences. However, in Experiment 2, we did see a small difference in the ERP between the Scrambled and Structural Only sequences. In contrast to the widespread N300/N400 effects seen in association with the other contrasts, this effect had a localized left anterior scalp distribution and we suggest that it may have reflected a distinct localized component—the LAN. In studies of language, the LAN has been associated with structural violations (Friederici, 2002; Neville et al., 1991), even in the absence of a semantic context (Münte et al., 1997). We therefore suggest that the LAN observed here indexed an analogous neurocognitive process that is sensitive to the narrative structure of sequential images.

We emphasize again that the narrative structure used to comprehend sequential images is quite different from syntactic structure used to comprehend single sentences. As the saying goes, "a picture is worth a thousand words": by and large, images can convey at least as much meaning as entire sentences, and we therefore conceive of the narrative structure of sequential images as more analogous to narrative structure at the discourse level of language. Nonetheless, there are several structural commonalities shared by narrative structure and syntax (see Appendix A). For example, both involve sequencing a semantic representation into a linear representation, which can account for phenomena

like distance dependencies and structural ambiguities, most likely through a hierarchic system of constituencies (Cohn, 2003, 2010). It is therefore possible that computing structure across sequential images recruits processing mechanisms that are shared with language. An analogous argument has been made for the processing of structure in music. A comparable effect to the LAN has been found over the right hemisphere to violations in musical structure (Koelsch et al., 2000; Patel et al., 1998). In discussing these findings, Patel (Patel, 2003; Patel et al., 1998) has proposed the "shared integration resource hypothesis", which posits the engagement of analogous neurocognitive mechanisms for syntax in language and structure in music. We argue that this hypothesis may be extended further from music to the processing of narrative structure in sequential images.

4.4. General implications and open questions

In sum, we have argued that these results support the idea that the comprehension of sequential images is guided by an interaction between narrative structure and semantic relatedness. A separation between narrative structure and semantic theme is predicted by Cohn's (in preparation) model of narrative structure for sequential images. While the narrative structure itself interfaces with aspects of semantics in predictable ways, it is conceived of as a system that is independent from semantic content. This is analogous to grammatical structure at a sentence level, which formed the inspiration for the design of these experiments: syntactic categories like nouns and verbs are different than semantic categories like objects and events, though they have prototypical correspondences with each other (Jackendoff, 1990, 2002). These experiments support the existence of such a structure and suggest that it can play a role in sequential image comprehension. Many questions, however, remain.

A first set of questions asks how the narrative categories discussed here relate to other constructs used to describe sequences of images. As discussed in Section 1, the most prominent approach to studying sequential image processing has taken its cue from the event-indexing model of discourse comprehension, which focuses on various types of local coherence relationships. For example, theories of discourse applied to film (Zacks & Magliano, 2011; Zacks et al., 2009, 2010) acknowledge that events are structured globally, but focus on the linear changes between characters, times, and locations (Zwaan & Radvansky, 1998). It is important to emphasize that the theory of visual narrative structure on which these experiments were based should not be viewed as an alternative to local coherence models. Rather, we conceptualize global narrative structure as a distinct system that sometimes aligns with local coherence relationships, but which can be dissociated from such relationships. This is once again analogous to models of language (particularly Jackendoff, 2002) in which syntactic and semantic structure often map homomorphically onto one another, but sometimes diverge.

In the present study, an example of where the global narrative structure aligns with local coherence relationships is in the comparison between the Normal and the Semantic Only sequences. We conceptualized these as differing in terms of presence or absence of a global narrative structure. However, they also clearly differ in their local coherence relationships: despite both having a common semantic theme, the Normal sequences have temporal (and sometimes causal) and referential coherence between sequential panels, whereas the Semantic Only sequences do not have any such local coherence relationships and have more referential shifts between panels than the Normal sequences. An example of where global narrative structure and local coherence relationships can be dissociated is in the contrast between the Structural Only and Scrambled sequences. As discussed above, these two types of sequences are very similar in terms of local relationships across their individual panels: in both types, there are very few links across panel pairs in characters, time, and location. Yet, they are distinguished in terms of the presence or absence of a global narrative structure. A challenge for future theoretical work will be to describe these associations and dissociations between local coherence relationships and global narrative structure in more detail.

Another important set of questions relates to the character of the narrative structure itself. Cohn has theorized that narrative structure is inherently hierarchical in nature. The present study did not address this directly. Previous work on story grammars in language sought evidence for hierarchy using statistical techniques of hierarchic clustering models (Gee & Grosjean, 1984; Mandler, 1987) or memory tasks (Mandler & Goodman, 1982). However, few approaches have examined how narrative constituents influence the processing of visual sequences. The study of hierarchy in sequential

images can again be inspired by methods of studying hierarchy in sentences. For example, will a disruption within versus between narrative constituencies facilitate or inhibit processing of a visual narrative? Will the processing of long distance dependencies in sequential images lead to similar neurocognitive effects as long distance dependencies in sentences? Examining these questions using contemporary psycholinguistic methods will be the subject of future experiments.

4.5. Conclusion

To conclude, these two experiments support the idea that a narrative "grammar" is indeed used in the visual-graphic modality, and offer evidence for the psychological validity of Cohn's (in preparation) theory of narrative structure across sequential images. Collectively, they point to a system of comprehension guiding sequential images that is broadly analogous to what is involved in processing verbal language: both sentences and sequential images require the combination of meaning (semantic relatedness) and structure (narrative structure/syntax) to build context across a sequence, and such context can be used incrementally during online comprehension. This opens the door to further research studying other aspects of structure and meaning, and their interface, within visual narratives.

Acknowledgments

Appendix A. Supplementary material

Supplementary data associated with this article can be found, in the online version, at doi:10.1016/j.cogpsych.2012.01.003.

References

Barrett, S. E., & Rugg, M. D. (1990). Event-related potentials and the semantic matching of pictures. *Brain and Cognition, 14*, 201–212.
Bentin, S., McCarthy, G., & Wood, C. C. (1985). Event-related potentials, lexical decision and semantic priming. *Electroencephalography and Clinical Neurophysiology, 60*, 343–355.
Black, J. B., & Bower, G. (1980). Story understanding as problem-solving. *Poetics, 9*, 223–250.
Black, J. B., & Wilensky, R. (1979). An evaluation of story grammars. *Cognitive Science, 3*, 213–230.
Bower, G. H., Black, J. B., & Turner, T. J. (1979). Scripts in memory for text. *Cognitive Psychology, 11*(2), 120–177.
Bransford, J. D., & Johnson, M. K. (1972). Contextual prerequisites for understanding: Some investigations of comprehension and recall. *Journal of Verbal Learning and Verbal Behavior, 11*(6), 717–726.
Brewer, W. F., & Lichtenstein, E. H. (1981). Event schemas, story schemas, and story grammars. In J. Long & A. D. Baddeley (Eds.), *Attention and performance IX* (pp. 363–379). Hillsdale: Erlbaum.
Brown, G., & Yule, G. (1983). *Discourse analysis*. Cambridge: Cambridge University Press.
Butcher, S. H. (1902). *The poetics of Aristotle*. London: Macmillian and Co. Ltd.
Carroll, J. M., & Bever, T. G. (1976). Segmentation in cinema perception. *Science, 191*(4231), 1053–1055.
Chatman, S. (1978). *Story and discourse*. Ithaca: Cornell University Press.
Chomsky, N. (1965). *Aspects of the theory of syntax*. Cambridge, MA: MIT Press.
Clark, H. H. (1972). Difficulties people have in answering the question "Where is it?". *Journal of Verbal Learning and Verbal Behavior, 11*, 265–277.
Cohn, N. (2003). *Early writings on visual language*. Carlsbad, CA: Emaki Productions.
Cohn, N. (2007). A visual lexicon. *Public Journal of Semiotics, 1*(1), 53–84.
Cohn, N. (2010). The limits of time and transitions: Challenges to theories of sequential image comprehension. *Studies in Comics, 1*(1), 127–147.
Cohn, N. (in preparation). Visual narrative structure. *Cognitive Science*.
de Beaugrande, R. (1982). The story of grammars and the grammar of stories. *Journal of Pragmatics, 6*, 383–422.
Delong, K. A., Urbach, T. P., & Kutas, M. (2005). Probabilistic word pre-activation during language comprehension inferred from electrical brain activity. *Nature Neuroscience, 8*(8), 1117–1121.
Federmeier, K. D., & Kutas, M. (1999). Right words and left words: Electrophysiological evidence for hemispheric differences in meaning processing. *Brain Research: Cognitive Brain Research, 8*, 373–392.

Federmeier, K. D., & Kutas, M. (2001). Meaning and modality: Influences of context, semantic memory organization, and perceptual predictability on picture processing. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 27*(1), 202–224.

Friederici, A. D. (2002). Towards a neural basis of auditory sentence processing. *Trends in Cognitive Sciences, 6*(2), 78–84.

Ganis, G., & Kutas, M. (2003). An electrophysiological study of scene effects on object identification. *Cognitive Brain Research, 16*(2), 123–144.

Ganis, G., Kutas, M., & Sereno, M. I. (1996). The search for "common sense": An electrophysiological study of the comprehension of words and pictures in reading. *Journal of Cognitive Neuroscience, 8*, 89–106.

Gee, J. P., & Grosjean, F. (1984). Empirical evidence for narrative structure. *Cognitive Science, 8*, 59–85.

Genette, G. (1980). In J. E. Lewin (Trans.) *Narrative discourse.* Ithaca: Cornell University Press.

Gernsbacher, M. A. (1983). *Memory for surface information in non-verbal stories: Parallels and insights to language processes.* Unpublished dissertation, University of Texas at Austin, Austin, Texas.

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

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

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

Glanzer, M., Fischer, B., & Dorfman, D. (1984). Short-term storage in reading. *Journal of Verbal Learning and Verbal Behavior, 17*, 467–486.

Greeno, J. G., & Noreen, D. L. (1974). Time to read semantically related sentences. *Memory and Cognition, 2*, 117–120.

Gunter, T. C., Stowe, L. A., & Mulder, G. (1997). When syntax meets semantics. *Psychophysiology, 34*(6), 660–676.

Haberlandt, K. (1984). Components of sentence and word reading times. In D. E. Kieras & M. A. Just (Eds.), *New methods in reading comprehension research.* Hillsdale, NJ: Erlbaum.

Hagoort, P., Brown, C. M., & Groothusen, J. (1993). The syntactic positive shift (SPS) as an ERP measure of syntactic processing. In S. M. Garnsey (Ed.), *Language and cognitive processes: Event-related brain potentials in the study of language* (pp. 439–483). Hove: Lawrence Erlbaum Associates (special issue).

Hahne, A., & Jescheniak, J. D. (2001). What's left if the jabberwocky gets the semantics? An ERP investigation into semantic and syntactic processing during auditory sentence comprehension. *Cognitive Brain Research, 11*, 199–212.

Halliday, M. A. K., & Hasan, R. (1976). *Cohesion in English.* London: Longman.

Halliday, M. A. K., & Hasan, R. (1985). *Language, context, and text: Aspects of language in a social-semiotic perspective.* Victoria: Deakin University Press.

Haviland, S. E., & Clark, H. H. (1974). What's new? Acquiring new information as a process in comprehension. *Journal of Verbal Learning and Verbal Behavior, 13*, 512–521.

Hobbs, J. R. (1985). *On the coherence and structure of discourse.* CSLI Technical report 85-37, Stanford, CA.

Jackendoff, R. (1990). *Semantic structures.* Cambridge, MA: MIT Press.

Jackendoff, R. (2002). *Foundations of language: Brain, meaning, grammar, evolution.* Oxford: Oxford University Press.

Kaufman, A. S., & Lichtenberger, E. O. (2006). *Assessing adolescent and adult intelligence* (3rd ed.). Hoboken: Wiley.

Keenan, J. M., Baillet, S. D., & Brown, P. (1984). The effect of causal cohesion on comprehension and memory. *Journal of Verbal Learning and Verbal Behavior, 23*, 115–126.

Kehler, A. (2002). *Coherence, reference, and the theory of grammar.* Stanford: CSLI Publications.

Kieras, D. E. (1978). Good and bad structure in simple paragraphs: Effects on apparent theme, reading time, and recall. *Journal of Verbal Learning and Verbal Behavior, 17*, 13–28.

Kintsch, W., & van Dijk, T. (1978). Toward a model of text comprehension and production. *Psychological Review, 85*, 363–394.

Koelsch, S., Gunter, T., Friederici, A. D., & Schröger, E. (2000). Brain indices of music processing: "Nonmusicians" are musical. *Journal of Cognitive Neuroscience, 12*(3), 520–541.

Kunzle, D. (1973). *The history of the comic strip* (Vol. 1). Berkeley: University of California Press.

Kuperberg, G. (2007). Neural mechanisms of language comprehension: Challenges to syntax. *Brain Research, 1146*, 23–49.

Kuperberg, G., Paczynski, M., & Ditman, T. (2011). Establishing causal coherence across sentences: An ERP study. *Journal of Cognitive Neuroscience, 23*(5), 1230–1246.

Kutas, M., & Federmeier, K. D. (2011). Thirty years and counting: Finding meaning in the N400 component of the event-related brain potential (ERP). *Annual Review of Psychology, 62*(1), 621–647.

Kutas, M., & Hillyard, S. A. (1980). Reading senseless sentences: Brain potential reflect semantic incongruity. *Science, 207*, 203–205.

Kutas, M., & Hillyard, S. A. (1984). Brain potentials during reading reflect word expectancy and semantic association. *Nature, 307*, 161–163.

Kutas, M., & Van Petten, C. (1994). Psycholinguistics electrified: Event-related brain potentials investigations. In M. A. Gernsbacher (Ed.), *Handbook of psycholinguistics* (pp. 83–143). San Diego, CA: Academic Press.

Lau, E. F., Phillips, C., & Poeppel, D. (2008). A cortical network for semantics: (De)constructing the N400. *Nature Reviews Neuroscience, 9*, 920–933.

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

Linde, C., & Labov, W. (1975). Spatial networks as a site for the study of language and thought. *Language, 51*, 924–939.

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

Mandler, J. M. (1978). A code in the node: The use of story schema in retrieval. *Discourse Processes, 1*, 14–35.

Mandler, J. M. (1984). *Stories, scripts, and scenes: Aspects of schema theory.* Hillsdale, NJ: Lawrence Earlbaum Associates.

Mandler, J. M. (1986). On the comprehension of temporal order. *Language and Cognitive Processes, 1*, 309–320.

Mandler, J. M. (1987). On the psychological reality of story structure. *Discourse Processes, 10*, 1–29.

Mandler, J. M., & DeForest, M. (1979). Is there more than one way to recall a story? *Child Development, 50*, 886–889.

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

Mandler, J. M., & Johnson, N. S. (1977). Remembrance of things parsed: Story structure and recall. *Cognitive Psychology, 9,* 111–151.

Mann, W. C., & Thompson, S. A. (1987). *Rhetorical structure theory: A theory of text organization.* Marina del Rey, CA: Information Sciences Institute.

Marslen-Wilson, W. D. (1987). Functional parallelism in spoken word-recognition. *Cognition, 25*(1–2), 71–102.

Marslen-Wilson, W. D., & Tyler, L. K. (1975). Processing structure of sentence perception. *Nature, 257,* 784–785.

Marslen-Wilson, W. D., & Tyler, L. K. (1980). The temporal structure of spoken language understanding. *Cognition, 8,* 1–71.

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

McPherson, W. B., & Holcomb, P. J. (1999). An electrophysiological investigation of semantic priming with pictures of real objects. *Psychophysiology, 36*(1), 53–65.

Morrow, D. G., Greenspan, S. L., & Bower, G. (1987). Accessibility and situation models in narrative comprehension. *Journal of Memory and Language, 26,* 165–187.

Münte, T. F., Matzke, M., & Johannes, S. (1997). Brain activity associated with syntactic incongruencies in words and psuedo-words. *Journal of Cognitive Neuroscience, 9,* 318–329.

Nakazawa, J. (2004). Manga (comic) literacy skills as determinant factors of manga story comprehension. *Manga Studies, 5,* 7–25.

Nakazawa, J., & Nakazawa, S. (1993). How do children understand comics? Analysis of comic reading comprehension. *Annual of Research in Early Childhood, 15,* 35–39.

Neville, H. J., Nicol, J. L., Barss, A., Forster, K. I., & Garrett, M. F. (1991). Syntactically based sentence processing classes: Evidence from event-related brain potentials. *Journal of Cognitive Neuroscience, 3*(2), 151–165.

Newtson, D., & Engquist, G. (1976). The perceptual organization of ongoing behavior. *Journal of Experimental Social Psychology, 12,* 436–450.

Nieuwland, M. S., & Kuperberg, G. R. (2008). When the truth is not too hard to handle. *Psychological Science, 19*(12), 1213–1218.

Osterhout, L., & Holcomb, P. (1992). Event-related potentials elicited by syntactic anomaly. *Journal of Memory and Language, 31,* 758–806.

Otten, M., & Van Berkum, J. J. A. (2007). What makes a discourse constraining? Comparing the effects of discourse message and scenario fit on the discourse-dependent N400 effect. *Brain Research, 1153,* 166–177.

Patel, A. D. (2003). Language, music, syntax and the brain. *Nature Neuroscience, 6*(7), 674–681.

Patel, A. D., Gibson, E., Ratner, J., Besson, M., & Holcomb, P. J. (1998). Processing syntactic relations in language and music: An event-related potential study. *Journal of Cognitive Neuroscience, 10*(6), 717–733.

Rugg, M. D. (1984). Event-related potentials in phonological matching tasks. *Brain and Language, 23,* 225–240.

Rumelhart, D. E. (1975). Notes on a schema for stories. In D. Bobrow & A. Collins (Eds.), *Representation and understanding.* New York, NY: Academic Press.

Saraceni, M. (2000). *Language beyond language: Comics as verbo-visual texts.* Unpublished dissertation, University of Nottingham.

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

Saraceni, M. (2003). *The language of comics.* New York, NY: Routeledge.

Schank, R. C. (1975). The structure of episodes in memory. In D. Bobrow & A. Collins (Eds.), *Representation and understanding: Studies in cognitive science* (pp. 237–272). New York: Academic Press.

Schank, R. C., & Abelson, R. (1977). *Scripts, plans, goals and understanding.* Hillsdale, NJ: Lawrence Earlbaum Associates.

Sitnikova, T., Holcomb, P. J., & Kuperberg, G. (2008a). Neurocognitive mechanisms of human comprehension. In T. F. Shipley & J. M. Zacks (Eds.), *Understanding events: How humans see, represent, and act on events* (pp. 639–683). Oxford University Press.

Sitnikova, T., Holcomb, P. J., & Kuperberg, G. (2008b). Two neurocognitive mechanisms of semantic integration during the comprehension of visual real-world events. *Journal of Cognitive Neuroscience, 20*(11), 1–21.

Sitnikova, T., Kuperberg, G., & Holcomb, P. (2003). Semantic integration in videos of real-world events: An electrophysiological investigation. *Psychophysiology, 40*(1), 160–164.

Sitnikova, T., Paczynski, M., & Kuperberg, G. R. (2007). Context motivates comprehenders attempts to make sense of novel events: Evidence from event-related potentials. *Journal of Cognitive Neuroscience Supplement.*

Sitnikova, T., West, W. C., Kuperberg, G., & Holcomb, P. (2006). The neural organization of semantic memory: Electrophysiological activity suggests feature-based segregation. *Biological Psychology, 71*(3), 326–340.

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.

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

Stein, N. L., & Glenn, C. G. (1979). An analysis of story comprehension in elementary school children. In R. Freedle (Ed.), *New directions in discourse processing* (pp. 53–119). Norwood, NJ: Ablex.

Stein, N. L., & Nezworski, T. (1978). The effects of organization and instructional set on story memory. *Discourse Processes, 1*(2), 177–193.

Thorndyke, P. (1977). Cognitive structures in comprehension and memory of narrative discourse. *Cognitive Psychology, 9,* 77–110.

Tomashevsky, B. (1965). Thematics. In L. T. Lemon & M. I. Reis (Eds.), *Russian formalist criticism: Four essays* (pp. 61–95). Lincoln: University of Nebraska Press.

Trabasso, T., Secco, T., & van den Broek, P. (1984). Causal cohesion and story coherence. In H. Mandl, N. L. Stein, & T. Trabasso (Eds.), *Learning and comprehension of text* (pp. 83–111). Hillsdale, NJ: Erlbaum.

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

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

van Berkum, J. J. A. (2009). The neuropragmatics of "simple" utterance comprehension: An ERP review. In U. Sauerland & K. Yatsushiro (Eds.), *Semantics and pragmatics: From experiment to theory* (pp. 276–316). Basingstoke: Palgrave.

van Berkum, J. J. A., Hagoort, P., & Brown, C. M. (1999). Semantic integration in sentences and discourse from the N400. *Journal of Cognitive Neuroscience, 11,* 657–671.

van Dijk, T. (1977). *Text and context*. London: Longman.

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

Van Petten, C., & Kutas, M. (1991). Influences of semantic and syntactic context on open- and closed-class words. *Memory and Cognition, 19*, 95–112.

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

Yang, C. L., Perfetti, C. A., & Schmalhofer, F. (2007). Event-related potential indicators of text integration across sentence boundaries. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 33*(1), 55–89.

Zacks, J. M., & Magliano, J. P. (2011). Film, narrative, and cognitive neuroscience. In D. P. Melcher & F. Bacci (Eds.), *Art and the senses*. New York: Oxford University Press.

Zacks, J. M., Braver, T. S., Sheridan, M. A., Donaldson, D. I., Snyder, A. Z., Ollinger, J. M., et al (2001). Human brain activity time-locked to perceptual event boundaries. *Nature Neuroscience, 4*(6), 651–655.

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.

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*, 1–15.

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, 130*(1), 29–58.

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

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