Balancing Grammar and Semantics in "Comics":

Global structure in sequential image processing

A thesis

submitted by

Neil Thomas Cohn

In partial fulfillment of the requirements

for the degree of

Master of Science

in

Psychology

TUF TS UNIVERSITY

February

2011

ADVISERS: Ray Jackendoff, Gina Kuperberg, Phil Holcomb

Abstract

Image sequences

Grammar, Meaning — separate?

RTs, ERPs.

Acknowledgements

Thanks go to my professors, Ray Jackendoff, Gina Kuperberg, and Phil Holcomb for their help and support through all stages of this project. Martin Paczynski aided tremendously both in research design and in the gathering of data. Research assistants Martina Santarsieri, Lauren Rose, Adrienne Ing all helped greatly in the creation of stimuli and/or gathering of data. My assistant Natalie Buzzeo should also be recognized individually for investing an amazing amount of time and effort into the creation of stimuli for this project. Finally, Fantagraphics Books is thanked for their generous donation of the volumes of *The Complete Peanuts.* This work was supported by NIMH (R01 MH071635) and NARSAD (with the Sidney Baer Trust).

Table of Contents

Abstract ii
Acknowledgements iii
Table of Contents iv
Table of Figures v
Table of Tables v
Introduction 1
Background 2
EXPERIMENT 1: Target Monitoring 20
Experiment 1: Methods 24
Construction of stimuli 24
Rating Study 1: Global Ratings 29
Rating Study 2: Local Ratings 29
Participants in the Panel Monitoring experiment 32
Procedure 32
Experiment 1: Results 35
Accuracy 35
Reaction Times for Target Monitoring 37
Experiment 1: Discussion 42
EXPERIMENT 2: Event Related Potentials 48
Experiment 2: Methods 56
Participants 56
Stimuli 57
Procedure 57
ERP Recordings 58
Behavioral Data Analysis 59
ERP Data Analysis 59
Experiment 2: Results 60
Behavioral Data 60
ERP Data 61
300-400: N300 61
400-600: N400 64
600-900ms: Late Negativity 67
Effects of Sequence Position 69
Effects of Fluency 71
Experiment 2: Discussion 74
General Discussion 79
Appendix: Additional Examples of Stimuli 85
References 88

Table of Figures

Figure 1. Transitions in a narrative sequence **6**
Figure 2. Semantic field in an infelicitous sequence **9**
Figure 3. Rearrangement of panels in sequences **10**
a. Rearranged strip of a semantic field **10**
b. Rearranged strip from a narrative sequence **10**
Figure 4. Narrative grammar in sequential images **15**
Figure 5. Example Stimuli. Critical panels have thick borders **26**
Figure 6. Mean false presses for sequence types across ordinal positions **36**
Figure 7. Mean reaction times for sequence types across all positions **38**
Figure 8. Mean reaction times for panels in sequence types across ordinal positions **39**
Figure 9. Waveforms and voltage maps for event related potentials to all sequence types **63**
Figure 10. Contrasts and voltage maps for the Scrambled versus Structural Only sequence types **65**
Figure 11. Amplitudes of the N300 across sequence position **70**
Figure 12. Amplitudes of the N400 across sequence position **70**
Figure 13. Amplitudes of sustained negativity in the 600-900ms time window across sequence position **71**
Figure 14. Correlation between Structural Only and Scrambled sequence type amplitudes with Comic Reading Fluency for the 300 to 400 ms time window **72**
Figure 15. Correlation between Structural Only and Scrambled sequence type amplitudes with Comic Reading Fluency for the 400 to 600 ms time window **72**
Figure 16. Correlations (r values) between the difference between Scrambled versus Structural Only and Comic Reading Fluency plotted by electrode site **73**

Table of Tables

Table 1. Mean ratings for experimental stimuli (on a scale of 1-7) **31**
Table 2. Regressions across panel positions for each sequence type for a subjects analysis **40**
Table 3. Mean accuracy for plausibility judgments for differing sequence types (standard deviation shown in parentheses) **60**
Table 4. Statistical analysis of waveforms in the 300-400ms time window **62**
Table 5. Statistical analysis of waveforms in the 400-600ms time window **66**
Table 6. Statistical analysis of waveforms in the 600-900ms time window **68**
Table 7. Regressions for amplitudes across sequence position **69**

Introduction

Drawings have been conveying narratives through sequences of images for millennia, whether painted on cave walls, carved into reliefs, or hung on medieval tapestries (Kunzle, 1973; McCloud, 1993). In their modern context, this visual language of sequential images appears in modern day comic books, which have provided an excellent forum for the study of narrative structure outside the context of verbal discourse (e.g. Cohn, 2003, 2010; McCloud, 1993; Saraceni, 2000; Weber, 1989). However, while ample research has focused on the structure and comprehension of verbal and written narratives, scholarship has barely examined the driving forces behind the understanding of sequential images: What are the representations and mechanisms engaged in sequential image comprehension? How is structure processed across a sequence of images? This research addresses these questions by examining online processing of narrative structures in sequential images using both reaction time measures and Event-Related Potentials (ERPs). First, we will review the existing research and theories pertaining to the structure and processing of sequential image comprehension. This will set the stage for the experiments described in Experiments 1 and 2, which aim to determine whether and how these theoretical principles play out in influencing the neurocognitive processes engaged during online processing of sequential images.

Background

Comprehension of sequential images

Researchers on verbal discourse have assumed that visual narratives are processed in comparable ways (Gernsbacher, 1983, 1985; Gernsbacher, Varner, & Faust, 1990; Robertson, 2000; Trabasso & Nickels, 1992; West, 1998; West & Holcomb, 2002), and therefore it is useful to review the foundational ideas of discourse processing. Van Dijk and Kintsch (1983) have described discourse comprehension as passing through four distinct levels: 1) its "textbase" of the surface form of the words and grammar, 2) its "microstructure" of the textbase's propositional structure, 3) a "macrostructure" of the gist or theme of the meaning, and finally 4) the situation model that additively retains the meaning of the discourse in long-term memory after the reading of a text. In brief, a reader engages the words and grammar (or, by extension, images and sequence) as the textbase, from which they acquire the microstructure of its meaning, and then extract the overall macrostructure of its gist. This information is then stored in a situation model that additively constructs the broader meaning of the whole discourse. Importantly, this model distinguishes between the surface aspects of a discourse's *structure* (its textbase) and the aspects of its *meaning* — both the extraction of micro/macrostructure from the textbase or storing it in the situation model. Furthermore, while comprehension for information about the textbase is lost to recall, semantic information in the situation model is retained in long-term memory.

Given this overall orientation, Gernsbacher has examined visual narratives of "picture stories" in studying discourse (Gernsbacher, 1983, 1985; Gernsbacher, et al., 1990). Her *structure-building framework* (Gernsbacher, 1990) argues that the situation model is built up across the course of reading a discourse — be it written, verbal, or pictorial. According to this framework, the situation model is continually updated with new information as a discourse is read or heard. As new information is acquired, the previously acquired data is continually referenced until it reaches a boundary of the global structure. At this point, it resets and the process repeats for the next structure. Thus, along the progression of a discourse, a situation model continuously aggregates more information, building the active information in working memory until a boundary is crossed and a new aggregation begins.

While it has primarily been applied to written discourse, the initial evidence for this structure-building framework came directly from studies examining how meaning is built up across sequential images in "picture stories." Notably, in her dissertation, Gernsbacher (1983) showed that seemingly continuous graphic stories can be divided up into parts. A fully pictorial story from a children's book was presented to participants, who were asked to mark where they thought boundaries divided up the scenes. Participants showed high agreement on where to mark episode boundaries for segments of the overall story, also consistent with the experimenters' expected boundary locations. These boundaries divide the overall graphic discourse into sub-episodes to break up sequences of images within a story. Cohn (2003) has offered a theoretical notion of a "visual sentence" that appears similar to the constituents marked by these boundaries. Each "visual sentence" is viewed as a sequence of images conveying a specific interaction or interrelation of characters in a scene or narrative. Because the graphic form does not use any notation to mark these "sentence" boundaries, the sequence of a graphic narrative would appear to end only at the end of the story. However, Gernsbacher's findings support the hypothesis that such groupings divide up the overall graphic discourse into smaller parts.

Beyond just showing that people can locate boundaries for graphic constituencies, Gernsbacher (1983, 1985) has found effects on recall for the crossing of the boundaries of these sub-episodes. When asked to remember the visual composition of particular images (normal vs. flipped frames) within the sequence of images in a picture story, the accuracy of participants' recall decreased when the target image appeared after as opposed to before the "visual sentence" boundary, as defined by the segmentation process described above. Because recall for the visual composition was more difficult after a boundary, it shows that boundaries only interfere with comprehension when beginning a new structure. This implies that memory builds up until the ending of a structure (marked by a boundary) and resets with the beginning of the next. Indeed, in the original dissertation, Gernsbacher (1983) notes that serial position of images within the sequence showed a relationship to accuracy of recall. Memory of surface information is lost more for earlier versus later parts of the "visual sentence" as structure builds up sequentially. An additional task showed that memory for visual composition was generally better for images from normal sequences than those from a picture story where the sequence was scrambled (Gernsbacher, 1983). Similar results have been found for recall of changes in camera angle or shot type at the segment boundaries in filmic narratives (Carroll & Bever, 1976). As in the picture stories, these film results show a build-up of structure along the sequence until viewers reach a boundary, at which point the segmentation begins anew with the next structure.

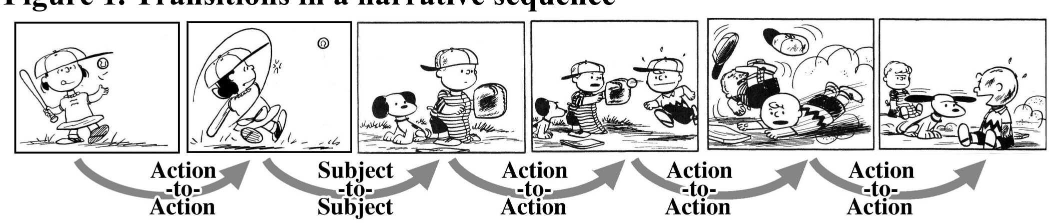
Further studies by Gernsbacher again compared the comprehension for written, auditory, and picture stories using recall tasks (Gernsbacher, et al., 1990). Based on assessments of reading comprehension, skilled readers' memory for scrambled picture stories was worse than for normal sequences in picture stories. Meanwhile, less skilled readers showed little difference in comprehension between scrambled and normal image sequences. Comparable results appeared in both the written and auditory stories as well. These findings were interpreted as indicating that less skilled readers lose access to information earlier in the sequence than skilled comprehenders, who can more easily build structures sequentially. As with other studies of situation models, these results also suggest that the same structure-building operations may be drawn upon for narrative comprehension in verbal and non-verbal domains.

All told, Gernsbacher's studies construct a solid case for the buildup of structure across sequential images. However, they leave open the question as to what the character of that structure might be. For this, we turn to theories of the structure of sequential images to describe just what might be building up across a sequence.

Theories of sequential image structure

Early theories of sequential image comprehension have focused on the relationship of one panel to another. The most prominent of these models emerged from comic author and theorist Scott McCloud (1993), who proposed six types of linear relationships between panels based on various dimensions of spatial and temporal changes. For example, Moment-to-Moment transitions feature a small scope of temporal change (two panels showing the blinking of an eye) while Action-to-Action transitions depict a wider scope (two panels showing the hitting of a baseball). Subject-to-Subject transitions shift between different characters, Scene-to-Scene transitions change between broader places and locations, while Aspect-to-Aspect panels show features of the background scene or environment. Finally, Non-Sequitur transitions have no relation between panels at all.

Figure 1. Transitions in a narrative sequence



A sample transitional analysis of a sequence is depicted in Figure 1. This sequence depicts the progression of a baseball game that primarily uses Action-to-Action transitions, save for the change in characters between panels two and three, which uses a Subject-to-Subject transition. Each of these "transitions" facilitates what McCloud called "closure" — a process where the mind "fills in the gap" of the unseen information that may connect each pairing of panels. Closure functions for McCloud similarly to the linguistic notion of inference — the mind providing unexpressed meaning — which here motivates *all* comprehension of image pairings. Panel relationships that are easily integrated in meaning require "less closure" (i.e. inference) than those requiring more difficult connections.

Theories of verbal discourse have also focused on the inference created by pairwise relationships between clauses or sentences (Halliday & Hasan, 1976; Hobbs, 1985; Mann & Thompson, 1987), so it is perhaps unsurprising that such approaches have been meshed with McCloud's transitions that connect panels (Saraceni, 2000, 2001; Stainbrook, 2003). McCloud did not describe the mechanisms driving closure or transitions, though fusing his work with discourse theories provided such elaborations. For example, Saraceni (2000, 2001, 2003) reinterprets McCloud's model by placing particular importance on the balance between repeated referential information (similar characters appearing across panels) and semantic associative fields (semantic themes repeated throughout panels). Referential information draws upon the tradeoff of new versus given information (Haviland & Clark, 1974) in a sequence, such as showing similar characters across panels or introducing new ones. Meanwhile, semantic associative fields bind individual elements to a common meaning, for example disparate panels of a horse's head and legs, a jockey, and spectators additively convey the concept of a "horse-racing track"(Saraceni, 2000, 2001, 2003).

Saraceni re-conceptualizes McCloud's transitions as balancing referential information and semantic fields (Saraceni, 2000, p. 131)[1]:

1. Moment-to-Moment – all referential elements of one panel are present in the next
2. Action-to-Action – most of the referential elements of one panel are present in the next
3. Aspect-to-Aspect – elements of the panels occur in the same semantic field
4. Subject-to-Subject – little referential repetition, elements share a common semantic field
5. Non-Sequitur – no markers of relatedness at all

Additionally, Saraceni's listed order for transitions ranks them based on how much inference they demand. Listed first, Moment-to-Moment transitions are assumed to require less inference than Non-Sequitur ones, listed last. However, he emphasizes that Non-Sequitur transitions are not necessarily fully disconnected panels, but that they simply require a covert connection with greater inferential demand. That is, Non-Sequitur relations are still "meaningful", though that meaning comes entirely from the reader's interpretive inference. Finally, like McCloud, Saraceni believes that the strongest inference is given between immediately juxtaposed panels, though he acknowledges that distant connections

[1] Note that Scene-to-Scene transitions are missing here. Saraceni incorporates Scene-to-Scene transitions as just a different type of Subject-to-Subject transitions only at a scope of locations instead of individuals.

between panels may involve weaker ties (Saraceni, 2001). Thus, Saraceni's application of discourse to sequential images balances the repetition of referential and thematic information, fused together through inference.

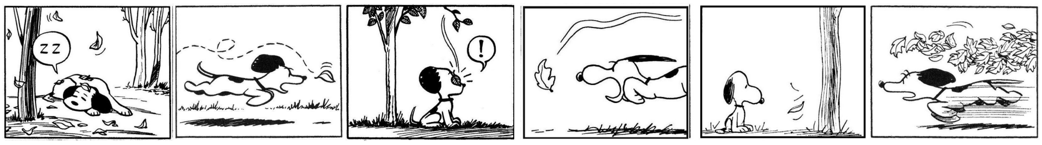
However, referential information and semantic fields alone cannot explain how a progression of images conveys a narrative. Take for instance the sequence in Figure 2.

Figure 2. Semantic field in an infelicitous sequence

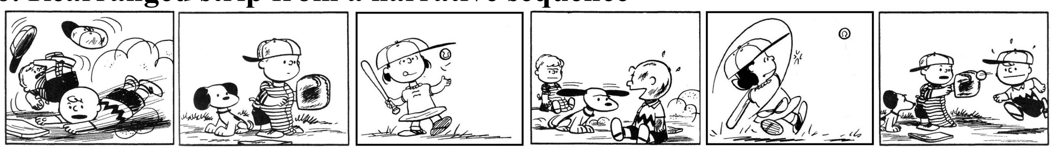
Here, all of the panels feature the same referential entities: Snoopy and leaves. By Saraceni's analysis, this would constitute a sequence of Aspect-to-Aspect transitions — bound solely by the semantic association pertaining to "Snoopy's interaction with leaves." However, despite this repetition creating a thematic semantic field, the sequence does *not* convey a felicitous sequence. As a narrative it makes no sense. Indeed, the types of thematic fields described by Saraceni could just be considered "graphic lists": "Things that you find in a racetrack" or here, "Glimpses of Snoopy interacting with leaves." Lists do not require a particular ordering of their units, and indeed sequences like Figure 2 that feature such a semantic field can be rearranged with no violation in how well-formed they feel, as depicted in the scrambled version of Figure 2 in Figure 3a. On the other hand, other sequential images quite clearly can be violated by scrambling the orders in their narrative sequence, as in the rearranged version of Figure 1 in Figure 3b.

Figure 3. Rearrangement of panels in sequences

a. Rearranged strip of a semantic field



b. Rearranged strip from a narrative sequence



Indeed, the rearrangement of panels in Figure 3b leaves no meaningful progression, but does yield a semantic field (here, about the category "baseball"). As Gernsbacher and colleagues (Gernsbacher, 1983; Gernsbacher, et al., 1990) showed, comprehension of scrambled narrative image sequences impairs comprehension, meaning that the mind is recognizing that scrambled narratives violate some form of cognitive system. However, such a scrambled sequence would be legal in Saraceni's (2000, 2001, 2003) model, since it would just be considered a series of ("meaningful") Non-Sequitur or Subject-to-Subject transitions. Indeed, since only local connections are accounted for in this approach, there is no consideration of a system guiding sequences to be a meaningful progression. This is precisely the limitation shown by Figure 2: there are meaningful associative relationships between panels (both locally and globally), but there is nothing guiding the reading to be a coherent narrative across the whole sequence.

Indeed, it has long been noted that narrative has a *global structure* that extends beyond linear, local relationships and broad scale semantic associations. Such observations go back at the level of plotlines to Aristotle's *Beginning-Middle-End* structure from *Poetics* (1902) and Freytag's (1894) progression of *Set up-Rising action-Climax-Falling action-Resolution,* a triangular story structure intended to describe 5-act plays. Just as syntax in language allows us to differentiate coherent sentences from scrambled strings of words, the comprehension of sequential images must also use a cognitive system to distinguish a coherent narrative sequence from a random string of images (or a string of random yet semantically connected images). Such a structure in the graphic domain would parallel grammars in the other conceptual expressive modalities of verbal and manual languages. However, because images contain more conceptual information than syntactic units like words, a grammar for the graphic form would operate at (more or less) a discourse level and thereby be appropriate for narratives in any domain.

Models of grammar for discourse have been attempted previously in theories of "story grammars," which abounded in psycholinguistics in the late 1970s and early 1980s (e.g. Mandler & Johnson, 1977; Rumelhart, 1975; Stein & Glenn, 1979; Thorndyke, 1977). Story grammars used traditional narrative categories based around characters' achievements of goals to characterize the parts and processing of written stories. This formalism provided an organization of story structure based around "problem solving schema" for characters' navigation through goal-directed events (Mandler, 1984; Rumelhart, 1980;

Thorndyke, 1977). For example, by and large the various models seem to agree that stories begin with the establishment of the Setting of characters and the environment, along with the backdrop of the narrative. An Initiating event sets the narrative in motion, which results in a protagonist to have 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 a story. This narrative progression was formally described in a canonical mental schema that allowed for embeddings of these segments into larger Episodes and a full Story.

Story grammar research largely used memory paradigms to compare individuals' recall of stories with the predictions made by the models. For example, this research has found that the canonical episode structure is recalled better than when alterations are made to it, such as stories with changes made to their temporal order (Mandler & Johnson, 1977) or when the order of sentences is inverted (Mandler, 1978, 1984; Mandler & DeForest, 1979). As in Gernsbacher et. al's (1990) studies using sequential images, scrambled orders of texts appear to be stored in memory worse than episodic structures (Mandler, 1984). However, the reliance on this schema to guide recall appears to change as people age. Adults remember the surface structure of altered stories better than do children, who are more likely to reorder stories into canonically ordered descriptions (Mandler, 1978; Mandler & DeForest, 1979). Similar studies have found that when the sequence of events in a story are scrambled, recall worsens along the degree to which the story was altered (Stein & Nezworski, 1978). However, other studies indicate that memory for broad scale aspects of story organization are recalled better than the lower-level surface details (Thorndyke, 1977). Altogether, these studies point towards an entrenched canonical schema playing a role in narrative understanding which is processed more easily than variations to it.

Story grammar categories have been extended to sequential images in one study looking at children's comprehension of sequential "picture stories." Trabasso and Nickels (1992) analyzed the narrative structure of an illustrated children's book, and compared it to those generated by children's descriptions of the story. They found that children pass through several distinct stages of awareness of goals and actions in sequential images that is similar to story understanding in the verbal domain (Stein, 1988). Children at three years old start only able to recognize referential information like objects and characters, but by four years old can recognize the relations between characters and can describe the actions represented directly in the image. However, it is not until five years old that children can identify the goals and intentions motivating those actions and tie the content of images together. These findings indicate that comprehension of sequential images progresses in stages of development that move from the establishment of referential information through the understanding of actions, events, and their motivating causes (see also Pallenik, 1986 in this regard).

Narrative structure in sequential image

While story grammars have been applied limitedly to sequential images, a grammatical theory of narrative structure designed particularly for sequential images has been proposed by Cohn (In prep-a). Like story grammars' treatment of sentences, this model describes panels as playing narrative roles in relation to a global sequence, and could potentially be extended to describe verbal discourse as well. This narrative grammar consists of five core "parts of speech" that play different functional roles, again somewhat similar to aspects of Freytag's (1894) narrative categories (Cohn, In prep-a). Notice that these are more like grammatical functions than like syntactic categories. The same image can function in different roles given the constraints of the sequence:

Establishers (E) – set up interrelations without acting on them

Initials (I) – initiate the interrelation or event

Prolongations (L) – act as a medial extender between categories

Peaks (P) – mark the highest point of tension of the interrelation or event

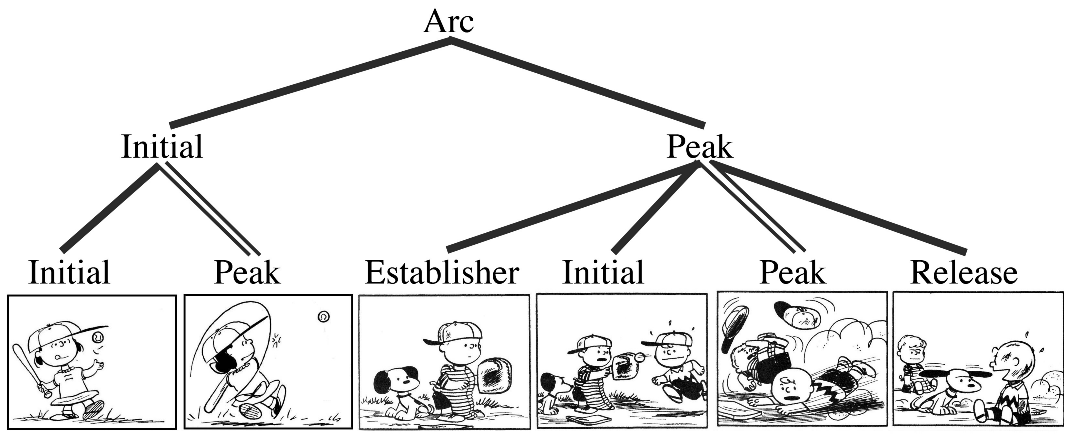
Releases (R) – release the tension of the interrelation

These categories comprise parts of an overarching narrative Arc, a maximal structure similar to the maximally high level of a "sentence" in the hierarchy of syntax. Like grammatical functions, these narrative roles can be used in various patterns. However, just as syntactic structure might use a canonical pattern of Subject-Object-Verb in English, a canonical narrative Arc progresses through the categories *Establisher-Initial-Peak-Release* in its own structure. However, importantly, each narrative category can be elaborated into its own sub-Arc. This makes the grammar inherently recursive, since, for example, a whole

“Initial-Peak-Release” phrase can act as the Initial for a higher Peak phrase. This can best be understood with an example.

Consider Figure 4, which illustrates the narrative structure in the same sequence as in Figure 1.

Figure 4. Narrative grammar in sequential images



This sequence shows a baseball game where Lucy hits the ball so that Charlie Brown can run home and score while escaping getting tagged out by Schroeder. It begins with a small node of an Initial and a Peak where Lucy hits the ball, which then sets up the remainder of the strip. The second clause follows a canonical pattern, as a subordinate clause 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 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. This first clause (Lucy hitting the ball) facilitates the second (Charlie scoring) and thus becomes an Initial at a higher level of processing, motivated by it's "head": the Peak inside it. Set up by the Initial of the first clause, the second clause itself is a Peak, again motivated by the Peak inside it.

Notice also that the structure in example in Figure 2 cannot be accounted for with this narrative grammar. Since they feature just semantic associations, the panels in that sequence do not play any narrative roles in relation to the whole. This is why the panels in Figure 2 could be rearranged with little effect on its felicity, while reordering the panels in Figure 4 would result in an odd interpretation: Figure 4 features a narrative structure, while Figure 2 does not.

Empirical evidence for these narrative roles of panels comes from previous psychological studies on comics that use a "reconstruction task," where participants are asked to order a set of unordered comic panels. People appear highly proficient in accurately reconstructing the original orders of strips (Bresman, 2004; Cohn, In prep-b; Lynch & Hagen, 1996; Nakazawa, 2004). These findings indicate that people have intuitions about the global structure of a sequence, and not just individual panel relations. Some unpublished studies further indicate that panels play roles in relation to this sequence. Bresman (2004) found that panels that participants moved further away from their original positions appeared to have less relevance to the overall meaning. In contrast, panels that participants moved less featured information more central to the strip. Similarly, Lynch and Hagen (1996) found that certain semantic traits accompany panels that fall in certain positions of a sequence. Panels starting a strip often "set the stage", begin an event, and may feature a wide viewpoint (a "long shot").

Panels ending a strip often show the completion or coda of an event. These results were taken to indicate that the ordinal positions in strips ideally correlates with the perceptual organization of events.

Specific testing of the visual grammar described above has echoed these previous studies with trends shown by the distributions of narrative categories (Cohn, In prep-b). In the reconstruction task, Establishers, Prolongations, and Releases appear to be moved more and have more flexibility in their distribution than Initials and Peaks, which are more essential to an Arc's comprehension. Subjects also move these same "non-essential" categories also into positions associated with each other — Establishers, which usually appear at the beginning of sequences, are often moved to the end, while Releases, which fall at the end of sequences, are moved to the beginnings. This implies that the same panel can play different functional roles at different places in the sequence. Additional tasks in this experiment looked at which categories participants choose to exclude when asked to delete a single panel from the reconstruction task. Again, Establishers, Prolongations, and Releases are more likely than the "essential" ones to be deleted. Conversely, when a comic strip is given with one panel deleted, participants correctly identify missing Initials and Peaks far more than they recognize deleted Establishers, Prolongations, or Releases. Because panels show consistent trends in their distribution, these results imply that panels are not simply involved in ad hoc linear relations that are locally determined, but may play more abstract functional roles in relation to a global structure.

Outstanding questions

Together, the theories about sequential images and narratives imply that comprehension of graphic sequences involves tracking referential information through a combination of semantic association and narrative structure. Findings from experiments on sequential image comprehension (and verbal story grammars) imply that the union of these elements in a normal sequence yields a buildup of structure across the sequence of images (Gernsbacher, 1983, 1985), which does not seem to occur for the comprehension of scrambled sequences of images (Gernsbacher, et al., 1990). However, while these studies showed a buildup of comprehension in *memory*, these methods do not tell us about how sequential images are processed online: Can we find a "structural buildup" in online comprehension as people are actively processing sequential images? Furthermore, since the materials used in previous studies compared only fully scrambled sequences with normal narratives, it is unclear what aspects of understanding might motivate such a buildup: Semantic association? Narrative structure? To further address these questions requires online measures that extend beyond memory paradigms and that control for the contributions of semantic association and narrative structure.

This study presents two experiments to investigate the online processing of semantic association and narrative structure in sequential images. In both experiments, four types of novel *Peanuts* comic strips were created that fully crossed semantic association and narrative structure. Two online techniques were used to explore the neurocognitive processes engaged during panel-by-panel comprehension — target monitoring of panels with reaction times and event related potentials (ERPs). The findings suggest that a narrative structure is used, alongside semantic relationships, during sequential image processing.

EXPERIMENT 1: Target Monitoring

As reviewed, most psychological studies of verbal or graphic narrative have hinted that comprehenders build structure across the ordinal sequence of images. However, most support for the comprehension of discourse and sequential images has come from psychological studies using memory tasks demanding that participants recall the content of a narrative that has been disrupted (Gernsbacher, 1985; Gernsbacher, et al., 1990; Mandler & Johnson, 1977; Stein & Nezworski, 1978; Thorndyke, 1977), from online tasks where participants are asked to arrange unordered sequences (Bresman, 2004; Cohn, In prep-b; Lynch & Hagen, 1996; Nakazawa, 2005) or to define breaks in a seemingly continuous discourse (Gernsbacher, 1983, 1985). However, as has been established with 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. Indeed, this dependence on memory tasks has been a critique of story grammars being based primarily on semantic rather than truly structural relationships (Black & Wilensky, 1979; de Beaugrande, 1982). Because Gernsbacher's experiments showed a buildup in memory, this would imply that semantics alone facilitate such an effect. However, different implications have been found at the sentence level interactions between structure and semantics, as will be discussed shortly. The theory of comprehension presented here incorporates the idea that sequential images draw upon a narrative structure that is distinct from, but that functions alongside, semantic associations. The psychological reality of this theoretical distinction was explored by drawing upon a classic online psycholinguistic paradigm originally used by Marslen-Wilson and Tyler (1975, 1980) to examine how distinctions between syntax and semantics play out during language processing.

In their seminal study, Marslen-Wilson and Tyler (1975, 1980) asked participants to monitor for target words (e.g. "ideas") in normal prose (e.g. "The boy's ideas formed silently"), syntactic-but-not-semantic prose ("Colorless green ideas sleep furiously"), and randomly scrambled sentences ("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 structures influence online comprehension, even in the absence of semantic information, with processing maximally facilitated when both semantics and syntax are present. This study further showed that, within sentences containing some syntactic structure (with or without semantics), reaction times became progressively faster as the target word was positioned further along in the sentence. Scrambled sentences, in contrast, showed no decreasing trend in reaction times across target word position. These findings were interpreted as suggesting that structural buildup across a sentence increasingly facilitates processing of target words.

In the present study, an analogous "panel-monitoring" paradigm was designed that measured reaction times as readers monitored for target panels in each four types of comic panel sequences that independently manipulated narrative structure and semantic association: 1) Normal sequences were like a standard comic strip, balancing both narrative and semantics. 2) Semantic Only sequences featured panels related through semantic association, but no narrative, such as disparate images of characters playing baseball, but with no narrative structure guiding them. 3) Structural Only sequences used a narrative Arc, based on the model of narrative grammar described above, but displayed no coherent semantic relationships between individual panels (analogous to the syntactic "Colorless green ideas..." sentences). 4) Scrambled sequences use randomly ordered panels without meaningful nor narrative connections between panels. These sequences are modeled directly after the stimuli in Marslen-Wilson and Tyler (1975, 1980), except that the Semantic Only sequence types were added to counterbalance the presence or absence of both structure and semantics.

If the transitional model of sequential images is correct, then this would predict semantics to guide comprehension with no impact from structure. Because this approach would consider the Structural Only strips to be a sequence of Non-Sequitur transitions, reaction times to these types should be indistinguishable from Scrambled sequences. We would expect both of these types (Scrambled and Structural Only) to have the slowest reaction times, because of the greater demand of inference for Non-Sequitur transitions. On the other hand, Normal strips which feature more Action-to-Action and Subject-to-Subject transitions would be expected to have the fastest reaction times because of the lesser demand for inference. Semantic Only sequences, with only a semantic field, would be predicted between Scrambled/Structural only and Normal sequences.

Additionally, a transitional theory of meaning would predict minimal buildup across sequence position for all sequence types, because inference is kept locally between each pairing of panels with no expectation of additive structure. However, if themes were recognized as extending across wider distances than just local relationships, then only sequences with semantic associations (i.e. Normal and Semantic Only sequences) would be expected to show a buildup along ordinal position. With no semantic associations, Scrambled and Structural Only sequences would just be guided by local Non-Sequitur relationships, which would not contribute towards any buildup.

In contrast, if a narrative structure guides comprehension, this would predict, as in Marslen-Wilson and Tyler (1975, 1980), that participants would respond fastest to target panels in Normal sequences, due to expectations built up from a combination of both meaning and structure. Participants should be slowest to respond to panels in Scrambled sequences because no expectations are built on the basis of either semantics 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 guides a buildup of 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 increases in reaction time to monitor panels at later versus earlier ordinal positions.

Experiment 1: Methods

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 (Schulz, 2004a, 2004b, 2004c, 2005, 2006a, 2006b). *Peanuts* was chosen because 1) they have systematic panel sizes and content with repeated characters and situations; 2) their content is recognizable to most readers; 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.). In order to eliminate any influence of written language on comprehension, panels without text were selected, or panels with text were edited by deleting the text using Adobe Photoshop CS3. All individual panels were adjusted to a single uniform size.

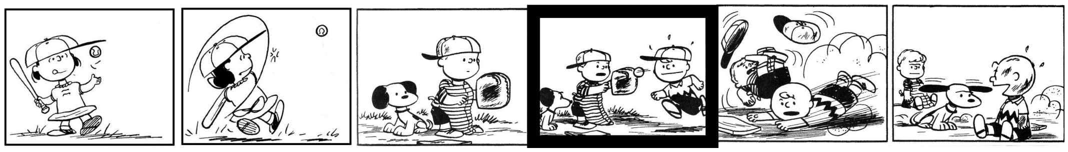
Two-hundred novel 6-panel coherent Normal sequences were initially created. Because the standard daily Peanuts strips are four panels long, experimental sequences were created by combining panels from existing comic strips. As described with its tree structure in Figure 4, the example Normal sequence type shown in Figure 4 first uses an Initial clause of Lucy hitting a baseball, which sets up a second Peak clause in which Charlie Brown slides into a base as Schroeder tries to tag him out. Across these Normal sequences, the pattern of narrative structure was varied so as not to bias the experimental results toward only one grammatical pattern (such as the basic canonical pattern).

To ensure that these novel Normal sequences were 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 the 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, see Figure 5 for examples.

In the Semantics Only sequences, each panel shared the same overall semantic or thematic field (Saraceni, 2000, 2001, 2003), 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. In the example shown in Figure 5, 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. A variety of semantic fields were used, reflecting the common themes in *Peanuts* strips: baseball, football, golf, piano playing, kite flying, weather (snow, rain, sunshine), winter activities (making snowmen, throwing snowballs, etc.), leaves falling, and others. These sequences were constructed by assigning panels to semantic fields based on cues within their images. These panels were then distributed throughout sequences in such a way to match the semantic field of the Normal sequence, yet to not make sense as a linear narrative progression.

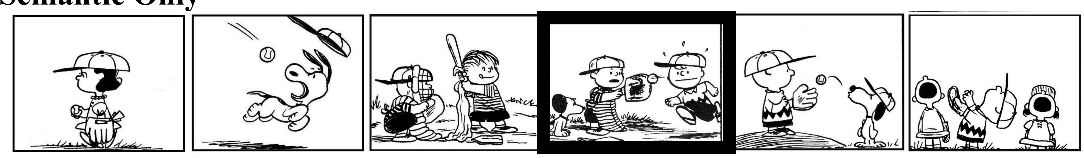
Figure 5. Example Stimuli. Critical panels have thick borders

Normal

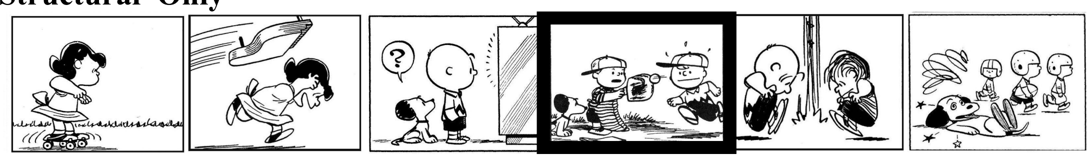


[$_{Arc}$ [$_{Initial}$ Initial Peak][$_{Peak}$ Establisher Initial Peak Release]

Semantic Only



Structural Only



[$_{Arc}$ [$_{Initial}$ Initial Peak][$_{Peak}$ Establisher Initial Peak Release]

Scrambled

In the Structural Only sequences, panels retained the same global narrative structure as the Normal sequence type but they were not semantically associated and featured different characters and themes. Thus, in combination, the sequence lacked coherent meaning. In the example shown in Figure 5, the sequence begins 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. Structural Only sequences were created using narrative categories for panels that were assigned in the context of their Normal sequences. Full sequences were constructed by matching the same narrative Arc pattern as the Normal strips, but with panels of categories drawn from numerous original sequences.

Finally, the Scrambled sequences used neither narrative structure nor semantic association between panels. In the example, the panels feature vastly different themes and no narrative arc, leaving nothing at all to unite them. These would be a sequence of "Non-Sequitur" transitions. These sequences combined disparate panels that did not share semantic fields with each other and violated the expectations of a coherent narrative progression.

In total, 191 sets of six-panel sequences were generated (764 sequences in total). Further examples of these sets are depicted in the Appendix. Within each quadruplet, the same "target panel" appeared at the identical position (in the examples in Figure 5, the target panel was panel 5, bolded). The target panels appeared from the second to sixth panel positions, with equal numbers of targets at each position. Within each quadruplet, the four sequence types were matched for number of characters per panel in each panel position.

Two additional behavioral rating studies provided objective measures of the relatedness between local panels and were used to further constrain the stimulus set, as described below.

Rating studies

Participants in rating studies were recruited either from Tufts University or through the Internet via links on the author's blog and picked up by other comic related websites. In these rating studies, participants' comic reading expertise was assessed using a pretest questionnaire asking how often they read a variety of forms of comics on a scale of 1 to 7 (comic books/sequences, graphic novels, Japanese comics, etc.), both as a child and adult. The questionnaire also gauged their familiarity with other forms of both written and film narrative. A "fluency rating" was then computed using the following formula:

$$\left(\begin{array}{c}\text{Mean Comic reading frequencies} \\ \times \;\; \text{Comic reading expertise}\end{array}\right) + \left(\frac{\text{Comic drawing freq. x Drawing Ability}}{2}\right)$$

This formula weighted fluency towards comic 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 with a mean fluency was very high, at 22.11 (SD=11.35). Self-defined "comic readers" were chosen in order to reduce the heterogeneity in the population, and ensure that participants were familiar with the materials and this manner of assimilating sequential pictures.

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 to Tufts University guidelines, and they were compensated by entrance into a raffle.

Rating Study 1: Global Ratings

40 "comic fluent" individuals with a mean age of 29.5 (29 male, 11 female) participated. Their mean fluency was high, at 22.11 (SD=11.35). 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 to 7 for how much they made sense as a whole.

Rating Study 2: Local Ratings

This study examined the local coherence of immediately juxtaposed panel pairs within the sequences. 100 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). For each of the 191 quadruplets of six-panel sequences, a set of five pairs of panels depicting the immediately juxtaposed units 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, so longer sequences could not be inferred and thereby bias ratings. However, across all participants, all pairings for a given panel were viewed. Participants were asked to rate the panel pairs along a 1 to 7 scale for how “related in meaning” the panels were.

Construction of final set of stimuli

The two rating 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, and 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 160 quadruplets. The mean ratings for the four sequence types are given in Table 1. An ANOVA revealed significant differences across the four sequence types in both global ratings, $F(3,477)=7492.1$, $p<.001$, and local ratings,

$F(3,477)=8036.94$, $p<.001$, with all pairwise comparisons between sequence types significantly different from one another both in global ratings (all $t>31$, all $p<.001$) and local ratings (all $t<26$, all $p<.001$).

Table 1. Mean ratings for experimental stimuli (on a scale of 1-7)
Standard deviations shown in parentheses

	Normal	Semantic Only	Structural Only	Scrambled
Global Ratings	6.13 (.72)	3.56 (1.05)	2.66 (.73)	2.39 (.69)
Panel positions 1--2	5.85 (.93)	4.44 (1)	3.05 (.99)	2.54 (1.05)
Panel positions 2--3	3.81 (1.07)	3.21 (1.02)	2.48 (.93)	4.94 (.86)
Panel positions 3--4	2.97 (.89)	2.56 (1.01)	4.70 (.78)	3.59 (.94)
Panel positions 4--5	2.60 (1.09)	4.05 (.76)	3.78 (.91)	2.65 (.99)
Panel positions 5--6	2.60 (1.09)	4.19 (1.04)	2.84 (.96)	2.46 (1.02)
Average Local	5.60 (.67)	3.66 (.82)	2.70 (.54)	2.46 (.44)

The stimuli were then counterbalanced using a Latin Square design across four lists, each to be seen by an individual participant. This allowed for each participant 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) 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 are 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 Scrambled. With both experimental and filler sequences, a balance overall between semantic and non-semantic strips was sought, with the overall proportion of strips being 30% Normal strips, 20% Semantic Only, 20% Structural Only, and 30% Scrambled. Within each list, the order of experimental and filler sequences was randomized.

Participants in the Panel Monitoring experiment

54 experienced comic readers (30 male, 24 female), recruited from the Tufts University undergraduate population, with a mean age of 20.4 (SD=1.68), were paid for participation. All participants gave their informed written consent to Tufts University 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). Data from two participants was discarded due to their not reaching a threshold of 80% accuracy in the task.

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 grey lettering. When the participant pressed the keypad, a fixation-cross appeared in the center of the screen for 1500ms with a 300ms ISI, followed by the target panel. Target panels remained on screen for 2500ms to allow participants to examine its features carefully. Following another 300ms ISI, another fixation cross appeared for 1500ms followed by successive panels, each remaining on the screen for 1500ms with an ISI of 300ms. This duration of ISI prevented the appearance of the sequences turning into a "flip-book" style animation. The 1500ms duration was used because a pilot self-paced reading study showed that this was the average time spent reading 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. At the end of the sequence, a screen reading READY again appeared for 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 was asked about various properties of the sequence (e.g. "Was Snoopy scared?", "Did Snoopy swallow the ball?", etc.). These questions were aimed at keeping participants attentive to reading the sequences for comprehension as opposed to 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.

Analysis of data

Accuracy for button presses was computed as the percentage of responses in which the participant 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 1) omissions or 2) presses to a panel other than the target.

Analysis of the RTs only used correctly answered responses and, in each participant, outlier reaction times – more than 2.5 standard deviations from the mean response within a given condition – were discarded. Collapsed across target panels, RTs to button presses were analyzed using a 4 (Sequence type) x 5 (Position) repeated-measures ANOVA for both subjects and items. In the subjects analysis both Sequence type and Position were within-subjects/items factors, while in the items analysis they were between-subjects factors. Main effects of Sequence type were followed up using planned t-tests comparing each Sequence Type with one another. Main effects of Position in both the subjects and items analyses were followed up using polynomial contrasts to determine whether there were linear trends across each ordinal position in the sequence. Regressions placing position as the predictor then followed this trend analysis to examine the trends of individual sequence types across ordinal position. Finally, the effect of comic reading fluency on reaction times and accuracy were analyzed using a Pearson’s correlation set to an alpha level of .05.

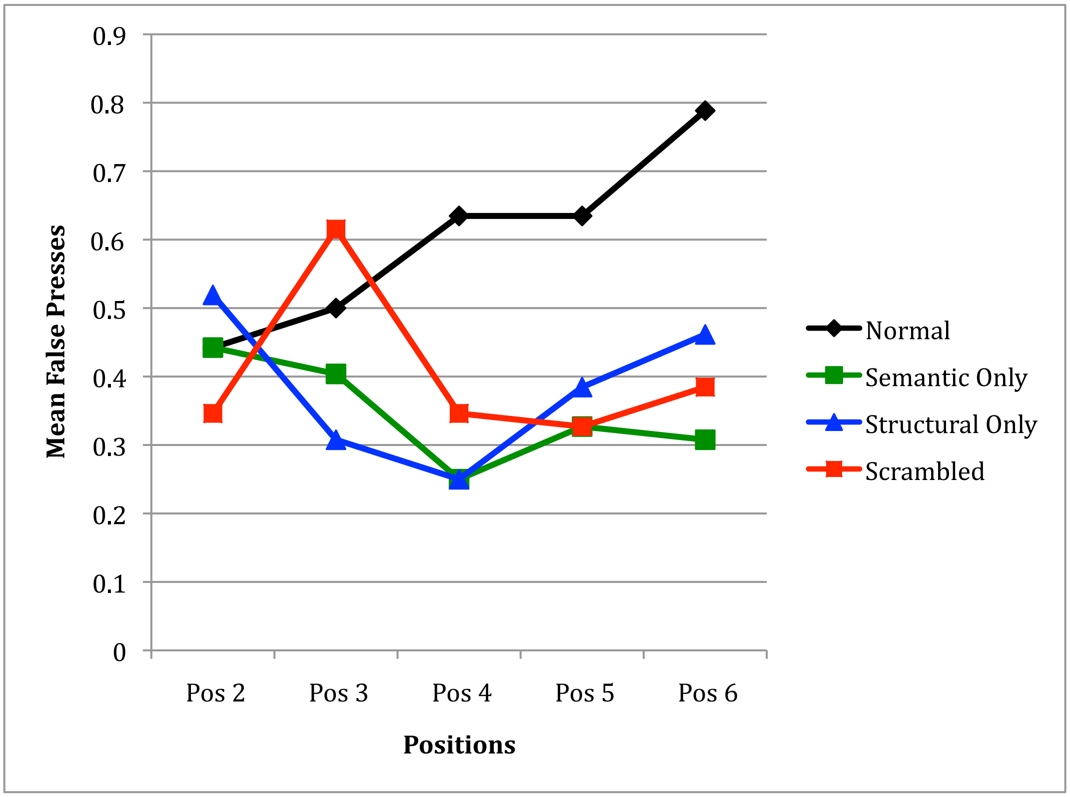
Experiment 1: Results

Accuracy

Participants’ accuracy across all sequence types was 93%. Accuracy to the Normal sequences was worst (90% correct; mean: 36 out of 40 (SD=2.90)), compared to 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). A four-way ANOVA confirmed significant differences in participants’ accuracy in monitoring targets across the four Sequence Types, $F(3,153)=14.53$, $p<.001$, with significantly reduced accuracy on Normal sequences compared to all other Sequence Types (all $p<.001$), but no significant differences between any other Sequence Types (all $ps>.600$).

The Normal sequences also had the highest rates of false presses (7.5% of trials), compared with 4.8% for Semantic Only, 4.3% for Structural Only, and 5% for Scrambled. A 4 (Sequence type) x 5 (Position) ANOVA for false presses yielded significant main effects between sequence types, $F(3,153)=8.14$, $p<.001$, but not for positions, $F(4,204)=.845$, $p=.498$, though did find a significant interaction between Sequence Types and Positions, $F(12,612)=2.11$, $p<.05$. This interaction reflected differences between sequence types in terms of how many false presses occurred across ordinal sequence position, as depicted in Figure 6.

Figure 6. Mean false presses for sequence types across ordinal positions



Normal sequences steadily increased in false presses across sequence positions. Scrambled sequences first increased between second and third positions, rapidly decreased at fourth position, and then increased through the sixth position in the amount of false presses. Semantic Only and Structural Only both first decreased in false presses through the third position, only to show a trend of increasing again through later positions. Follow up regressions confirmed these observed trends, setting Position as the predictor for each individual Sequence Type. Significant linear effects were found for the false presses to Normal sequences $b=.141$, $t(258)=2.291$, $p<.05$, with significant variance being attributed for in positions, $R^2=.02$, $F(1, 258)=5.25$, $p<.05$. Additionally, significant quadratic effects were shown for Structural Only sequences, $b=-.984$, $t(257)=-2.33$, $p<.05$, with a trending proportion of variance across positions, R^2=.021, F(2,257)=2.74, p=.066. Finally, Scrambled sequences also showed a significant cubic trend, b=.491, t(256)=2.099, p<.05, though positions were not able to account for a significant variance of the data for this sequence type, R^2=.019, F(3,256)=1.67, p=.174.

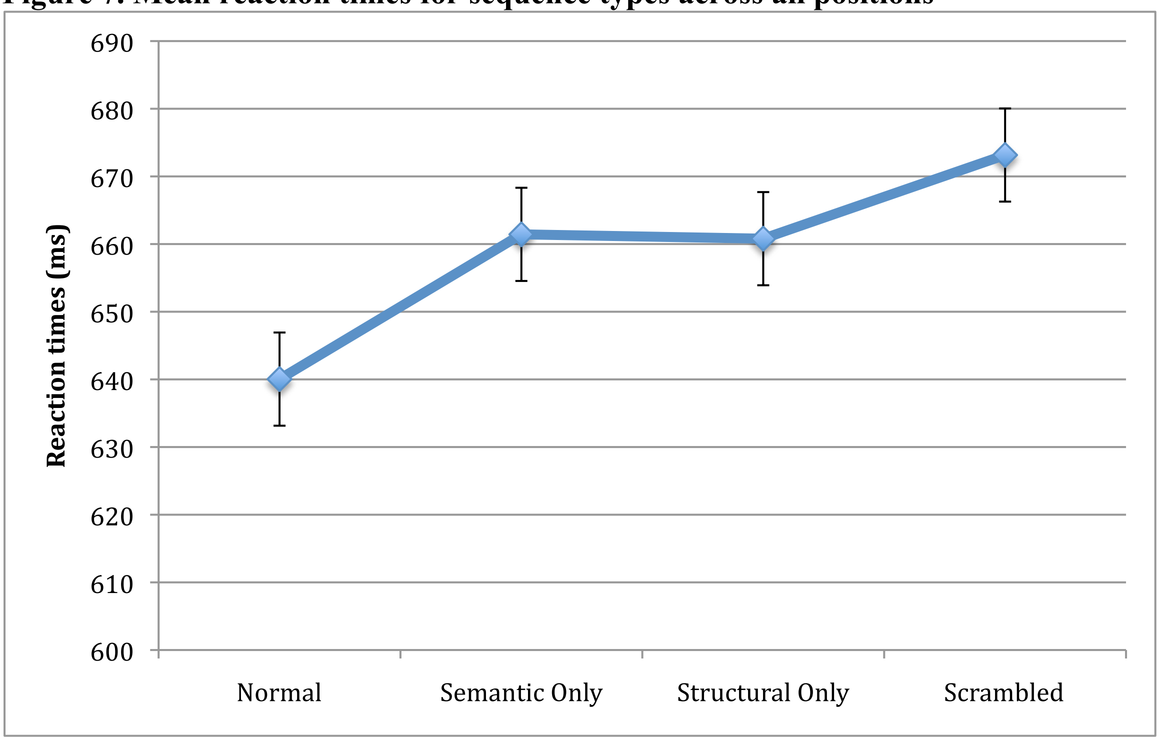
Finally, the post-test questionnaires indicated that most participants noticed a distinction between the Scrambled and Normal strips. 46% of participants explicitly commented that the Semantic Only sequences featured "themes" of meaning, though no participant picked up on any difference between the Structural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

Reaction Times for Target Monitoring

An overall 4 (Sequence Type) x 5 (Position) repeated-measures ANOVA showed a main effect of Sequence Type (F1(3, 153)=7.29, p<.001; F2(3, 93)=2.33, p=.079) (although not in the items analysis); a main effect of Position (F1(4, 204)=36.76, p<.001; F2(4, 124)=22.05, p<.001); and a significant interaction between Sequence Type and Position (F1(12, 612)=2.19, p<.05; F2(12, 372)=2.1, p<.05).

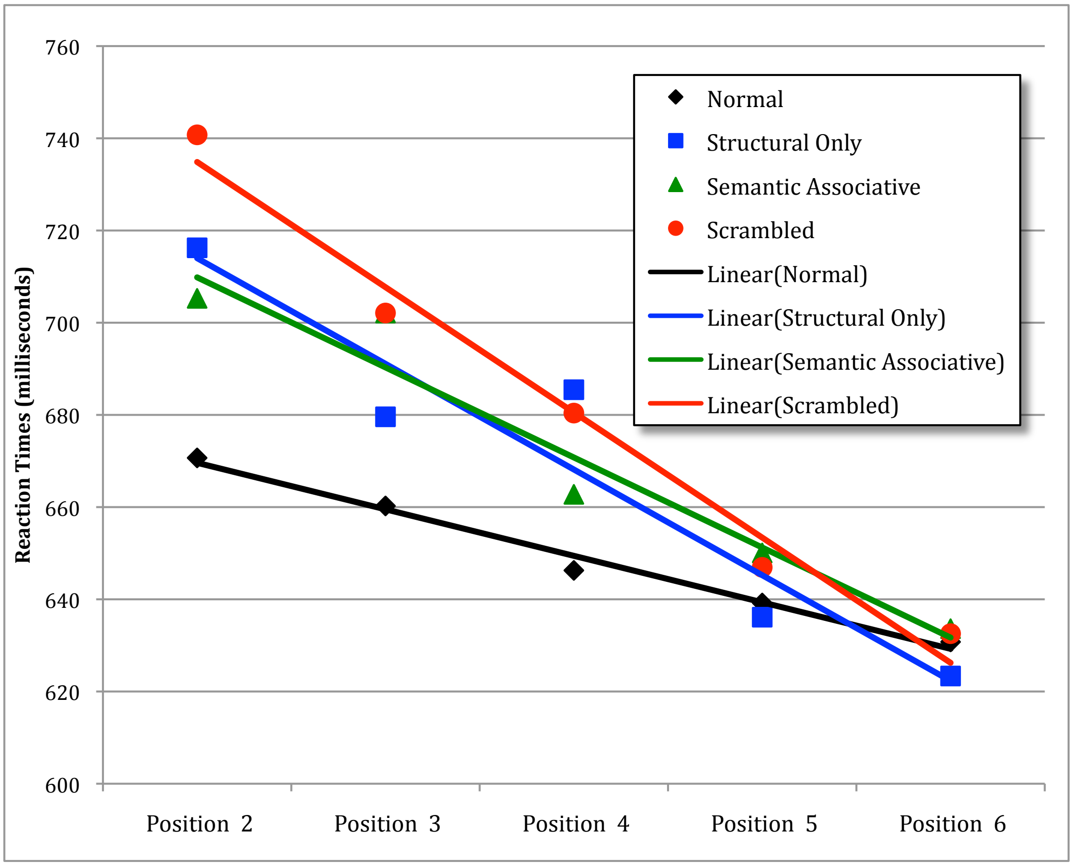
The main effect of Sequence Type arose because targets in the Normal sequences were recognized fastest (though least accurately, see above), while targets in the Scrambled sequences were recognized slowest, with RTs to targets in the Structural Only and Semantic Only sequences in-between (see Figure 7). Planned pairwise comparisons using paired t-tests confirmed this pattern of findings: there were significant differences between RTs to targets in Normal strips and all other Sequence Types (all ts < -3.2, all ps < .005). RTs to targets in Scrambled sequences were longer than in the Structural Only sequences (reaching significance, t(51)=-2.14, p<.05), and also longer than in the Semantic Only sequences (trending towards significance, t(51)=-1.99, p=.051). There were no significant differences in RTs to targets in the Structural Only and Semantic Only sequences, t(51)=.133, p=.895.

Figure 7. Mean reaction times for sequence types across all positions



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 Figure 8). This was confirmed by examining the linear trends of the polynomial contrasts in both the subjects and items analyses, which revealed significant overall linear effects of Position ($F1(1, 51)=138.26$, $p<.001$; $F2(1, 31)=76.90$, $p<.001$).

Figure 8. Mean reaction times for panels in sequence types across ordinal positions



The Sequence Type x Position interaction in the overall ANOVA reflected differences between the four sentence types in the rate of RT decrease across linear position, as illustrated in Figure 8 which shows the linear effects of position for each Sequence Types. The Normal sentences showed the least decrease in RT across position; the Semantic Only and Structural Only sequences had steeper slopes, and the steepest reduction in RT across Position was seen in the

Scrambled sequences. This observation was confirmed by significant differences across the four Sequence Type using polynomial contrasts for the linear trend component (F1(1, 51)=7.183, p<.05; F2(1,31)=10.89, p<.005. Follow up regressions for each individual Sequence Type, setting Position as the predictor, showed significant linear effects for all Sequence Types, summarized in Table 2.

Table 2. Regressions across panel positions for each sequence type for a subjects analysis

	Normal	Semantic Only	Structural Only	Scrambled
r=	.132	.330	.274	.375
b=	-.002	-.005	-.004	-.005

r= correlation coefficient, b= slope in milliseconds (n=5)

As a result of these differences between Sequence Type in the linear trends across Position, differences across Sequence Types were maximal at the second position, and were absent by the fifth position. To characterize this further, four-way ANOVAs were carried out examining the effect of Sequence Type at each Position. Significant main effects of Sequence Type were 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, there were significant differences in RTs between Normal sequences and all other types (all t<-2.2, all t<.05), and between Semantic Only and Scrambled sequences, t(52)=-2.32, p<.05. At Position 3, there were significant differences in RTs between Normal sequences and Semantic Only, t(52)=-2.69,

$p<.05$, and Scrambled sequences, $t(52)=-2.84$, $p<.05$. There were trends for differences between Structural Only and Semantic Only, $t(52)=1.76$, $p=.085$, and between Structural Only and Scrambled, $t(52)=-1.93$, $p=.059$, sequences. At Position 4, Normal sequences were significantly different from Structural and Scrambled sequences (all $t<-2.6$, all $p<.05$).

Relationships with comic reading fluency

Correlations were carried out between ratings of comic reading fluency and each individual's mean RTs in Sequence Type. A significant negative correlation was found between Fluency and RTs in the Structural Only sequences, $r(52)=-.29$, $p<.05$, RTs in the Semantic Associative sequences, $r(52)=-.28$, $p<.05$, and RTs in the Scrambled sequences, $r(52)=-.34$, $p<.05$, i.e. the more experience with reading comics, the faster the RTs in these sequences. No such correlation was seen in the Normal sequences, $r(52)=-.14$, $p=.336$.

Additionally, correlations were run on the differences for reaction times between Sequence Types compared with Fluency. A significant positive correlation was found for Fluency with the difference in RTs between Normal and Scrambled sequence types, $r(52)=.302$, $p<.05$. This correlation indicated that the disparity in reaction times between these sequence types (Scrambled being slower than Normals) increased with greater comic reading fluency. No significant correlations were found for the differences between any other types.

Experiment 1: Discussion

In this study, reaction times were measured as viewers monitored for a target panel while reading various sequences of images one panel at a time. As predicted, panels in Normal sequences received the fastest reaction times while those in Scrambled sequences were the slowest. Panels in both Semantic Only and Structural Only sequences showed comparable reaction times, falling directly between those of the Normal and Scrambled sequence types. The times for Semantic Only and Structural Only sequences were not significantly different from each other, but Normal and Scrambled sequences showed significant contrasts with all other types. This gradation in reaction times for target panels across Normal, Structural Only, and Scrambled sequences mirrors the times observed by Marslen-Wilson and Tyler (1975, 1980) for target words in studies using verbal sentences. Additionally, all sequence types showed a decrease in reaction times across ordinal sequence positions. Relevant contrasts will be addressed in more detail below, and these data will be argued to support the hypothesis that sequential images use a narrative grammar that extends beyond semantic associations to guide comprehension.

In Marslen-Wilson and Tyler (1975, 1980) reaction times were faster to critical words in Normal sentences than reaction times to Syntactic Only sentences. This was taken to support the idea that syntax acts together with semantics to construct the overall comprehension of sentences. The results of Experiment 1 suggest an analogous result, with critical panels in Normal sequences showing faster reaction times to those in Structural Only sequences.

Since these sequence types feature a narrative structure without semantics, it analogously suggests that structure and semantics combine in the comprehension of Normal sequential images. However, an alternative interpretation is that, since the Structural Only and Normal sequences were distinguished *only* by the presence of an overall semantic theme in the Normal sequences and its absence in Structural Only sequences, any facilitation of reaction time for the Normal sequences reflected non-specific priming between panels. Because of this, the experiment added the control of the Semantic Only condition, which was not used in the Marslen-Wilson and Tyler studies. The faster reaction times to target panels in the Normal than in the Semantic Only sequences suggests that they used information over and above simple semantic association in comprehension.

The most compelling evidence that participants were using a narrative structure comes from their faster reaction times to critical panels in Structural Only than to Scrambled sequences. Local coherence ratings between the Scrambled and Structural Only sequences differed only marginally, though it was significant. However, in contrast, Semantic Only sequences had nearly the same reaction times as those with only structure, though they with a far greater difference in coherence relations from Scrambled sequences than did Structural Only sequences. Thus, if the effects of local coherence relations were attributed solely to semantic effects, Semantic Only sequences would be expected to show faster times than Structural Only sequences. This suggests that such a structure was being used regardless of semantic content. Structural Only sequences were distinguished from Scrambled sequence types in being derived from a theoretical structure of abstract narrative categories. These sequences had grammatical structures that were matched to their Normal counterparts, while the Scrambled sequences lacked structuring. This therefore provides evidence that such a structure is employed during sequential image comprehension.

Despite the reaction time data distinguishing between the non-semantic strips with and without structure, no participants reported that they noticed any difference between the Scrambled and Structural Only sequence types. This suggests that readers were using this grammar implicitly during processing rather than through conscious awareness.

Finally, the observation that monitoring times for target panels decreased across ordinal position in sequences with narrative structure mirrors that of Marslen-Wilson and Tyler (1975, 1980) who showed a similar decrease in reaction time across ordinal position to words in sentences with structure. However, contrary to the results in Marslen-Wilson and Tyler (1975, 1980), who showed no decrease in reaction times to Scrambled sentences, times for all sequence types in the current study sped up across ordinal position. Here, Normal sequences actually resulted in the shallowest increase in speed across ordinal position, perhaps because they maxed out in possible speed more than the other types. In the current study, the strips without narrative structure (i.e. the Scrambled and Semantic Only sequences) actually showed the greatest decrease in reaction times in monitoring targets across ordinal position: differences between sequence types were maximal at the second position but by the final positions, there were no differences in RTs across sequence types.

Overall though, the largest separation between sequence types appears at the beginning of the sequence. This shows that readers are able to identify characteristics of the sequence even based on the relationship of a single pair of panels. However, Gernsbacher (1983) has shown that the first images of a picture story are viewed slower than later occurring pictures, a trait she attributes to "laying the foundation" of what the sequence is about (Gernsbacher, 1990). These findings are similar to self-paced reading pilot data for this present study that showed longer reading times towards the beginning of strips. Here, the slower reaction times overall at the start of the sequence may reflect this greater need for "laying the foundation." Establishing a recognition of the structure would thereby maximize reading time at the beginning of the sequence, since only one panel's influence could hint at the overall strip's character, thereby resulting in greater separation between reaction times for different sequence types. However, as the sequence progressed, the convergence of the faster times may reflect the establishment of what to expect out of the sequence structure. Such a result may even disregard an attempt at comprehension in favor of the probe task once a participant realizes a sequence makes little sense.

Furthermore, aspects of the probe task may have negatively influenced the interpretability of the reaction time data across sequence position. Given that false presses on the whole increased along serial panel position, participants' memory for the target panels may have been reduced as strips progressed. Such a degradation in memory for target images would be consistent with studies using picture stories showing worse accuracy for recall of images along sequence position (Gernsbacher, 1983). Indeed, false presses for Normal sequences featured a steady increase across sequence position. These sequences likely had greater false presses because images in Normal sequences can be mistaken for one another with more frequency because of similar content, however, the increasing trend further along a sequence indicates problems retaining the accurate panel in memory. Meanwhile, Structural Only and Semantic Only sequences appear to be U-shaped — first decreasing in false presses in the second through fourth positions, then increasing in the fifth and sixth positions. However, all sequence types showed a relative increase in false presses in the final two positions over the fourth position — the same place that reaction times decrease and converge for all sequence types. These results would imply that the demands of keeping the target in memory might be impairing accurate button presses. The decreased speed in reaction times across ordinal position could then be reflecting an abandonment of the task of comprehending strips in favor of successful completion of the probe task, thereby yielding similar fast times for all sequence types.

The reaction time data across ordinal position here indicates that viewers are sensitive to sequence type even with the cues from a single pairing of panels. However, with the rapid speeding up of reaction times across ordinal position for all sequence types, it remains unclear whether all types are facilitating a buildup of structure, or if additional confounds are impairing any differentiation. This could be due to the memory task in target monitoring or an expectation of sequence type lead to a disregard of comprehension across sequence position. If such limitations were providing confounds, they could be overcome with a more sensitive measure that does not rely on behavioral responses like reaction time. As discussed below, this limitation was overcome in Experiment 2, in which neural activation was measured at all panels without requiring participants to monitor for specific targets in a behavioral task.

EXPERIMENT 2: Event Related Potentials

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 but has some limitations. The task of monitoring for specific target panels may have interfered with the comprehension of the sequence types. As discussed above, this may have led to all sequence types converging on similar reaction times by the end of sequences, masking any differences between types. To overcome these limitations, a second experiment used Event related potentials (ERPs) — multidimensional measures with excellent temporal resolution that directly measure underlying neural processes and which, in principle, are not dependent on the performance by participants in a behavioral task.

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) identified this component as a negative deflection in the waveform peaking around 400ms that was smaller (less negative) in amplitude to words that were semantically congruous relative to those that were semantically anomalous or unexpected with their preceding contexts. The amplitude of the N400 is attenuated by featural overlap or semantic association 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 & Ditman, In Press; St. George, Mannes, & Hoffman, 1997; van Burkum, Hagoort, & Brown, 1999; Yang, Perfetti, & Schmalhofer, 2007). The modulation of N400 amplitude to semantic congruity is termed the 'N400 effect' and 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, processing of each upcoming word is progressively facilitated. Critically, however, just as in the RT 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 et al. (1975, 1980), Van Petten and Kutas (1991) showed that N400 evoked by a given 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 a semantically unexpected stimuli, but rather that that it reflects default semantic processing of all meaningful stimuli, which is facilitated 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 in the N400 amplitude evoked by words in Scrambled sentences and Structural Only sentences. This suggested that it was not sensitive to structure alone, in the absence of semantics.

Taken together, all these findings suggest that the N400 reflects a process of relating the meaning of an incoming word with its preceding context and with information stored within semantic memory. Although the amplitude of the N400 is influenced by the combination of structure and semantics in building a coherent context, it does not itself *directly* reflect structural processing or the integration of structure and meaning.

The component that is thought to be sensitive to the demands of integrating structure and meaning in language is a late Positivity or P600 waveform – a centro-parietally distributed positive deflection that extends beyond the N400 time window and that peaks from 600-800ms. The P600 was first reported in association with syntactic ambiguities in garden-path sentences (Lee Osterhout & Holcomb, 1992), as well as to frank syntactic anomalies within sentences (P. Hagoort, 1993). Later studies found a P600 to the resolution of question words (“who”) with verb phrases connected through a long-distance dependency (Kaan, Harris, Gibson, & Holcomb, 2000). Importantly, the P600 appears to be evoked by structural violations only in the presence of some semantic information: Munte et al. (1997) found that agreement violations between nouns and verbs showed a P600 only in sentences with semantic content, but not sentences constructed with pseudo-words (though Hahne and Jescheniak (2001) did find a P600 when pseudo-word sentences were presented auditorily). In addition, the amplitude of the P600 to syntactic violations is modulated by the preceding semantic context

(Gunter, Friederici, & Schriefers, 2000). Indeed, recent studies have shown that the P600 can be evoked by severe semantic violations within sentences, particularly in semantically constraining contexts (Kuperberg, 2007; van de Meerendonk, Kolk, Vissers, & Chwilla, 2010). Based on these observations, the P600 has been interpreted as reflecting continued analysis or reanalysis operations that combine structure with semantic content (Kuperberg, 2007), and which can be triggered by conflict between a highly implausible syntactically-determined interpretation and a match in the semantic memory-based analysis (Kuperberg, 2007).

A second component that sometimes (although not always) accompanies the P600 effect that has been associated with structural processing in language is a *left anterior negativity* (LAN) falling between 300 and 500ms, and distributed over frontal electrode sites (sometimes in left lateralized regions). The LAN has been tied to a number of syntactic operations. It is sometimes (although not always — see Peter Hagoort, Brown, & Groothusen, 1993; Kuperberg, Caplan, Sitnikova, Eddy, & Holcomb, 2006; Kuperberg, Kreher, Sitnikova, Caplan, & Holcomb, 2007; Kuperberg, Sitnikova, Caplan, & Holcomb, 2003) seen in association with morphosyntactic agreement violations between noun phrases and verbs (Peter Hagoort & Brown, 2000; L. Osterhout & Mobley, 1995), and long-distance dependencies in non-grammatically violated sentences (Kluender & Kutas, 1993). Unlike the P600, the LAN has been observed to violations in the structure of "Jabberwocky" sentences that use nonsense words (i.e. no semantics vs. sentences with semantics) but that still have syntactic structure (Münte, et al., 1997). Because a P600 was found to sentences with semantics, but only the LAN was found for Jabberwocky sentences that lacked semantics, they suggest that this negativity more directly indexes a property of structural incongruity. Functionally, the LAN has been linked with working memory operations that look both forward and backwards within a sentence to reconcile structure from various places, be it for phenomena like displaced constituencies (Lee Osterhout & Holcomb, 1992) or distance dependencies (Kluender & Kutas, 1993).

In the visual modality, most ERP studies have focused on static images. Just like words, meaningful images evoke an N400. The N400 effect is less for target pictures congruous to their context than for those that are incongruous, whether the context is (a) preceding single pictures in priming paradigms (McPherson & Holcomb, 1999), (b) the context of an overall surrounding scene (Ganis & Kutas, 2003) or (c) a verbal sentence (Ganis, Kutas, & Sereno, 1996). Along with this N400, a preceding negative peak, onsetting at around 200-250ms and peaking at around 300 milliseconds after the onset of the stimulus, termed the N300, has also been observed in the visual domain alone (Barrett & Rugg, 1990; Ganis, et al., 1996; McPherson & Holcomb, 1999). This N300 often overlaps with the N400 and has a more frontal distribution than the N400 observed in most language studies (McPherson & Holcomb, 1999), possibly reflecting a more rapid access to the semantic features of objects than to 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 for 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. The N400 had a more anterior distribution (peaking in right centro-frontal regions) and a longer duration than observed in language studies, but it was still more widespread and lateralized than its preceding anteriorly distributed N300. The N300 had a slightly later onset than that seen to static images, appearing at around 275-300ms and peaking around 325ms, suggesting that sequential image processing may engage similar mechanisms but demands more effortful retrieval.

These sequential image results have also been observed to ERPs from written narratives, in two experiments by West (1998), repeating the same paradigm of congruous versus incongruous sequence endings. One experiment examined the ERPs to written narratives that substituted verbal descriptions of each image of the same sequential image stimuli. While full sentences were presented in the story stem of the experiment, the sentences ending in text were presented word by word, and words were deemed as "critical" by participants' intuitions as to what cued their congruity judgments. These critical words elicited an N400 effect, but no N300, in a more posterior scalp distribution. The second experiment used the same verbal descriptions, but ended the sequence with a critical image taken from the imagistic stimuli. Here, incongruous endings again produced a sustained N400 preceded by an N300, distributed in a more posterior area on the scalp. While both images and text had a more rightward distribution, the anterior distribution of the image sequences differed from the more posterior distribution for textual sequences with either text or images as endings. Because the differing distribution of N400 effects of these narratives, these results were interpreted as evidence that the semantic processing involved in discourse comprehension uses modality specific structures.

Another set of studies by Sitnikova, Holcomb, and Kuperberg (2008b) examined the processing of visual images appearing within short, silent movie clips depicting everyday events. Final scenes of movie clips that were congruous with their context (e.g. a man preparing to cut bread and then cutting the bread with a knife) were contrasted with incongruous final scenes (e.g. a man preparing to cut bread and then cutting the bread with a knife followed by a scene of him ironing a shirt). Again, an anteriorly-distributed N300/N400 effect was observed (Sitnikova, et al., 2008b), which was once again interpreted as reflecting a mapping of the meaning of the final event on to the meaning of the context. Interestingly, when the authors introduced a final scene depicting an 'action-violating' event in which the central action was being carried out with an object that did not possess the semantic properties to carry it out (such as a man attempting to cut bread with an iron), an additional posterior positivity starting around 500ms and lasting until 800ms was observed (Sitnikova, et al., 2008b; Sitnikova, Kuperberg, & Holcomb, 2003). This was interpreted as being somewhat analogous to the P600 evoked to highly implausible events in constrained contexts in language (see above); the authors speculated that it once again reflected additional processing as comprehenders attempted to relate the semantic properties of the object to structural semantic constraints of the predicted central action (Sitnikova, Holcomb, & Kuperberg, 2008a).

Taken together, these studies indicate that, just as in language, comprehenders are able to use sequential visual information to influence semantic processing of an upcoming segment, and that prolonged processing can be engaged when the depicted image violates expectations about central actions and events. It remains unclear, however, how and if such expectations are being built additively across sequence positions. Are comprehenders able to draw upon a narrative structure in combination with semantic information to influence processing of each upcoming image?

The present study aimed to address this issue by determining whether the narrative structure of a visual discourse influenced the semantic processing of an upcoming image over and above simple semantic associations. The main focus was on the N300/N400 complex evoked by each visual image in the sequence. The same stimuli were used as in Experiment 1 but this time, rather than measure ERPs only to target panels, neural activity was measured to all panels in the sequence, mirroring the study design and logic of the language study reported by Van Petten and Kutas, described above. An N300/N400 complex was predicted to be evoked by all visual images in coherent sequences and the amplitude of this complex would decreases to successive panels in the sequence. This would be taken as evidence that, as context is built up the sequence, semantic processing of each panel is progressively facilitated. If the buildup of this visual context is dependent on a progressive combination of structure with the semantics of individual panels, this would predict two main results: First, there should be no such progressive attenuation of the N300/N400 with increasing ordinal position in either Structural Only or Scrambled sequences, i.e. those sequences with structure but no semantics. Second, the N300/N400 evoked by a given panel (collapsed across all ordinal positions) should be smaller in coherent sequences than in either Structural Only or Scrambled sequences. Moreover, given that the N300/N400 in language does not reflect structural integration per se (see Van Petten and Kutas' findings in language), there should be no difference of the N300/N400 between panels in Scrambled versus Structural Only sequences.

Experiment 2: Methods

Participants

24 Tufts University undergraduates with a mean age of 19.4 (SD=1.67) (12 male, 12 female) 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 psychiatric drugs. All participants completed the comic fluency questionnaire (described under Experiment 1) and had a mean fluency rating of 16.99 (SD=6.36).

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. The longer fillers were also altered to consistently be six panels long, reducing the overall length of the experiment and the likelihood of participants' blinking due to the challenge of maintaining open eyes during presentation of lengthy experimental stimuli. Each list of experimental and filler stimuli was randomized within lists.

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 a screen reading 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 1500ms, followed by a 300ms ISI, and then the first panel of the sequence appeared centered on the screen. Each panel remained on the screen for 1500ms with an ISI of 300ms. An ISI of 500ms 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 were reading the sequences carefully for full comprehension. Questions were designed to address events related to target 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.

ERP Recordings

ERPs were measured using an elastic cap with twenty-nine 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), four lateral sites on each hemisphere (FC1/FC2, C3/C4, CP1/CP2), and five peripheral sites (FP1/FP2, F7/F8, T3/T4, T5/T6, O1/O2). All electrodes were referenced to an electrode placed on the left mastoid, while differential activity was monitored in the right mastoid.

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

Behavioral Data Analysis

Accuracy for the plausibility judgment 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 scores had an accuracy rate less than 80%.

ERP Data Analysis

ERPs were time-locked to each panel (at positions 2 – 6). Analyses of ERPs used ANOVAs along the midline column (five electrodes), medial columns (three electrodes each), lateral columns (four electrodes each), and peripheral columns (five electrodes), depicted in Figure ERP. Within-subject factors were the four levels of Sequence Type, Anterior-Posterior (AP) Distribution (with levels corresponding to electrodes in a column), and Hemisphere (two levels) for analyses off the midline. Main effects and interactions were followed by simple effects ANOVAs when necessary.

Following West and Holcomb (2002), mean voltage analysis was conducted within the windows of 300-400ms, 400-600ms, and 600-900ms to investigate the presence of the N300, N400, and sustained negativity effects.

To further examine the ERPs across a sequence, for each sequence type, amplitudes at each position were averaged across select electrode sites and compared using regressions placing position as the predictor. Finally, participants' ratings of comic reading fluency were compared to the difference amplitudes between sequence types to determine any effect of fluency. These analyses will be undertaken using a Pearson’s correlation set to an alpha level of .05.

Experiment 2: Results

Behavioral Data

Participants’ accuracies in judging the plausibility of each type of sequence are summarized in Table 3. A four-way ANOVA found a significant main effect of Sequence Type, $F(3,69)=8.80$, $p<.001$, and follow-up pairwise analyses showed 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 or <-5.27, all ps<.005). This may be partially attributed to semantic associations within individual strips being construed as “making sense” more than the other infelicitous sequence types. There were no significant differences in accuracy between the Normal versus Structural Only, and Normal versus Scrambled sequences (ts <.390, ps >.40).

Table 3. Mean accuracy for plausibility judgments for differing sequence types (standard deviation shown in parentheses)

Sequence type:	Normal	Semantic Only	Structural Only	Scrambled	Filler
Mean Correct	.92 (.08)	.70 (.28)	.87 (.25)	.89 (.24)	.92 (.07)

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. (Note: almost half of all participants (10 of 24) made no explicit notice of particular properties on the questionnaire).

ERP Data

300-400: N300

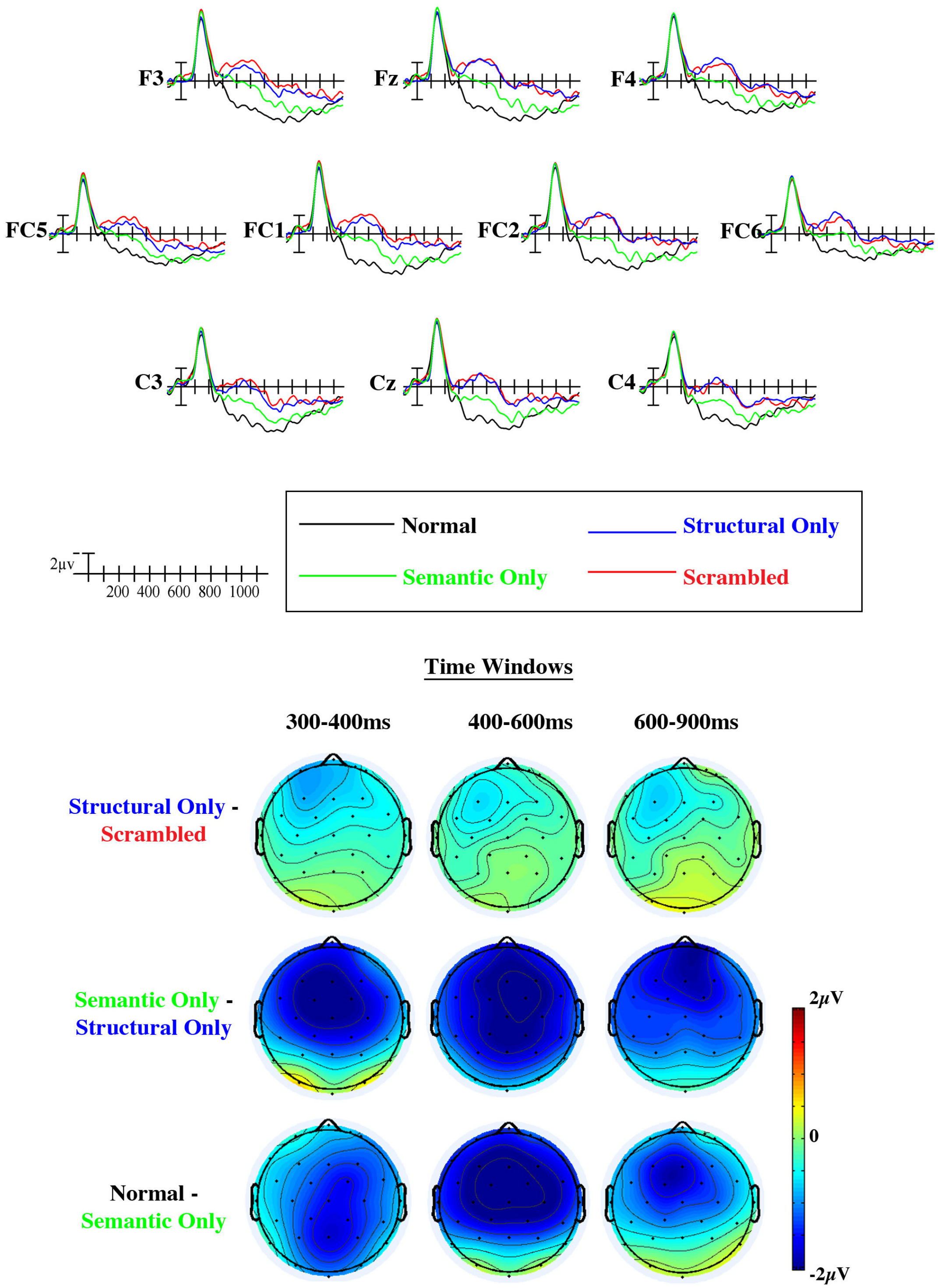
Averaged across all panels, a significant negative deflection was observed starting between 250-300ms and lasting until 400ms, with a peak around 300ms after the onset of the stimuli panel – the N300 component. A four-way repeated measures ANOVA, comparing the four sequence types (collapsed across ordinal position) showed significant main effects of Sequence Type (all $F>36.05$, all $p<.05$), and/or interactions with AP distribution (all $F>6.62$, all $p <.05$) at all electrode columns. The amplitude of the N300 appeared to increase progressively across the four conditions. The N300 effect was greater in the Semantic Only than in the Normal sequences. The difference was widespread but maximal at anterior sites (main effects of Sequence Type and interactions between Sequence Type and AP Distribution at all columns), see Table 4, Figure 9. In turn, the N300 was greater in the Structural Only sequences than in the Semantic Only sequences, and this effect was again widespread (main effects of Sequence Type at all columns), but maximal at anterior sites (interactions between Sequence Type and AP Distribution at all columns) and at slightly and rightward medial and lateral sites (interactions between Sequence Type and Hemisphere at medial and lateral columns). N300 amplitudes were slightly larger still in the Scrambled than in the

Structural Only sequences, but this effect was less widespread (no main effects of Sequence Type) and apparent only at anterior sites (interactions between Sequence Type and AP Distribution at all columns).

Table 4. Statistical analysis of waveforms in the 300-400ms time window ST= Sequence Type, H= Hemisphere, AP= AP Distribution, *= p < .05, ^=p<.10

	Midline	*Medial*	*Lateral*	*Peripheral*
Norm-Sem	**ST** [F(1,23)=28.43]* **ST x AP** [F(4,92)=18.07]*	**ST** [F(1,23)=49.20]* **ST x AP** [F(2,46)=8.31]* **ST x AP x H** [F(2,46)=2.77]^	**ST** [F(1,23)=34.34]* **ST x AP** [F(3,69)=12.28]*	**ST** [F(1,23)=12.78]* **ST x AP** [F(4,92)=18.36]*
Norm-Str	**ST** [F(1,23)=96.03]* **ST x AP** [F(4,92)=20.90]*	**ST** [F(1,23)=134.65]* **ST x AP** [F(2,46)=6.41]*	**ST** [F(1,23)=105.18]* **ST x AP** [F(3,69)=9.07]*	**ST** [F(1,23)=44.53]* **ST x AP** [F(4,92)=17.74]*
Norm-Scram	**ST** [F(1,23)=100.26]* **ST x AP** [F(4,92)=24.96]*	**ST** [F(1,23)=133.89]* **ST x AP** [F(2,46)=10.82]*	**ST** [F(1,23)=118.14]* **ST x AP** [F(3,69)=18.25]* **ST x AP x H** [F(3,69)=2.96]*	**ST** [F(1,23)=66.59]* **ST x AP** [F(4,92)=22.38]*
Sem-Str	**ST** [F(1,23)=36.58]* **ST x AP** [F(4,92)=3.42]*	**ST** [F(1,23)=41.82]* **ST x AP** [F(2,46)=3.38]* **ST x H** [F(1,23)=12.11]*	**ST** [F(1,23)=37.66]* **ST x H** [F(1,23)=4.79]*	**ST** [F(1,23)=23.14]* **ST x AP** [F(4,92)=4.03]^
Sem-Scram	**ST** [F(1,23)=47.62]* **ST x AP** [F(4,92)=4.76]*	**ST** [F(1,23)=48.33]* **ST x AP** [F(2,46)=2.92]^	**ST** [F(1,23)=56.58]* **ST x AP** [F(3,69)=4.08]*	**ST** [F(1,23)=45.61]* **ST x AP** [F(4,92)=3.48]* **ST x H** [F(1,23)=3.36]^
Str-Scram	**ST x AP** [F(4,92)=3.69]*	**ST** [F(1,23)=3.57]^ **ST x AP** [F(2,46)=3.86]*	**ST** [F(1,23)=5.47]* **ST x AP** [F(3,69)=4.99]*	**ST** [F(1,23)=4.24]^ **ST x AP** [F(4,92)=4.29]*

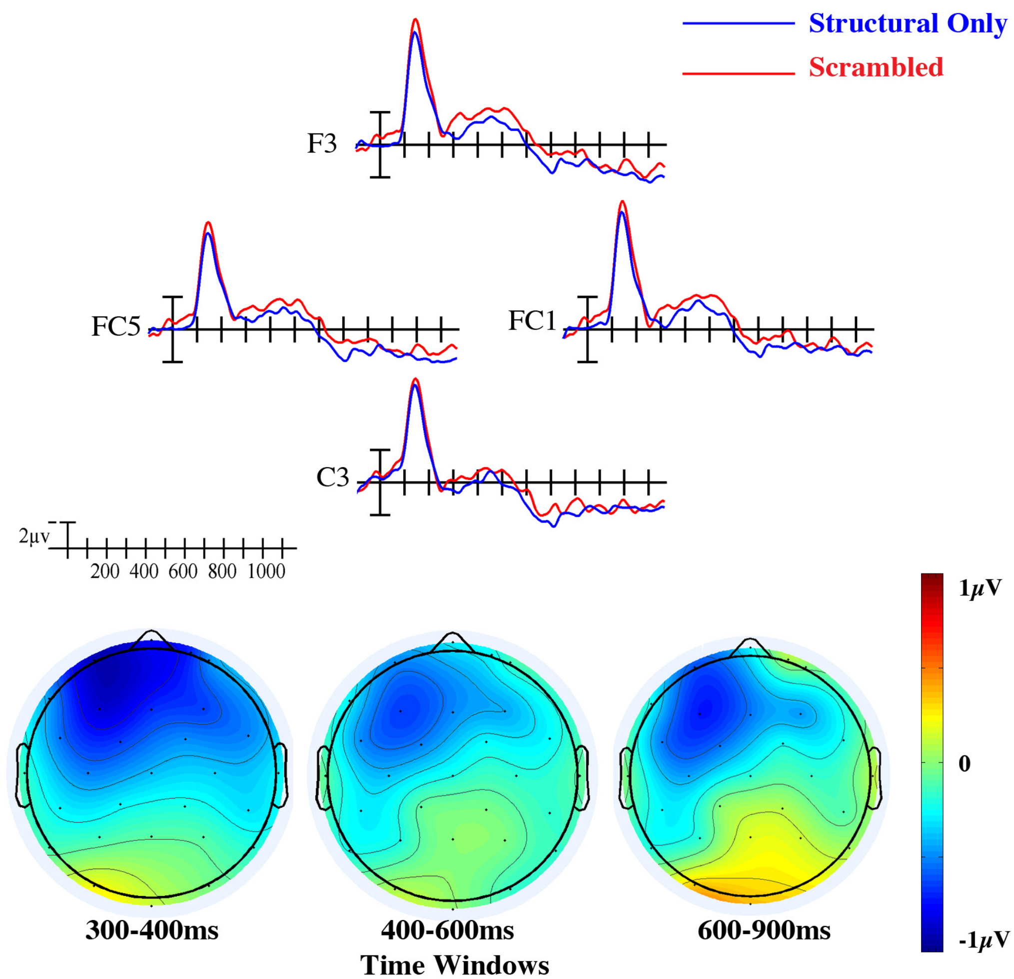
Figure 9. Waveforms and voltage maps for event related potentials to all sequence types



400-600: N400

Starting at around 400ms and lasting until roughly 600ms, there appeared to be a negative deflection peaking at 450ms, consistent with an N400. Repeated-measures ANOVAs once again revealed significant main effects of Sequence Type (all $F>48.5$, all $p<.05$) and/or interactions between AP distribution and Sequence Type at all columns (all $F>10.25$, all $p<.05$). Just as for the N300, the N400 increased from the Normal to Semantic Only and further in Structural Only sequences; these effects were widely distributed and again maximal at anterior electrode sites and, for the Semantic Only versus Structural Only also at right hemisphere sites see Figure 9, Table 5. Once again, the N400 was slightly greater in the Scrambled than Structural Only sequences, but only at localized sites (a near-significant effect of Sequence Type was found in the lateral column, an interaction between Sequence Type and AP Distribution at both medial and lateral columns, and a near-significant interaction between Sequence Type and Hemisphere at the medial column, but no effects at midline or peripheral columns) Follow-up of these interactions revealed effects only at left anterior sites in the medial and lateral columns (F3, FC5, FC1, C3). These contrasts are depicted in Figure 10.

Figure 10. Contrasts and voltage maps for the Scrambled versus Structural Only sequence types



(Note that the scale of voltages has been reduced by half to 1µV)

Table 5. Statistical analysis of waveforms in the 400-600ms time window
ST= Sequence Type, H= Hemisphere, AP= AP Distribution, *= p < .05

	Midline	*Medial*	*Lateral*	*Peripheral*
Norm-Sem	**ST** [F(1,23)=24.45]* **ST x AP** [F(4,92)=17.49]*	**ST** [F(1,23)=46.90]* **ST x AP** [F(2,46)=6.39]* **ST x AP x H** [F(2,46)=2.60]^	**ST** [F(1,23)=36.66]* **ST x AP** [F(3,69)=8.13]*	**ST** [F(1,23)=16.45]* **ST x AP** [F(4,92)=11.80]*
Norm-Str	**ST** [F(1,23)=127.69]* **ST x AP** [F(4,92)=40.53]*	**ST** [F(1,23)=176.38]* **ST x AP** [F(2,46)=9.58]*	**ST** [F(1,23)=156.89]* **ST x AP** [F(3,69)=12.86]*	**ST** [F(1,23)=87.48]* **ST x AP** [F(4,92)=25.50]*
Norm-Scram	**ST** [F(1,23)=130.13]* **ST x AP** [F(4,92)=44.99]*	**ST** [F(1,23)=172.91]* **ST x AP** [F(2,46)=21.79]*	**ST** [F(1,23)=160.31]* **ST x AP** [F(3,69)=28.2]*	**ST** [F(1,23)=96.95]* **ST x AP** [F(4,92)=27.08]*
Sem-Str	**ST** [F(1,23)=45.78]* **ST x AP** [F(4,92)=7.39]*	**ST** [F(1,23)=65.43]* **ST x AP** [F(2,46)=3.23]* **ST x H** [F(1,23)=4.43]* **ST x AP x H** [F(2,46)=3.42]*	**ST** [F(1,23)=65.53]* **ST x H** [F(3,69)=2.61]^	**ST** [F(1,23)=32.67]* **ST x AP** [F(4,92)=3.51]*
Sem-Scram	**ST** [F(1,23)=57.60]* **ST x AP** [F(4,92)=15.15]*	**ST** [F(1,23)=61.14]* **ST x AP** [F(2,46)=12.24]^	**ST** [F(1,23)=62.73]* **ST x AP** [F(3,69)=12.52]*	**ST** [F(1,23)=37.41]* **ST x AP** [F(4,92)=6.89]*
Str-Scram	**n.s.**	**ST x AP** [F(2,46)=3.75]* **ST x H** [F(1,23)=3.58]^	**ST** [F(1,23)=2.97]^	**n.s.**

600-900ms: Late Negativity

The negativity effect continued throughout the 600 to 900ms time window where again ANOVAs revealed significant main effects of Sequence Type (all $F>24.36$, all $p<.05$), and/or interactions between Sequence Type, AP distribution and/or hemisphere (all $F>3.8$, all $p<.05$). Again, the Semantic Only sequences showed a greater negativity than the Normal; this effect was once again widespread and maximal at anterior sites (main effects of Sequence Type and interactions between Sequence Type and AP Distribution at all columns) and appeared to become more left lateralized (interactions between Sequence Type and Hemisphere approaching significance at medial, lateral and peripheral columns), see Table 6, Figure 9. Once again, Structural Only showed a greater amplitude of late negativity than Semantic Only sequences, and again this effect was widespread (main effects of Sequence Type at all columns), but maximal at anterior sites (interactions between Sequence Type and AP Distribution at all columns). Finally, late negativity was slightly greater to panels in the Scrambled than Structural Only sequences, but only at anterior and left lateralized sites (interactions between Sequence Type and AP Distribution at midline, medial and lateral columns, interactions between Sequence Type and Hemisphere at the medial column and a three-way interaction between Sequence Type, Hemisphere and AP Distribution at the peripheral column. Follow-up of these interactions revealed effects only at left anterior medial sites (FC1, C3).

Table 6. Statistical analysis of waveforms in the 600-900ms time window
ST= Sequence Type, H= Hemisphere, AP= AP Distribution, *= p < .05

	Midline	*Medial*	*Lateral*	*Peripheral*
Norm-Sem	**ST** [F(1,23)=11.87]* **ST x AP** [F(4,92)=9.18]*	**ST** [F(1,23)=22.75]* **ST x AP** [F(2,46)=8.78]* **ST x H** [F(1,23)=4.57]* **ST x AP x H** [F(2,46)=2.49]^	**ST** [F(1,23)=13.98]* **ST x AP** [F(3,69)=8.96]* **ST x H** [F(1,23)=3.09]^	**ST x AP** [F(4,92)=5.06]* **ST x H** [F(1,23)=4.40]^
Norm-Str	**ST** [F(1,23)=47.31]* **ST x AP** [F(4,92)=27.28]*	**ST** [F(1,23)=67.86]* **ST x AP** [F(2,46)=14.62]*	**ST** [F(1,23)=63.72]* **ST x AP** [F(3,69)=14.95]* **ST x H** [F(1,23)=3.42]^	**ST** [F(1,23)=31.29]* **ST x AP** [F(4,92)=19.88]* **ST x H** [F(1,23)=3.32]^
Norm-Scram	**ST** [F(1,23)=49.69]* **ST x AP** [F(4,92)=31.47]*	**ST** [F(1,23)=73.17]* **ST x AP** [F(2,46)=27.54]* **ST x H** [F(1,23)=9.55]*	**ST** [F(1,23)=78.14]* **ST x AP** [F(3,69)=31.08]* **ST x H** [F(1,23)=11.52]*	**ST** [F(1,23)=32.57]* **ST x AP** [F(4,92)=21.49]* **ST x H** [F(1,23)=9.80]*
Sem-Str	**ST** [F(1,23)=32.20]* **ST x AP** [F(4,92)=9.43]*	**ST** [F(1,23)=47.38]* **ST x AP** [F(2,46)=4.87]*	**ST** [F(1,23)=42.41]* **ST x AP** [F(3,69)=3.13]*	**ST** [F(1,23)=28.20]* **ST x AP** [F(4,92)=6.90]* **ST x AP x H** [F(4,92)=2.73]*
Sem-Scram	**ST** [F(1,23)=71.72]* **ST x AP** [F(4,92)=27.56]*	**ST** [F(1,23)=79.59]* **ST x AP** [F(2,46)=18.65]*	**ST** [F(1,23)=81.62]* **ST x AP** [F(3,69)=15.59]* **ST x H** [F(1,23)=3.30]^	**ST** [F(1,23)=34.57]* **ST x AP** [F(4,92)=16.57]*
Str-Scram	**ST x AP** [F(4,92)=2.76]*	**ST x AP** [F(2,46)=6.88]* **ST x H** [F(1,23)=5.03]*	**ST** [F(1,23)=3.19]^ **ST x AP** [F(3,69)=3.32]*	**ST x AP x H** [F(4,92)=2.999]*

<u>Effects of Sequence Position</u>

ERPs to panels at each of the sequence positions were averaged at the six select electrode sites which showed the largest N300/N400 and late negativity amplitudes (F3, Fz, F4, FC1, FC2, Cz), and the modulation of these waveforms across ordinal Position was examined for each sequence type using linear regressions (see Table 7).

Table 7. Regressions for amplitudes across sequence position
r= correlation coefficient, b= slope in amplitudes (n=6), *= p<.05

	300-400 ms r, b	400-600 ms r, b	600-900 ms r, b
Normal	.271*, .116	.379*, .162	.206 *, .098
Semantic Only	.073, .032	.060, .029	.121, -.063
Structural Only	.009, .004	.171*, -.092	.308*, -.159
Scrambled	.044, -.023	.174*, -.097	.390*, -.199

There were significant linear trends across ordinal Position in the Normal sequences for all three components, reflecting a decrease in the amplitude of the N300/N400/late negativity across panel position. No other sequence type showed significant effects of ordinal position in the N300 time window, depicted in Figure 11. In the N400 and late negativity time windows, there were no significant effects of Position for the Semantics Only sequences, and panels in both Structural Only and Scrambled sequence types evoked small but significant increases in amplitude with increasing ordinal position, depicted in Figures 12 and 13 (note: as with ERP data, positive is plotted down).

Figure 11. Amplitudes of the N300 across sequence position

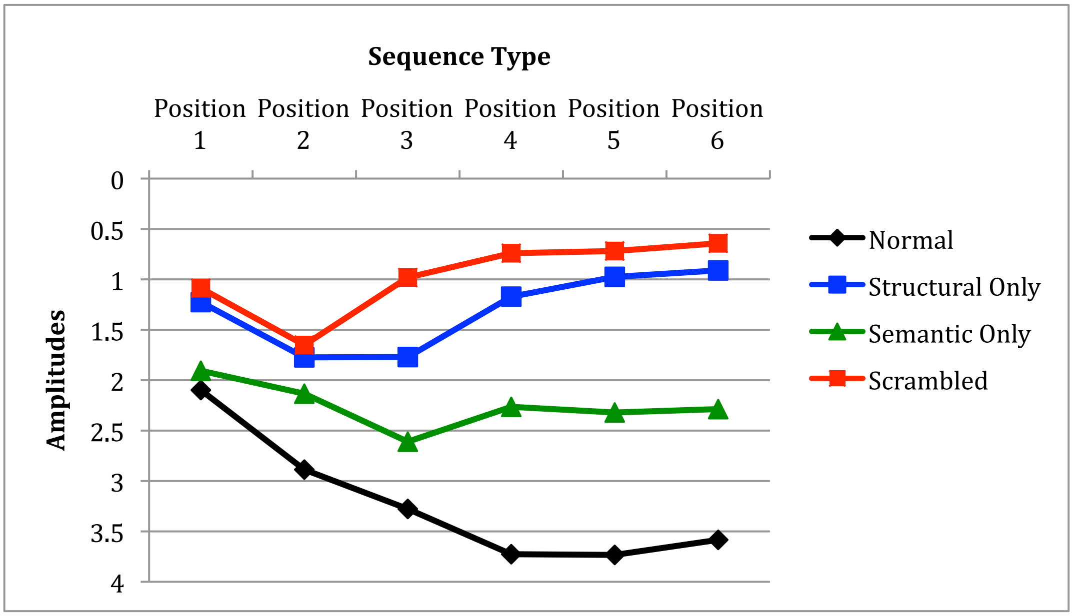


Figure 12. Amplitudes of the N400 across sequence position

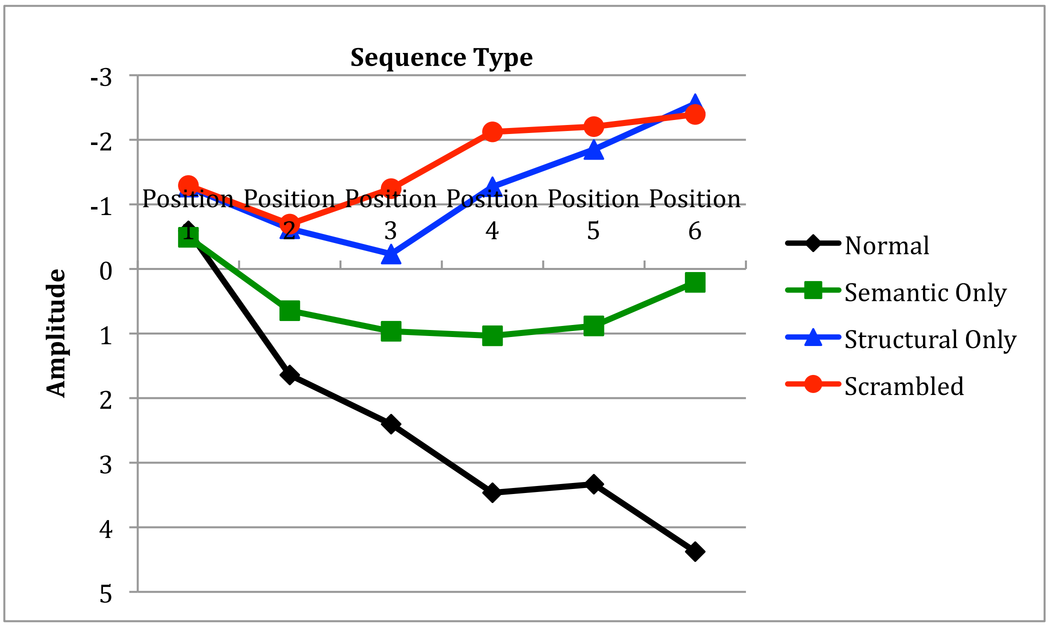
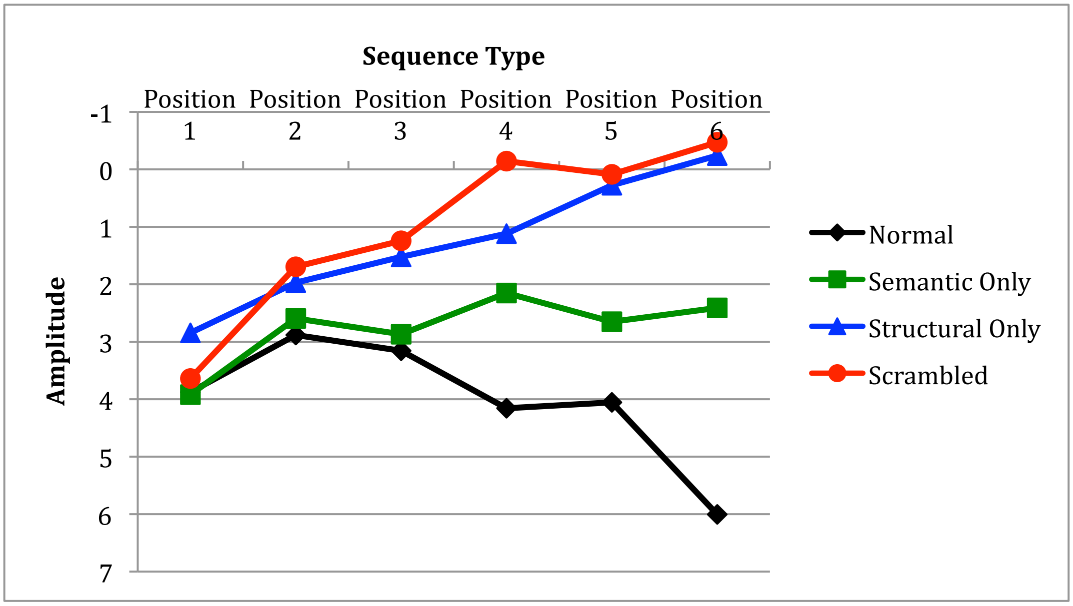


Figure 13. Amplitudes of sustained negativity in the 600-900ms time window across sequence position



Effects of Fluency

For each individual participant, we calculated differences in ERPs between the Normal vs Semantic Only, the Semantic vs. Structural Only and the Structural Only vs. Scrambled contrasts, averaged across the same six select electrode sites (F3, Fz, F4, FC1, FC2, Cz) at the N300, N400 and late negativity time-windows. These differences were correlated with these participants' comic reading fluency, using a Pearson's correlation set to an alpha level of .05.

Significant correlations were observed between Comic Reading Fluency and the magnitude of the N300 and N400 differences between Structural Only and Scrambled sequences (N300 effect: $r(24)=-.467$, $p<.05$; N400 effect: $r(24)=-.424$, $p<.05$). As depicted in Figures 14 and 15, these correlations indicated that higher fluency was associated with a larger negativity effect for this contrast.

Figure 14. Correlation between Structural Only and Scrambled sequence type amplitudes with Comic Reading Fluency for the 300 to 400 ms time window

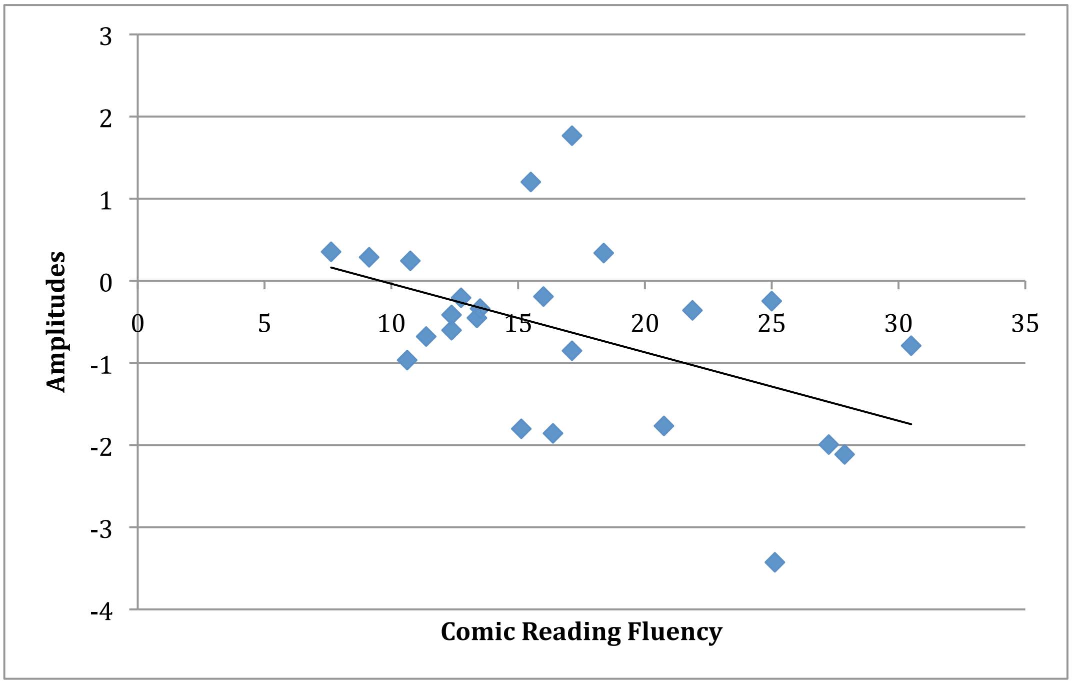
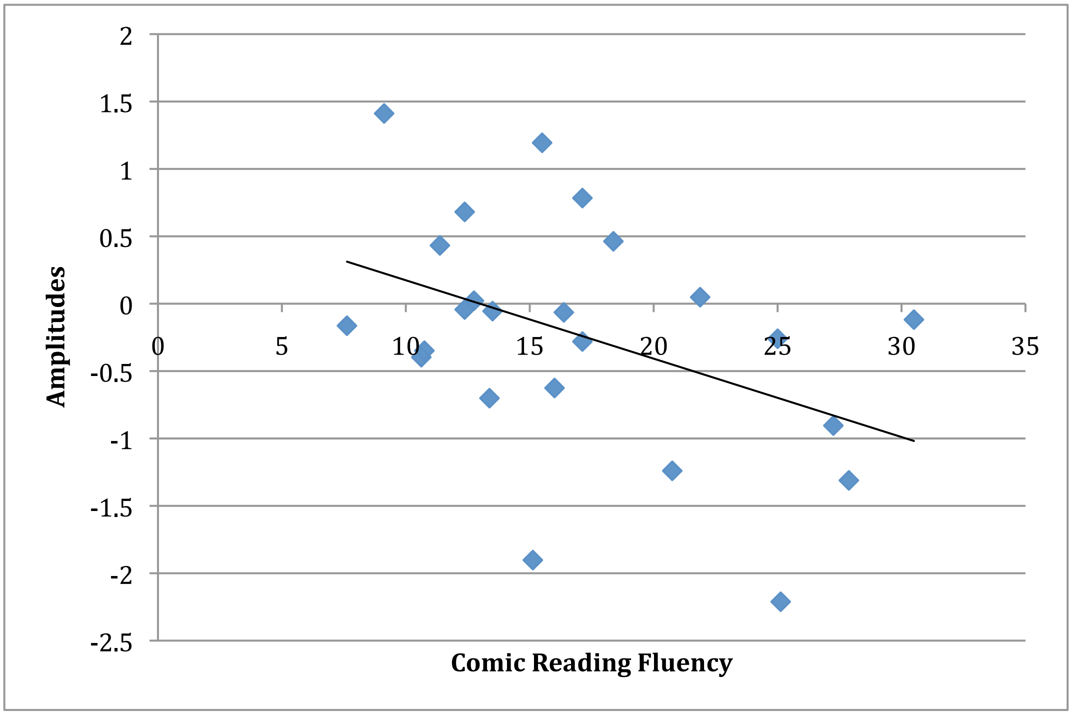
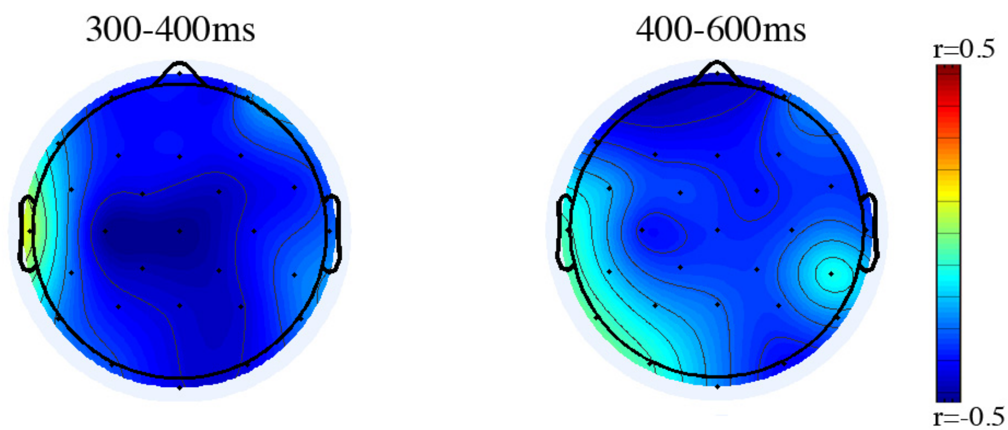


Figure 15. Correlation between Structural Only and Scrambled sequence type amplitudes with Comic Reading Fluency for the 400 to 600 ms time window



No other contrasts yielded significant correlations with fluency. When these correlations (r values) were plotted by electrode across head-maps, the distributions of the highest fluency on the scalp differed between effects, as shown in Figure 16. For the N300, the effects of high fluency were most contrasting at central and posterior sites. In contrast, the highest disparity in fluency for the N400 was found in a leftward frontal distribution.

Figure 16. Correlations (r values) between the difference between Scrambled versus Structural Only and Comic Reading Fluency plotted by electrode site



Experiment 2: Discussion

Experiment 2 examined the neural modulation to individual panels using ERPs as participants read comic strips one panel at a time. Clear differences in a negative waveform were observed, peaking around 500ms, depending on whether panels appeared in Normal, Semantic Only, Syntactic Only or Scrambled sequences. Consistent with the distribution of the N400 found by West and Holcomb (2002), this waveform had an anterior distribution and was preceded by a negative deflection peaking at 300ms, also with an anterior distribution, the N300 (see also Barrett & Rugg, 1990; Ganis, et al., 1996; McPherson & Holcomb, 1999). The amplitude of the N300/N400 complex was largest to panels in sequences without any semantic association (the Scrambled and Structural Only sequences), smallest in the Normal sequences, with the N400 to panels in the Semantics Only sequences falling in between. A differential effect was found for ordinal position on the modulation of the 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 panel within the sequence. In contrast, no decrease in N300/N400 amplitude appeared across ordinal position in the Scrambled and Structural Only sequences. Indeed, these sequence types featured an increasing trend in amplitude of the N300/N400 across panel position for the N400, but not for the N300. These results — both the modulation of the N400 across the Normal, Structural Only and Scrambled sequence types, as well as differences in its modulation across ordinal position — parallel the observations of Van Petten and Kutas (1991) to words within sentences within and without syntactic structure. They extend the findings of West and Holcomb (2002) who showed a larger N400 to incongruous panels falling at the end of a normal sequence panels than to congruous ones. West and Holcomb's (2002) results and this experiment show that structure in the presence, but not in the absence, of semantic association, impacts directly on the semantic processing of incoming visual images.

The fact that the N300/N400 is smaller in the Semantic Only strips (in which panels were linked by semantic association but no narrative structure) than in the Structural Only strips (in which panels were linked by structure but no semantic association) extends the findings of McPherson and Holcomb (1999) and Federmeier and Kutas (2001) who have shown similar facilitory effects of semantic association in semantic priming paradigms with picture pairs. The even smaller amplitude of the N300/N400 to panels in the Normal sequences (which contained both semantic association and structure), relative to those in the Semantic Only strips suggests that, in the presence of semantic association, the structure of the strips as a whole did confer a semantic processing advantage. The combination of semantic association and structure provided a coherent context that facilitated semantic processing of each upcoming panel. In other words, the smaller N400 to panels in the Normal sequences was driven by more than simple semantic association.

Importantly, there was much less separation between the amplitude of the N300/N400 to panels in sequences with structure but no semantic associations, and that to panels in the unstructured Scrambled sequences. Indeed, at most electrode sites, the N300/N400 did not differentiate between these two types of sequences – both without semantic association but with and without structure. These results are analogous to those of Van Petten and Kutas (1991), who reported that 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. Similarly, these findings suggest that, in comprehending sequential images, the N400 is insensitive to narrative structure, in the absence of coherent semantic links between panels.

There was, however, a small difference in the N300/N400 waveform evoked by panels in Scrambled greater than Structural Only sequences at some more localized anterior electrode sites — between 300-400ms at central and anterior sites, and between 400-600ms at left anterior sites. Between 300-400ms, the distribution of this difference was similar to that described in the comparisons described above, i.e. central and anterior, consistent with effects to anomalies in other picture studies (Barrett & Rugg, 1990; Ganis, et al., 1996; McPherson & Holcomb, 1999). Between 400-600ms, however, the distribution of the Structural versus Scrambled ERP difference appeared to be somewhat distinct: it was quite anterior and maximal at left sites. Further, the degree of divergence between panels in the Structural Only and Semantic Only sequences appeared to be modulated by a level of fluency. Fluent readers showed a greater separation in N300/N400 differences between the Structural Only strips and Scrambled sequences than for readers with less fluency. Between 300-400ms, the correlation with degree of fluency was maximal at central-anterior sites, but between 400-600ms the correlation was maximal once again at left anterior sites.

Given the left anterior scalp distribution of the ERP effect in the structural versus scrambled contrast, and the correlation with reading fluency at these sites, one possibility is that this ERP modulation does not reflect an N400 effect, but rather a left-anterior negativity (LAN). As discussed above, in studies of language, the LAN appears in the same time window as the N400, but exhibits a left anterior distribution, and is generally associated with syntactic rather than semantic violations. For example, during sentence processing, LAN effects have been shown to syntactic violations in meaningless sentences (Münte, et al., 1997). Moreover, other studies have described effects similar to the LAN outside the domain of language: a comparable effect to the right hemisphere has been found to violations in musical syntax (Maess, Koelsch, Gunter, & Friederici, 2001). If narrative in sequential images serves as a system of structural integration similar to syntax or musical “syntax” in their respective domains, a distinct effect sensitive to such structure would not be out of the realm of possibility. Further experiments are necessary to directly test this hypothesis by examining the effects of introducing pure structural violations as readers comprehend sequential images.

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. Like the normal sentences in Van Petten and Kutas (1991), panels in the Normal sequences showed a decrease in amplitude of both the N300 and

N400 across position. As in normal sentences, this finding suggests full comprehension relies on the buildup of *both* semantics and structure. The buildup created through this combination of structure and meaning facilitates semantic processing of each successive panel. Also, like the findings of Van Petten and Kutas (1991), the amplitude of the N300/N400 did not decrease throughout the Scrambled and Structural Only sequences. Rather, the amplitude of the N400 showed a slight increase across sequence position, with no trend for the N300. Once again, this has two implications. First, although semantic association can facilitate the buildup of coherence across sequential position, it does not occur in the absence of narrative structure. Second, structure, in the absence of semantic association, does little to facilitate semantic processing of successive panels. That is, as in memory tasks (Gernsbacher, 1983, 1985; Gernsbacher, et al., 1990) there does appear to be a buildup of structure across serial position in the online comprehension of a sequence of images, but only with the influence of *both* narrative structure and semantic association.

General Discussion

In this study, two experiments tested the hypothesis that sequential image comprehension involves an interaction between semantics and a narrative grammar. In Experiment 1, reaction times were measured while participants implicitly monitored for target panels in sequences that featured Normal meaningful narratives, a thematic/semantic field without narrative structure, a narrative structure without meaning, and totally scrambled strings of images. In Experiment 2, event-related brain potentials were measured across all panels of these same sequence types in order to more directly examine neurocognitive processing without relying on behavioral performance, thereby overcoming any strategy of expectation or interference from a probe task.

In Experiment 1, reaction times were fastest to panels in Normal sequences and slowest in Scrambled ones, while in Experiment 2 a negative deflection between 200-600ms – the N300/N400 complex – 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 were smaller than those to panels in the Scrambled sequences, but larger than those to panels in the Normal sequences. However, there were differences across Experiment 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 the Structural Only sequences were faster than those in the Scrambled sequences but the same as those in the Semantic Only sequences. In Experiment

2, the N300/N400 amplitude to panels in the Structural Only sequences were nearly the same as those in the Scrambled Sequences (except at some localized sites) and larger than to panels in the Semantic Only sequences. In both Experiments 1 and 2, there was modulation of the behavioral/ERP response across ordinal position which differed depending on the sequence type. In Experiment 1, reaction times decreased less fast along ordinal position in the Normal sequences than in other types, but in Experiment 2 the N300/N400 amplitude decreased more through Normal sequences than in other types. In Experiment 1, the decrease in reaction time across ordinal position in the other sequence types dropped far more, possibly due to an abandonment of comprehension in favor of simply doing the probe task. In Experiment 2, where such strategies were less likely, there was no such decrease and, indeed, in the Structural Only and Scrambled sequence types, there was a trend towards an increase in N300/N400 amplitude across ordinal position.

Taken together, the results of these experiments offer converging evidence that sequential image comprehension involves the union of separate structures of semantics and narrative, comparable to the system found in language. In both experiments, Normal sequences showed a processing advantage over sequences with either semantics and narrative but not both. Moreover, only Normal sequences featured a reduction in the N300/N400 across ordinal position, unlike sequences of other types. Given that panels in Semantic Only sequences had slower reaction times and higher amplitude ERP effects, this advantage cannot be attributed just to semantic association. Combined, these results suggest that sequential image processing builds up structure across a sequence position using a narrative grammar alongside semantic association to guide comprehension.

However, there were interesting differences between the results of the two experiments. In Experiment 1, a behavioral processing advantage appeared for Structural Only sequences relative to Scrambled ones. In Experiment 2, a slight ERP divergence emerged between these two sequence types, but this was localized at only a few sites and was much less marked than the differences in ERPs between the other conditions. This dissociation between behavioral and N400 modulation across these two conditions mirrors a similar dissociation seen in the studies of sentence processing. Marslen-Wilson and Tyler (1975, 1980) showed reaction times increased from normal sentences to those with only syntax and then scrambled sentences. Relative to normal sentences, Van Petten and Kutas (1991) showed a similar N400 effect for sentences with only syntax as to scrambled ones. Similar to Van Petten et al., this evidence is interpreted as showing that, while the combination of structure and semantics can facilitate semantic processing of an upcoming item, as reflected by an attenuated N400, the N400 itself doesn’t directly reflect structural integration costs.

This interpretation is supported by a similar dissociation across the behavioral and ERP experiments for the contrast between the Structural Only and Semantic Only sequences. In Experiment 1, there was no difference in reaction times to panels in these two types of sequences. In Experiment 2, however, the N300/N400 was clearly smaller to panels in the Semantic Only than to the Structural Only sequences. While reaction times were sensitive to structural and semantic constraints independently and in combination, the amplitude of the N300/N400 was sensitive to semantics and structural constraints in combination, but not structural constraints alone.

However, an alternative interpretation might not attribute any sensitivity to structure in the absence of semantic association, but suggest that the ERP manifested for the Structural Only sequence appears as modulation of a distinct more localized component – the LAN between 400-600ms at left anterior sites. This idea is supported by the observation that the modulation of the waveform in this time-window and at these sites was greatest for individuals with higher comic reading fluency. However, such an interpretation remains speculative and requires further experiments to test more definitively.

Nevertheless, fluency appeared to show advantages for both reaction times in Experiment 1 and in neurocognitive differences in Experiment 2. In the first experiment, readers with higher fluency responded with faster reaction times to panels in non-Normal sequence types than less fluent readers. Higher fluency readers also had a greater difference in reaction times between Normal and Scrambled sequences. These results imply that violations of a sequence affect comprehension more for those with higher fluency in the graphic structure. Beyond these results though, Experiment 2 showed a larger separation between the N300/N400 effects for Scrambled and Structural Only sequence types for more fluent readers compared to less fluent readers. This would indicate that fluency in reading comics is more sensitive to a narrative structure, but not semantic association.

These effects of fluency in the graphic form are consistent with other studies that indicate the comprehension of sequential images is modulated by age and expertise. Two common comprehension tasks using sequential images have asked participants to order a series of scrambled comic panels into a comic strip or to infer the contents of a missing panel. Both the ability to accurately reconstruct strips and fill-in-the-blank increased with age from kindergartners through 8^{th} graders (Nakazawa & Nakazawa, 1993). However, when comparing results between children, college students, and older adults, college students showed the highest accuracy for both tasks, attributed to this population having the highest expertise in comic reading (Nakazawa, 2004). Additionally, a case study comparing an expert versus novice reader showed significant differences in eye movements in the reading of comic pages (Nakazawa, 2002). The expert reader showed smoother, directed eye motions and often skipped word balloons with far faster reading times, while the novice reader read slower and focused more on text than the images, moving through pages with far more erratic eye movements. Recall for the content of the narrative was also found to be more accurate in the expert reader than the novice. Given these previous findings, the results from the present studies shed more light on the actual mechanisms involved with fluency in sequential image comprehension.

Collectively, these two experiments point to a system of comprehension guiding sequential images that is analogous to what is involved in processing verbal language. Both sentences and sequential images require the combination of meaning (semantic association) and structure (narrative/syntax) in the buildup of comprehension across a sequence. Furthermore, while it does not serve to facilitate comprehension on its own, the narrative structure does appear to be modulated by a degree of fluency. Through the effects on processing, these experiments provide evidence for a narrative grammar in the visual modality, and open the door to further research studying other complex aspects of this structure within narrative sequences.

Appendix: Additional Examples of Stimuli

Target panels are noted at the top of each set. Stimuli are listed in blocks of:

1. Normal
2. Semantic Only
3. Structural Only
4. Scrambled

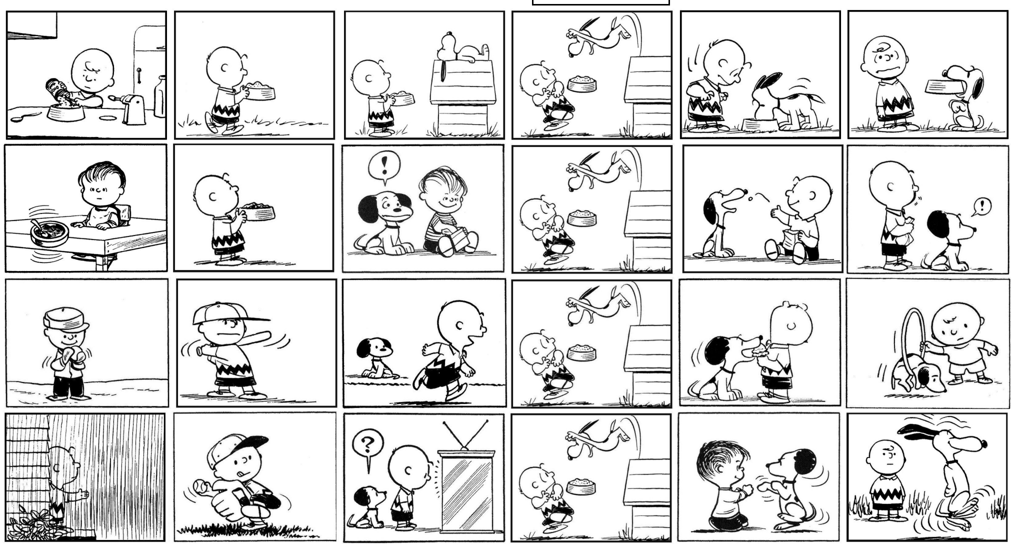
TARGET

1.
2.
3.
4.

TARGET

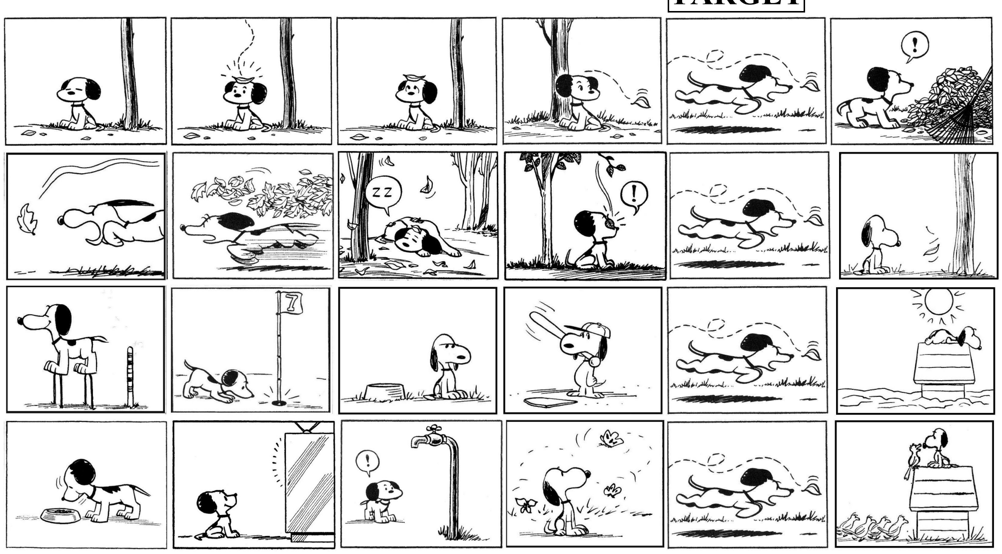
1.
2.
3.
4.

TARGET



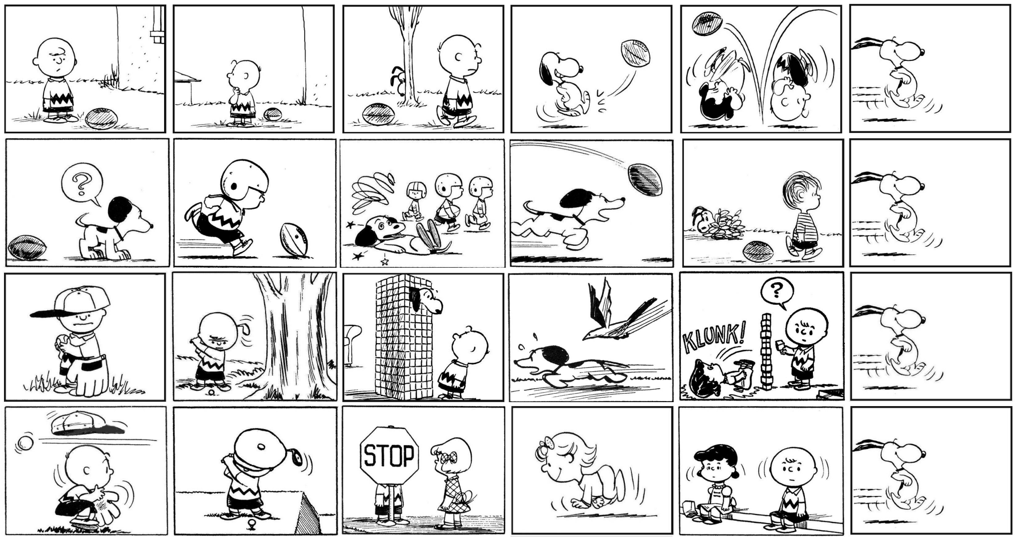
1.
2.
3.
4.

TARGET



1.
2.
3.
4.

TARGET



1.
2.
3.
4.

References

Aristotle (1902). *Poetics*. London: Macmillian and Co. Ltd.

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., & Wilensky, R. (1979). An Evaluation of Story Grammars. *Cognitive Science, 3*, 213-230.

Branigan, E. (1992). *Narrative Comprehension and Film*. London, UK: Routledge.

Bresman, J. (2004). *Test of Sequential “Sense” of Comics.* Unpublished Class Essay: MSTU 5510.008: Social and Communicative Aspects of the Internet and Other ICTs, Teacher's College, Columbia University, New York.

Carroll, J. M., & Bever, T. G. (1976). Segmentation in Cinema Perception. *Science, 191*(4231), 1053-1055.

Cohn, N. (2003). *Early Writings on Visual Language*. Carlsbad, CA: Emaki Productions.

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 prep-a). Foundations for a Visual Language Grammar.

Cohn, N. (In prep-b). You’re a Good Experiment, Charlie Brown: Narrative structure in comic strips.

de Beaugrande, R. (1982). The story of grammars and the grammar of stories. *Journal of Pragmatics, 6*, 383-422.

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, & Cognition, 27*(1), 202-224.

Freytag, G. (1894). *Technique of the Drama*: S.C. Griggs & Company.

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.

Gernsbacher, M. A. (1983). *Memory for Surface Information In Non-Verbal Stories: Parallels and Insights to Language Processes.* Unpublished Doctoral 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 Earlbaum.

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.

Gunter, T. C., Friederici, A. D., & Schriefers, H. (2000). Syntactic gender and semantic expectancy: ERPs reveal early autonomy and late interaction. *Journal of Cognitive Neuroscience, 12*, 556-568.

Hagoort, P. (1993). Impairments of lexical-semantic processing in aphasia: evidence from the processing of lexical ambiguities. *Brain and Language, 45*(2), 189-232.

Hagoort, P., & Brown, C. M. (2000). ERP effects of listening to speech compared to reading: The P600/SPS to syntactic violations in spoken sentences and rapid serial visual presentation. *Neuropsychologia, 38*, 1531-1549.

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. Special issue: Event-related brain potentials in the study of language* (Vol. 8, pp. 439-483). Hove: Lawrence Erlbaum Associates.

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.

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*. Stanford, CA: CSLI Technical Report 85-37.

Kaan, E., Harris, A., Gibson, E., & Holcomb, P. (2000). The p600 as an index of syntactic integration difficulty. *Language and Cognitive Processes, 15*, 159-201.

Kluender, R., & Kutas, M. (1993). Bridging the gap: Evidence from ERPs on the processing of unbound dependencies. *Journal of Cognitive Neuroscience, 5*(2), 196-214.

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. *Cognitive Brain Research, 1146*, 23-49.

Kuperberg, G., Caplan, D., Sitnikova, T., Eddy, M., & Holcomb, P. J. (2006). Neural correlates of processing syntactic, semantic and thematic relationships in sentences. *Language and Cognitive Processes, 21*(5), 489-530.

Kuperberg, G., & Ditman, T. (In Press). Neural Underpinnings of Constructing and Integrating Causal bridging inferences.

Kuperberg, G., Kreher, D. A., Sitnikova, T., Caplan, D., & Holcomb, P. J. (2007). The role of animacy and thematic relationships in processing active English sentences: Evidence from event-related potentials. . *Brain and Language, 100*, 223-238.

Kuperberg, G., Sitnikova, T., Caplan, D., & Holcomb, P. (2003). Electrophysiological distinctions in processing conceptual relationships within simple sentences. *Cognitive Brain Research, 17*, 117-129.

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.

Lynch, K., & Hagen, M. A. (1996). Towards a reality-based analysis of the world of comics. Unpublished Article.

Maess, B., Koelsch, S., Gunter, T. C., & Friederici, A. D. (2001). Musical syntax is processed in Broca's area: an MEG study. *Nature Neuroscience, 4*(5), 540-545.

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

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. In L. Polanyi (Ed.), *The Structure of Discourse*. Norwood, NJ: Ablex Publishing Corp.

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

Marslen-Wilson, W., & 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.

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. (2002). Analysis of manga (comic) reading processes: Manga literacy and eye movement during Manga reading. *Manga Studies, 5*, 39-49.

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

Nakazawa, J. (2005). Development of Manga (Comic Book) Literacy in Children. In D. W. Shwalb, J. Nakazawa & B. J. Shwalb (Eds.), *Applied Developmental Psychology: Theory, Practice, and Research from Japan* (pp. 23-42). Greenwich, Connecticut: Information Age Publishing.

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

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

Osterhout, L., & Mobley, L. A. (1995). Event-related brain potentials elicited by failure to agree. *Journal of Memory and Language, 34*, 739-773.

Robertson, D. A. (2000). *Functional neuroanatomy of narrative comprehension.* Unpublished Doctoral dissertation, University of Wisconsin, Madison, Madison, WI.

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* (Vol. 211-236). New York, NY: Academic Press.

Rumelhart, D. E. (1980). On Evaluating Story Grammars. *Cognitive Science, 4*, 313-316.

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.

Schulz, C. M. (2004a). *The Complete Peanuts: 1950-1952*. Seattle, WA: Fantagraphics Books.

Schulz, C. M. (2004b). *The Complete Peanuts: 1953-1954*. Seattle, WA: Fantagraphics Books.

Schulz, C. M. (2004c). *The Complete Peanuts: 1955-1956*. Seattle, WA: Fantagraphics Books.

Schulz, C. M. (2005). *The Complete Peanuts: 1957-1958*. Seattle, WA: Fantagraphics Books.

Schulz, C. M. (2006a). *The Complete Peanuts: 1959-1960*. Seattle, WA: Fantagraphics Books.

Schulz, C. M. (2006b). *The Complete Peanuts: 1961-1962*. Seattle, WA: Fantagraphics Books.

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., 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.

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.

Stainbrook, E. J. (2003). *Reading Comics: A Theoretical Analysis of Textuality and Discourse in the Comics Medium.* Unpublished Dissertation, Indiana University of Pennsylvania.

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*. Norwood, NJ: Ablex.

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

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

Trabasso, T., & Nickels, M. (1992). The Development of Goal Plans of Action in the Narration of a Picture Story. *Discourse Processes, 15*, 249-275.

van Burkum, 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 de Meerendonk, N., Kolk, H. H. J., Vissers, C. T. W. M., & Chwilla, D. (2010). Monitoring in language perception: Mild and strong conflicts elicit different ERP patterns. *Journal of Cognitive Neuroscience, 22*(1), 67-82.

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.

Weber, H. J. (1989, July 16-21 1984). *Elements of Text-Based and Image-Based Connectedness in Comic Stories, and Some Analogies to Cinema and Written Text.* Paper presented at the Text and Connectedness: Proceedings of the Conference on Connexity and Coherence, Urbino, Italy.

West, W. C. (1998). *Common versus multiple semantic systems: An electrophysiological examination of the comprehension of verbal and picture stories.* Unpublished Doctoral Dissertation, Tufts University, Medford, MA.

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, & Cognition, 33*(1), 55-89.