Comics on the Brain: Structure and Meaning in Sequential Image Comprehension

Neil Cohn (neil.cohn@tufts.edu), Martin Paczynski, Phil Holcomb
Department of Psychology, 490 Boston Ave
Medford, MA 02155 USA

Ray Jackendoff
Center for Cognitive Studies, Miner Hall
Medford, MA 02155 USA

Gina Kuperberg
Department of Psychiatry and Athinoula A. Martinos Center for Biomedical Imaging
Massachusetts General Hospital, Bldg 149, 13th Street
Charlestown, MA 02129 USA

Abstract

Just as syntax differentiates coherent sentences from scrambled word strings, sequential images must also use a cognitive system to distinguish coherent narratives from random strings of images. We conducted experiments analogous to two classic psycholinguistic studies to examine structure and semantics in sequential images. We compared Normal comic strips with both structure and meaning to sequences with Semantics Only, Structure Only, or randomly Scrambled panels. Experiment 1 used a target-monitoring paradigm, and found that RTs were slowest to target panels in Scrambled sequences, intermediate in Structural Only and Semantic Only sequences, and fastest in Normal sequences. Experiment 2 measured ERPs to the same strips. The largest N400 appeared in Scrambled and Structural Only sequences, intermediate in Semantic Only sequences and smallest in Normal sequences. Together, these findings suggest that sequential image comprehension is guided by an interaction between a structure and meaning, broadly analogous to syntax and semantics in language.

Keywords: Narrative; Comics; Visual Language; Discourse; ERPs; N400

Introduction

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

Background

Research on language comprehension has long drawn a distinction between meaning and structure. Most theoretical work pertaining to this separation has been at the level of the single sentence (Chomsky, 1965). However, this general divide is also relevant our understanding of how we make sense of text and discourse, beyond the single sentence.

Much work on text and discourse has emphasized the local relationships which draw together related concepts into common semantic fields (Halliday & Hasan, 1985; van Dijk, 1977) or through the establishment of coherence relationships across individual sentences (Halliday & Hasan, 1976; Zwaan & Radvansky, 1998). Theories of narrative with a more global scope have also been articulated in “story grammars” (e.g. Mandler & Johnson, 1977; Rumelhart, 1975; Thorndyke, 1977), which base individual narrative categories around characters’ goals. These categories were organized within a formalistic generative grammar which imposed a global structure for understanding how characters navigated through events.

Similar to verbal discourse, most theories describing sequential image structure have also focused on linear semantic relationships between individual images (McCloud, 1993). Recently though, Cohn (2003, In prep) has proposed a theoretical model for the structure of graphic narratives. Like story grammars’ treatment of sentences, this model describes panels as playing narrative roles in relation to a global sequence of images (although there are several important differences to story grammars and indeed, the model could potentially extend to describe the structure of verbal discourse, as an alternative approach).

Figure 1 illustrates the narrative structure for a 6-panel *Peanuts* comic strip. This sequence shows a baseball game in which Lucy hits the ball so that Charlie Brown can run home and score while escaping getting tagged out by Schroeder. It begins with a clause of two panels in which the Initial shows the start of Lucy's actions (tossing a ball) that she then hits in the Peak. The second clause progresses from the 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 climaxes as the event of catching the ball is interrupted with Charlie sliding into the base. Finally, the last panel features the resolution of this interaction (Release). This first clause (Lucy hitting the ball) facilitates the second (Charlie scoring) and thus becomes an Initial at a higher level of processing, motivated the Peak inside it. Set up by the Initial of the first clause, the second clause itself is a Peak, again motivated by its Peak. Thus, the narrative structure operates on both the panel level and the level of whole constituencies.

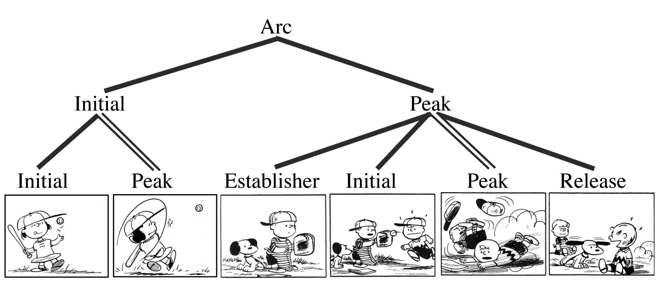

Figure 1: Narrative Structure in a Comic Strip.

While several experiments have looked at the processing of verbal discourse, fundamental questions remain about the online comprehension of sequential images: Is it possible to detect narrative structure during online image-by-image comprehension of sequential images? Can narrative structure be detected independently of semantic associations?

We conducted two experiments to examine the effects of narrative structure in processing sequential images.

Experiment 1: Target Monitoring

In a classic study of sentence comprehension, Marslen-Wilson and Tyler (1975) 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 data suggested that syntactic structure offered a processing advantage beyond lexico-semantics.

Experiment 1 used an analogous "panel-monitoring" paradigm that measured reaction times as readers monitored for target panels in each of four types of comic panel sequences that independently manipulated narrative structure and semantic association.

Methods

Participants

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. Based on their answers to a pre-experiment questionnaire asking about their comic reading and drawing experience as an adult and a child, a measure of participants' comic reading fluency was calculated.

Stimuli

Four types of novel 6-frame long comic strips were created (160 sets) using individual panels culled from several volumes of Charles Schulz's *Peanuts* (see Figure 2): 1) Normal sequences with both narrative structure and semantic relationships between panels, 2) Semantic Only sequences with semantically related panels but no narrative structure (such as panels sharing semantic associations, like baseball, snow, flying kites, etc.), 3) Structural Only sequences featuring a narrative structure but no semantic relationships between panels (maintaining the same narrative categories as in the normal sequences), 4) Scrambled sequences of randomly-ordered panels. Structure in the strips were formalized using Cohn's (2007, In prep) model of visual narrative. The target panels appeared from the second to sixth panel positions, with equal numbers of targets at each position.

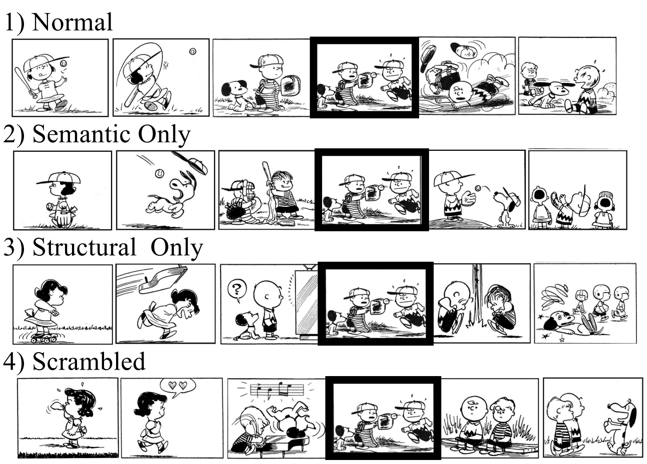

Figure 2: Example Stimuli.

Procedure

At a computer screen, participants monitored target panels as each sequence was presented one frame at a time (1500msec presentation, 300ms ISI).

Results

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

RTs to monitor target panels were slowest in Scrambled sequences, intermediate in both Structural Only and Semantic Only sequences, and fastest in Normal sequences, as depicted in Figure 3. RTs to panels in the Normal strips were shorter to panels in all other sequence types. RTs to targets in the Semantic Only sequences were shorter than in the Scrambled sequences, and RTs in the Structural Only sequences were also shorter than in the Scrambled sequences. There was, however, no significant difference in RTs to targets in the Structural Only and Semantic Only sequences.

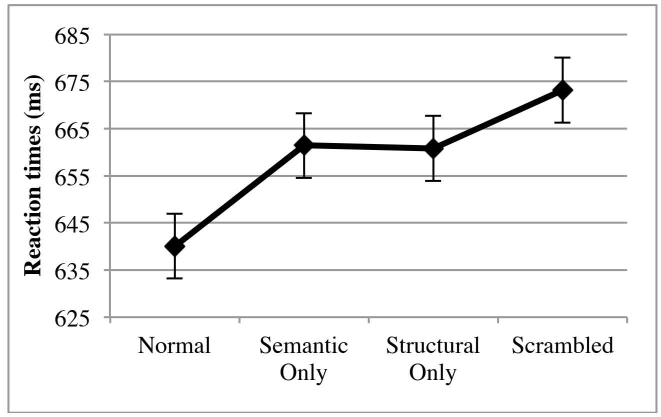

Figure 3: Reaction Times for Sequence Types.

Discussion

As predicted, RTs on target panels in the Normal sequences were fastest while those in Scrambled sequences were the slowest. RTs to target panels in both Semantic Only and Structural Only sequences fell directly in between the RTs in the Normal and Scrambled sequence types and did not significantly differ from one another. These results suggest that both semantics and structure offer advantages to sequential image processing.

Experiment 2: Event-Related Potentials

One of the most studied ERP components in both the verbal and visual domains is the "N400," which has been associated with semantic processing (Kutas & Hillyard, 1980; McPherson & Holcomb, 1999). In an ERP study using similar stimuli to the Marslen-Wilson and Tyler study, Van Petten and Kutas (1991) found that the amplitude of the N400 was larger to targets in scrambled and syntactic-only sentences than in normal sentences. These findings suggested that, in the absence of semantics, syntactic structure did not facilitate semantic processing, as reflected by the N400. Additionally, the amplitude of the N400 was found to decrease across ordinal position in normal sentences.

Syntactic operations in ERPs have been tied to both the "P600" (Osterhout & Holcomb, 1992) and the "left-anterior negativity" (LAN) (Neville, Nicol, Barss, Forster, & Garrett, 1991). However, the P600 only appears in the presence of some semantic information, while the LAN has been observed to violations in the structure of sentences without semantics but still have syntax (e.g. Münte, Matzke, & Johannes, 1997).

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. The authors observed a larger anteriorly-distributed N400 to sequences with incongruous (versus congruous) endings, indicating that semantic processing of the final image was facilitated when its preceding visual narrative was semantically consistent.

Analogous to the study by Van Petten and Kutas, Experiment 2 examined the ERPs to the same set of stimuli from Experiment 1.

Methods

Participants

24 Tufts University undergraduates (12 female, 12 male, mean age = 19.4) participated in the study for compensation. All participants completed the comic fluency questionnaire (described under Experiment 1).

Stimuli

The same stimuli were used in Experiment 2 as in Experiment 1.

Procedure

Participants viewed each sequence one panel at a time on a computer screen while ERPs were measured to all panels. After each sequence, participants judged whether or not the sequence made sense.

Results

Repeated-measures ANOVAs 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 $p<.05$). The N400 was largest to panels in both Scrambled and Structural Only sequences, intermediate in Semantic Only sequences and smallest in Normal sequences. Consistent with previous studies of image processing (West & Holcomb, 2002), this N400 effect had an anterior scalp distribution.

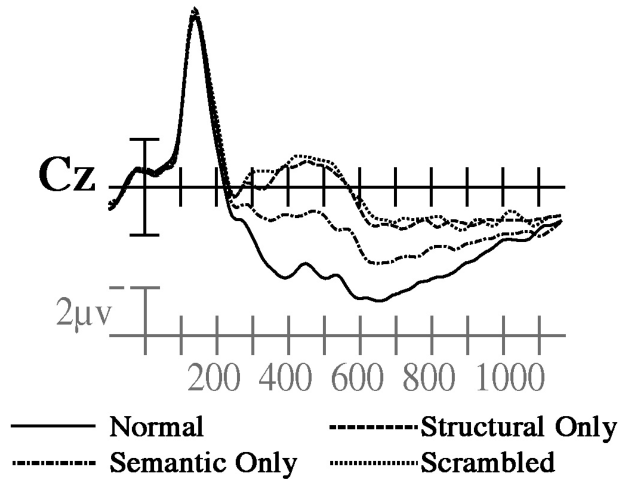

Figure 4: ERPs to different Sequence Types.

While large contrasts appeared for other comparisons between waveforms, the differences between the N400 for Structural Only and Scrambled sequences was small and appeared only at left anterior sites in the medial and lateral columns (F3, FC5, FC1, C3), as depicted in Figure 5.

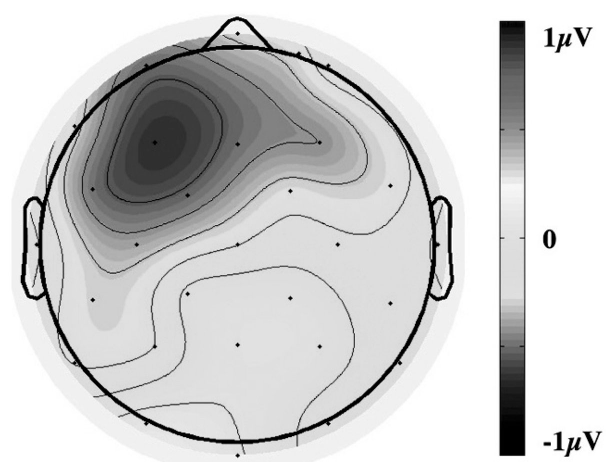

Figure 5: Amplitudes across the scalp for the difference between Structural Only and Scrambled sequence types.

Panels at each sequence position were averaged at the six select electrode sites that showed the largest N400 amplitudes (F3, Fz, F4, FC1, FC2, Cz). A 4 (Sequence Type) x 6 (Position) ANOVA found significant main effects for Sequence Type (all $F>42.0$, all $p<.001$), and a significant interaction between Sequence Type and Position was found for all components (all $F>3.1$, all $p<.001$).

As depicted in Figure 6, significant linear trends appeared across ordinal position in the Normal sequences, reflecting a decrease in the amplitude of the N400 across panel position. No significant effects of Position were seen for the Semantics Only sequences, but panels in both Structural Only and Scrambled sequence types evoked small but significant increases in amplitude with increasing ordinal position.

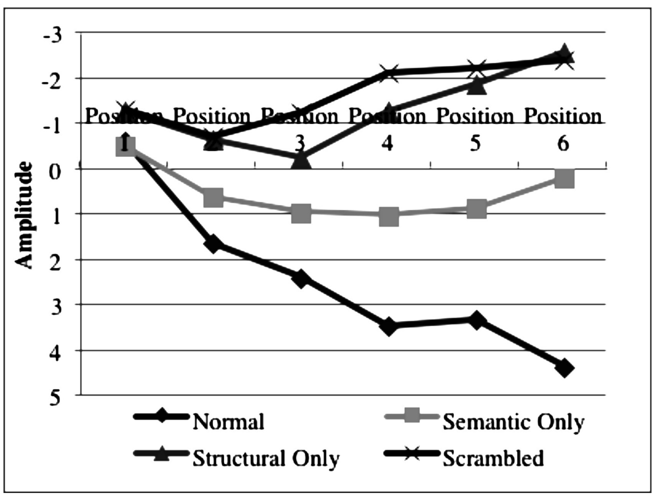

Figure 6: N400 Amplitude for Sequence Types across Sequence Position.

Additionally, significant correlations were observed between participants' scores of Comic Reading Fluency and the magnitude of the differences for the N400 between Structural Only and Scrambled sequences. Greater fluency was associated with a larger difference between these waveforms for this contrast. No other contrasts yielded significant correlations with fluency.

Discussion

The amplitude of the N400 was smallest to panels in the Normal sequences, larger in the Semantics Only sequences, and largest in sequences without any semantic association (the Scrambled and Structural Only sequences). These results suggest that the processing advantage offered by structure, in the absence of semantics, had little impact on reducing the N400 — a waveform sensitive to semantic processing.

No P600 effect appeared to panels in any of the sequence types, consistent with the ERPs reported for sentences in Van Petten and Kutas' (1991). The P600 has primarily appeared to structural anomalies that accompany semantic violations (Münte, et al., 1997). In this experiment, structure was not violated; sequences either were structurally well-formed (albeit possibly without semantics) or fully lacking structure altogether. Thus, given these manipulations, no P600 effect would be expected to these stimuli.

The left-lateralized scalp distribution to the difference in ERPs between Structural Only and Scrambled sequences appeared somewhat distinct from that which has been associated with the N400 in image processing studies. One possibility is that it reflects a partially distinct ERP component such as a left-anterior negativity (LAN), which appears in the same time window but is associated with syntactic processing. One avenue of future research is to determine whether experiments can evoke a LAN more distinctly, as well as elicitation of a P600…

Finally, a decrease in amplitude of the N400 appeared across ordinal position only for Normal sequences. This indicates that the build-up of context created through a combination of narrative structure and semantic theme facilitated the semantic processing of each successive panel.

General Discussion

The results of Experiment 1 suggest that the semantic relationships across individual panels and overall narrative structure across panels each offer advantages to sequential image processing. Experiment 2 further suggests that a combination of narrative structure and semantic relationships facilitate semantic processing of upcoming images, as reflected by the N400. However, structure, in the absence of semantics, had little impact on the N400, suggesting that it acts independently of semantic association.

Taken together, these findings suggest that sequential image comprehension uses a global narrative structure that extends beyond semantic associations between individual frames. We suggest that the comprehension of graphic narrative is guided by an interaction between structure and meaning, analagous to that between syntax and semantics in language.

Acknowledgments

This work was supported by NIMH (R01 MH071635) and NARSAD (with the Sidney Baer Trust), as well as funding from the Tufts Center for Cognitive Studies. Martina Santarsieri, Lauren Rose, Adrienne Ing, and especially Natalie Buzzeo aided greatly in the creation of stimuli and gathering of data.

References

Chomsky, N. (1965). *Aspects of the Theory of Syntax*. Cambridge, MA: MIT Press.

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

Cohn, N. (2007). *Foundations for a Natural Visual Language Grammar*. Paper presented at the Visual and Iconic Languages Conference.

Cohn, N. (In prep). *The Structure of Visual Narrative*.

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

Halliday, M. A. K., & Hasan, R. (1985). *Language, Context, and Text: Aspects of Language in a Social-Semiotic Perspective*. Victoria: Deakin University Press.

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

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

Mandler, J. M., & Johnson, N. S. (1977). Remembrance of Things Parsed: Story Structure and Recall. *Cognitive Psychology, 9*, 111-151.

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

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.

Neville, H. J., Nicol, J. L., Barss, A., Forster, K. I., & Garrett, M. F. (1991). Syntactically Based Sentence Processing Classes: Evidence from Event-Related Brain Potentials. *Journal of Cognitive Neuroscience, 3*(2), 151-165.

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

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.

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

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

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

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

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