Getting a cue before getting a clue:

Event-related potentials to inference in visual narrative comprehension

Neil Cohn and Marta Kutas

Department of Cognitive Science, University of California, San Diego. La Jolla, CA

Short title: Inference in visual narratives

Corresponding author:

Neil Cohn

Department of Cognitive Science

University of California, San Diego

9500 Gilman Dr. Dept. 0526

La Jolla, CA 92093-0526

Phone: (858) 822-0736

Fax: (858) 822-5097

Email: neilcohn@visuallanguagelab.com

Abstract

Inference has long been emphasized in the comprehension of verbal and visual narratives. Here, we measured event-related brain potentials to visual sequences designed to elicit inferential processing. In **Impoverished** sequences, an expressionless "onlooker" watches an undepicted event (e.g., person throws a ball for a dog, then watches the dog chase it) just prior to a surprising finale (e.g., someone else returns the ball), which should lead to an inference (i.e., the different person retrieved the ball). **Implied** sequences alter this narrative structure by adding visual cues to the critical panel such as a surprised facial expression to the onlooker implying they saw an unexpected, albeit undepicted, event. In contrast, **Expected** sequences show a predictable, but then confounded, event (i.e., dog retrieves ball, then different person returns it), and **Explicit** sequences depict the unexpected event (i.e., different person retrieves then returns ball). At the critical penultimate panel, sequences representing depicted events (Explicit, Expected) elicited a larger posterior positivity (P600) than the relatively passive events of an onlooker (Impoverished, Implied), though Implied sequences were slightly more positive than Impoverished sequences. At the subsequent and final panel, a posterior positivity (P600) was greater to images in Impoverished sequences than those in Explicit and Implied sequences, which did not differ. In addition, both sequence types requiring inference (Implied, Impoverished) elicited a larger frontal negativity than those explicitly depicting events (Expected, Explicit). These results show that neural processing differs for visual narratives omitting events versus those depicting events, and that the presence of subtle visual cues can modulate such effects presumably by altering narrative structure.

1. Introduction

Inference has long been emphasized as primary to the comprehension of visual narratives, such as those found in comics or films (Bordwell, 1985, 2007; Branigan, 1992; Chatman, 1978; Eisenstein, 1942; Magliano, Dijkstra, & Zwaan, 1996; McCloud, 1993; Saraceni, 2001; Yus, 2008). While inferences play an important role for making sense of all image-to-image relationships, some sequences make use of storytelling techniques that are aimed expressly at eliciting inferences from a reader. Consider Figure 1a.
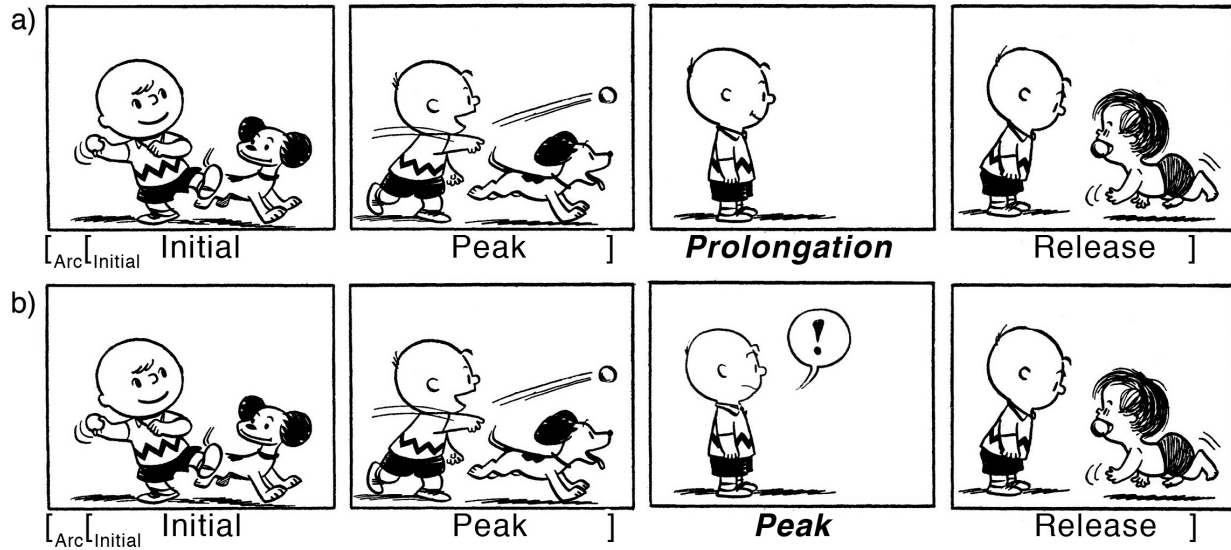

Figure 1. Visual narrative sequences requiring bridging inferences between the penultimate and final panels. a. Sequence with no visual cues in the third panel to suggest an event occurring off panel (Schulz, 2004). b. Sequence with subtle cues in third panel conveying that an event occurs off panel. *Peanuts* is © Peanuts Worldwide LLC.

This early *Peanuts* strip features a structure commonly used by Charles Schulz (Cohn, 2013a), where Charlie Brown reaches back (panel 1) and throws a ball, which Snoopy chases (panel 2) as Charlie Brown watches (panel 3). In the final panel, Charlie Brown (and the reader) is then surprised as it is Linus, not Snoopy, who returns with the ball in his mouth. The humor presumably comes from the reanalysis of prior events in light of the confounded expectations in the final panel (Attardo & Raskin, 1991; Coulson, 2001), but it is facilitated by the ambiguity of the third panel—if Linus had appeared there, the subsequent panel would not be as funny. This sequence uses a pattern common in visual narratives, namely, a penultimate "pause" panel that delays the punchline (Cohn, 2013a). However, because this "pause" only shows Charlie looking, it offers no clue that an unusual event is about to happen. Compare this to Figure 1b, which is identical to Figure 1a with the exception that the third panel not only shows Charlie Brown, but shows him with an exclamation mark in a balloon. This subtle change in the graphic elements that compose the image suggests *some* event outside of view, though it remains unclear exactly what that event is. In both cases, the reader must infer the missing event (i.e., Linus's retrieval) from the depiction in the final panel combined with the prior context.

In discourse terms, the final panel in both of these sequences requires a *bridging inference* (Haviland & Clark, 1974; McNamara & Magliano, 2009), a mental "filling-in" of

information that is necessary to maintain discourse/semantic coherence. If such an inference is not made about what Linus (and Snoopy) were doing "off-panel," the depicted sequence would make less sense overall. Inferential processes like these are a ubiquitous and necessary part of comprehending both verbal and visual discourses (Magliano et al., 1996; McCloud, 1993; McNamara & Magliano, 2009; Zwaan & Radvansky, 1998), filling in the unexpressed information required to build a *situation model* of the narrative (van Dijk & Kintsch, 1983). Generally, successful inference of missing information in a visual narrative increases with age and experience reading comics (for review, see Nakazawa, 2005), as does the capacity to produce visual narratives where the frame acts like a "window" on a scene (Wilson & Wilson, 1987).

Given that both sequences in Figure 1 rely on bridging inferences to be understood (indeed, the same bridging inference), what characterizes the difference between them? Although inferences are *semantic* in nature, they can be influenced by the *narrative structure* of the sequential images. The difference between Figure 1a and 1b can be characterized by ***Visual Narrative Grammar*** (VNG), which describes a system for packaging meaning into a sequence using architectural constraints similar to the way that meaning is packaged by syntax in written or spoken sentences (Cohn, 2013b). Note that the units of visual narratives (images) typically convey more semantic information than the units of sentences (words), giving them an information structure more akin to discourse-level semantics than to sentence-level semantics. That said, our analogy between syntax and visual narrative applies to the structural principles of the systems (categories, hierarchy, etc.) which organize this semantic information into a coherent message and not to their meaning (see Cohn, 2013b for details).

In VNG, a sequence canonically starts with an Establisher, which introduces the characters, often with a passive action. A sequence then gets underway with an Initial, climaxes at a Peak, and ends in a Release, which provides a resolution or coda for the Peak. Because Peaks hold the primary information of the sequence, readers are highly sensitive to their absence (Cohn, 2014). Yet, despite the culmination of a sequence in a Peak, punchlines often occur in Releases.

According to VNG, the first panel in Figure 1 introduces the characters, like an Establisher, but also displays characteristics of an Initial, since Charlie reaches back in a preparatory position presumably to throw the ball for Snoopy to chase. The second panel is a Peak, since Charlie completes the act of throwing the ball. However, this panel also starts a new interaction: Snoopy fetching the ball. Thus, this panel motivates a second role at a higher level—the first and second panels together form a constituent that serves as an Initial for the subsequent sequence. The final panel is a Release, where the sequence resolves humorously.

What then is the role of the third panel in each sequence? In Figure 1a, the third panel appears superficially to be a Prolongation—a spatial or temporal extension following an Initial that often *prolongs* the reader from reaching the Peak. This delay is fairly uninformative about the presumably ongoing undepicted event. Moreover, this sequence appears to have no "culminating" panel—i.e., no Peak—that shows the primary event of the sequence (Linus taking the ball instead of, or from, Snoopy). Although Figure 1b also does not explicitly show this event (Linus getting the ball), both Charlie's facial expression and the exclamation mark in a balloon imply an ongoing event off-image. Because these cues imply a culminating event (!), VNG would identify this panel as the Peak. In sum, according to VNG, Figure 1a and 1b both require a semantic inference but differ in the structural information supporting that inference: the third panel in Figure 1b is more informative than that in Figure 1a.

Despite the minimal differences between Panel 3 in Figure 1a and 1b, there are several reasons to believe that the panel plays a more structural role in 1b (as a Peak) than in 1a (as a Prolongation). For example, one can imagine a sequence adding Panel 3 of Figure 1b *after* the penultimate panel in Figure 1a—making a narratively coherent sequence where the impoverished Prolongation panel delays this slightly more informative Peak panel. Given the option of omitting panels from a narrative sequence, participants rarely delete Initials and Peaks, choosing to delete Establishers, Prolongations, and Releases instead (Cohn, 2014, Experiment 3). Likewise, in a complementary task, when asked to locate missing panels given a 3-panel sequence, participants recognized missing Initials and Peaks more often than other categories, like Prolongations (Cohn, 2014, Experiment 4). Peaks clearly play a central role in visual narratives.

As in this prior work, both Figures 1a and 1b omit content from the sequence. Moreover, they omit the *same* information—in both cases we do not see a panel of Linus taking the ball instead of Snoopy—and call for an inference. What differs is that one image suggests at an event via its content (Figure 1b) while the other does not (Figure 1a). We can ask to what extent, if at all, is the processing of Peaks impacted by this slight difference in two sequences both with impoverished semantic representations?

Previous work has examined the role of semantically impoverished, yet narratively felicitous, panels in sequential images that demand inference. "Action stars", for example, are a conventionalized panel using a star-shaped "flash"—often with an onomatopoeia like *Pow!*—to depict a sudden violent event, especially an impact (Cohn, 2013a, 2013b). These panels suggest an event, but provide no overt referential information about characters or specific actions, and thus demand inference. Participants' self-paced viewing times for action stars that replaced Peak panels were shorter than for empty panels in the same sequential position, and both of these panels were shorter than for fully coherent Peaks, which in turn were shorter than for anomalous Peak panels (Cohn & Wittenberg, 2015). That times are shorter for action stars, even though they contain more visual information than empty panels, suggests that an action star may play a narrative role as a Peak in the sequence. These results suggest that we may find processing differences between panels that play a more structural role as Peaks (like Figure 1b) than those that do not (like Figure 1a). In addition, we expect differences between both of these semantically impoverished images compared with those that fully depict events (such as an image explicitly showing Linus taking the ball).

Though structural differences may be assessed at the critical panel (i.e., the third panel of Figure 1), the full assessment of structure and inference should occur at the subsequent panel. Indeed, we found that panels following either action stars or blank panels were both viewed longer than panels after coherent Peaks with fully depicted scenes, suggesting that there may be a cost for omitting information—i.e., for demanding inference of the undepicted events (see also Baggett, 1975). However, viewing times to panels following action stars and blank panels did not differ from those following anomalous panels, despite their impacting the structure and semantics of the sequence in different ways. Anomalous panels should not just demand inference, but should be fully incongruous to the sequence (Sitnikova, Holcomb, & Kuperberg, 2008b; West & Holcomb, 2002). Perhaps viewing times were not sensitive enough to detect such differences.

Research on sequential images using event-related brain potentials (ERPs)—which are arguably more sensitive to functional processing than viewing times—indicates that differences may be expected to manipulations that impact semantics and structure in different ways. The

processing of meaning has typically been associated with the N400, a negative-going deflection from 300 to 500 milliseconds post-stimulus onset, thought to reflect the activation state of the input in semantic memory (Kutas & Federmeier, 2011). Though first described for the processing of language (Kutas & Hillyard, 1980), N400s also appear to individual (Barrett, Rugg, & Perrett, 1988) and sequential images. For example, larger N400s are evoked by incongruous and/or unexpected images in visual narratives and visual events compared to congruous images (Amoruso et al., 2013; Cohn, Paczynski, Jackendoff, Holcomb, & Kuperberg, 2012; Reid & Striano, 2008; West & Holcomb, 2002).

The manipulation of structure in sequential images elicits different neural responses. One such response, the P600, is a centro-parietal positivity peaking around 600ms, first associated with manipulations of syntactic structure in language (Hagoort, Brown, & Groothusen, 1993; Osterhout & Holcomb, 1992). P600s in sequential images are evinced when the narrative grammar, and not the semantics, of a sequence is manipulated, such as when a sequence requires a reanalysis of the constituent structures (Cohn, Jackendoff, Holcomb, & Kuperberg, 2014). Sometimes P600 amplitude modulations also appear, often in combination with N400 modulations, when internal aspects of images are violated, such as when actions are carried out with the wrong item, such as cutting a cake with an iron (Amoruso et al., 2013; Sitnikova et al., 2008b), when lines depicting motion are drawn in the wrong direction (Cohn & Maher, 2015), or when aspects of a scene are out of place (Võ & Wolfe, 2013). As in studies of language, these positivities seem to appear when the incoming stimulus demands an update or reanalysis of both structure and meaning built from the ongoing context (Kuperberg, 2007, 2013; Sitnikova, Holcomb, & Kuperberg, 2008a), and thereby may be connected to more general processes of mental model updating (Donchin & Coles, 1988).

Contrary to our findings of similar viewing times to panels following action stars, blank panels, and anomalous panels, these ERP results would suggest different mechanisms for the comprehension of critical panels with different contributions from structure and semantics. In Figures 1a and 1b, both sequences would seem to demand the same degree of *semantic* inference—given equivalent ambiguity about undepicted events—but they differ *narratively* in how they provide information about an unseen event to a comprehender. Given that such information involves an updating of a situation model of the broader context, we expect this difference would manifest as a larger P600 to the final panel of Figure 1a than 1b, as the former calls for further consideration of its narrative structure. Similar outcomes also would follow from the observed modulations of P600 amplitudes in language by unexpected, novel, or ambiguous referential information (such as mismatching pronouns or character changes) given the event structure of the prior discourse context (Burkhardt, 2006, 2007; Ferretti, Rohde, Kehler, & Crutchley, 2009; Nieuwland & Van Berkum, 2005; van Berkum, Koornneef, Otten, & Nieuwland, 2007). Posterior positivities similar to P600 effects also appear in good comprehenders for verbal jokes, where appreciating the humor calls for a reanalysis of the context (Coulson & Kutas, 2001). These results are consistent with theories that the P600 indexes the detection of a "prediction error" resulting from the mismatch of the top-down expectations of a broader context, and the bottom-up content of an incoming stimulus (Kuperberg, 2013). This view aligns with the idea that changes in referential and causal cohesion across sentences in a discourse will demand updating of a situation model in memory (Zwaan & Radvansky, 1998). In a similar vein, we hypothesize that the final panel of Figure 1a would elicit a larger P600 than the final panel of Figure 1b, because subtle cues (such as the exclamation

mark and facial expression) about an undepicted event in panel 3 of Figure 1b would lead to a more structurally felicitous sequence than the content of Figure 1a.

Nevertheless, sequences that demand inference, like those in Figures 1a and b, also should be processed differently than sequences that overtly depict events, such as those that replace the third panel with a depiction of the explicit event (Linus taking the ball) or the expected, but then confounded, event (Snoopy retrieving the ball). While inference may motivate a P600, given the cost of updating a situation model, research on verbal discourse has also reported N400s to discourse contexts that demand inference. St. George, Mannes, and Hoffman (1997), for example, found attenuated N400s to words in the final sentences of a discourse preceded by sentences that encouraged inferences versus those that did not. They argued that the causal inference facilitated the comprehension of the subsequent sentences. Others likewise have reported larger N400s to words that afforded bridging inferences across sentences than to those that merely repeated lexical information (Burkhardt, 2006; Yang, Perfetti, & Schmalhofer, 2007) as well as modulations of the N400 effect by the strength of the causal inference demanded by connecting clauses of a discourse (Kuperberg, Paczynski, & Ditman, 2011). These findings suggest we too might obtain larger N400 effects to sequences that demand inference at the final panel, like those in Figure 1, when compared to sequences that depict events overtly.

In addition to inference between sentences in discourse, a later frontal negativity starting around 500ms has been observed in contexts related to the inference of event information that arises from compositional demands within sentences. For example, frontal negativities have appeared when sentence structures imply additive or iterative events (e.g., *For several minutes the cat pounced on the rubber mouse* suggests repeated *pouncing*) though such information is not provided in the sentence outright (Bott, 2010; Paczynski, Jackendoff, & Kuperberg, 2014). Similar negativities have appeared in other contexts requiring complex mappings between semantics and syntax (Baggio, van Lambalgen, & Hagoort, 2008; Wittenberg, Paczynski, Wiese, Jackendoff, & Kuperberg, 2014), and they have been interpreted as reflecting an increase in working memory demands related to generating an appropriate event representation given complex mappings to sentence structures. Given that the sequences in Figure 1 also negotiate a complex relationship between narrative structure and the events being conveyed, we may hypothesize that such sequences would evince a similar anterior negativity. However, no directly comparable precedents have appeared in research of visual narratives, though left lateralized anterior negativities appear to violations of narrative structure in sequential images (Cohn et al., 2014; Cohn et al., 2012).

Given these precedents, we therefore set out to answer our two primary questions: First, can we find evidence of processing differences between visual sequences like those in Figure 1, which omit event information, when compared to overt depictions of events? Second, will processing vary between sequences that both require inference generation, but differ with regard to narrative structure, as in Figure 1a versus 1b? To examine these issues, we presented participants with four types of sequences while measuring event-related potentials: Impoverished sequences (Figure 1a), Implied sequences (Figure 1b), Expected sequences (Snoopy retrieving the ball in the third panel), and Explicit sequences (Linus taking the ball in the third panel).

2. Material and Methods

2.1. Stimuli

We constructed 120 novel visual sequences drawing from a corpus of panels culled from twelve volumes of the *Complete Peanuts* by Charles Schulz (1952-1974), as in several other studies of visual narrative comprehension (Cohn & Maher, 2015; Cohn & Paczynski, 2013; Cohn et al., 2012). The overall "sequence-frame" always featured four panels, with the opening two panels providing the context and initiation of the event, the third and critical panel differing for each of the four experimental conditions, and the critical final panel in the fourth position.

As depicted in Figure 2a, *Impoverished* sequences featured a critical panel with a "onlooker" character with a passive non-emotive expression on their face. In *Implied* sequences, as in 2b the third panel differed from that in the Impoverished condition by virtue of speech balloons containing either a question mark (?) or exclamation mark (!), and/or subtle changes in emotion of the face (exasperated eyes, open mouth) and/or body posture (outstretched arms, "surprised" body positions, etc.). We expected the inference of event information in both the Impoverished and Implied sequences. By contrast, two additional sequence types used critical panels in which the event information was provided and thus should not need to be inferred. In *Expected* sequences, the critical third panel showed an event that was congruous with the prior context, as in Figure 2c where Snoopy retrieves the ball. However, the final panel in these sequences is inconsistent with this information, and thus should appear as anomalous. The third panel of *Explicit* sequences overtly provides the event which presumably was being inferred in the Impoverished or Implied sequences. In Figure 2d, this event is Linus instead of Snoopy going towards the ball, which should be somewhat surprising at the critical panel, but should lead to a congruous final panel.
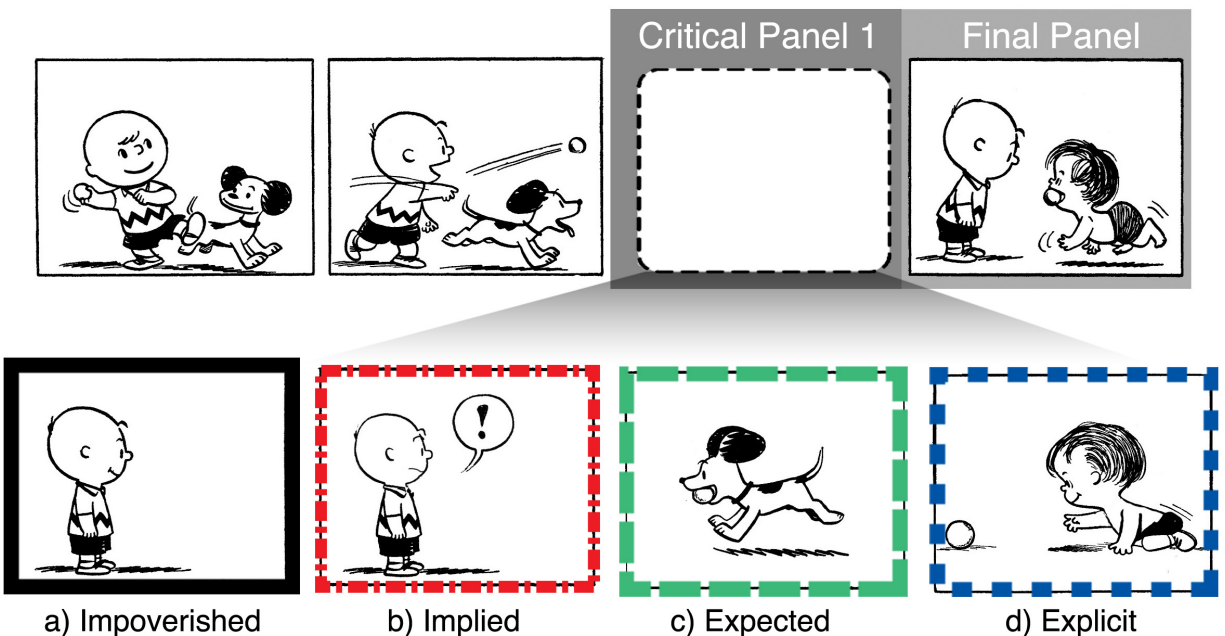

Figure 2. Example experimental stimuli using a single sequence-frame with multiple critical panels.

Note that only 30% (36 of 120) of all final events featured a change in referential information (different characters, changes in expected objects, etc.), as in the example in Figure 1; some showed unexpected causal and/or temporal event outcomes. For example, Charlie Brown hitting a baseball in the opening might reveal that by the end that the ball only landed a short distance away. In terms of the narrative structure, Impoverished and Implied critical panels were always categorized as Prolongations or Peaks respectively in VNG. However, Expected and Explicit critical panels were either Initials or Peaks, depending on whether the final panel was a Peak or a Release, respectively. Additional example stimuli can be found in the supplementary materials posted online: http://www.visuallanguagelab.com/P/gacbgac_supplement.pdf.

Critical panels from the Explicit and Expected conditions were crossed such that all panels appearing in the Explicit condition for one sequence-frame appeared in the Expected condition of a different sequence-frame. Sequence types were distributed into four lists using a Latin Square design such that each sequence-frame appeared only once per list, and participants viewed each critical panel only once. We then randomized these four lists into three different orders each, yielding twelve randomized orders overall. In addition to the experimental sequence types, each list contained 60 filler sequences ranging between 5 to 8 panels in length and ending with sequence-final anomalies. These fillers were included to prevent participants from predicting that unexpected endings would only appear in the fourth ordinal position.

2.2. Participants

Thirty-six comic readers (13 men, 23 women, mean age: 20.3) were recruited from the UC San Diego community. All participants gave their informed written consent according the guidelines of the UCSD Human Research Protections Program. Prior to experimentation, participants answered the Visual Language Fluency Index (VLFI) questionnaire which assessed their expertise in understanding visual narratives based on the frequency with which they read comic books, comic strips, graphic novels, and Japanese comics, as well as drew comics, both currently and while growing up (http://www.visuallanguagelab.com/resources.html). Frequency measurements used a seven-point scale (1=never, 7=always), and the questionnaire also gauged their self-assessed "expertise" at reading and drawing comics along a five-point scale (1=below average, 5=above average). A "VLFI score" was then computed according to the following formula:

$$\left(\begin{array}{c}Mean\ Comic\ Reading\ Freq.\\ x\ Comic\ reading\ expertise\end{array}\right) + \left(\frac{Comic\ Drawing\ Freq.\ x\ Drawing\ Ability}{2}\right)$$

This formula weighted fluency for comic reading comprehension, giving an additional "bonus" for fluency in comic production. Previous research has demonstrated that the fluency score produced by this metric correlates with neurophysiological effects in the online comprehension of visual narratives (Cohn & Maher, 2015; Cohn et al., 2012). Self-defined "comic readers" were recruited to ensure fluency in the visual language of comics (Cohn, 2013a). An idealized average score along this metric would be 12, with low being below 7 and high above 20. Participants' fluency was a high average, with a mean score of 17.43 (SD=7.1, range=7.5 - 37.5).

2.3. Procedure

Participants sat in a comfortable chair facing a computer screen in a copper shielded chamber separate from the experimenter and computers. A trial began with the word READY in the center of the screen. After a button press, a fixation-cross appeared followed by each panel of the sequence, one at a time on the center of the screen for 1350ms. Trials ended with a question mark, at which point participants rated the sequence as either easy or hard to understand by pressing a single button held in either hand. A 300ms ISI separated all screens, preventing the appearance of panels from becoming a "flip-book" animation. An equal number of participants held the "good" and "bad" rating buttons in their right and left hands across all lists. Ten practice sequences preceded the experimental trials to acclimate participants to the procedure and stimuli.

2.4. Data Analysis

EEG was recorded from 26 tin electrodes evenly distributed across the scalp in a quasi-geodesic design (Figure 3). The signal from each electrode site was referenced online to the left mastoid and then re-referenced offline to the average of the right and left mastoids. To monitor horizontal eye movements and blinks, electrodes were placed beneath and next to each eye. Impedances were kept below 5 kΩ for all electrodes. A sampling rate of 250 Hz was used to digitize the EEG, which was bandpass filtered between 0.01 and 100 Hz with James Long amplifiers (www.JamesLong.net).

Our analysis focused on the ERPs recorded to the manipulated critical panel in the third position, the subsequent final panel in the fourth position, and the amplitude differences between the ERPs to these panels in the epochs of 200-400ms, 400-600ms, 600-900ms. We focused our statistical analysis of Sequence Types (Impoverished, Implied, Explicit, Expected) across 16 electrode sites that evenly divided 8 electrodes each into factors of Hemisphere (left, right), Laterality (lateral, medial), and Anterior-Posterior Distribution (prefrontal, frontal, parietal, and occipital), as depicted in Figure 3. Our within-subjects ANOVA looked for main effects and interactions of Sequence Type, Hemisphere, and AP Distribution. Significant interactions were followed by targeted ANOVAs within each of the four quadrants (right/left - anterior/posterior) for pairwise relations between Sequence Types.

Finally, to investigate the influence of comic reading experience on our results, we ran Pearson's correlations with an alpha level set to .05 between VLFI scores and the mean amplitude differences averaged across all electrode sites at the critical panel and final panel.
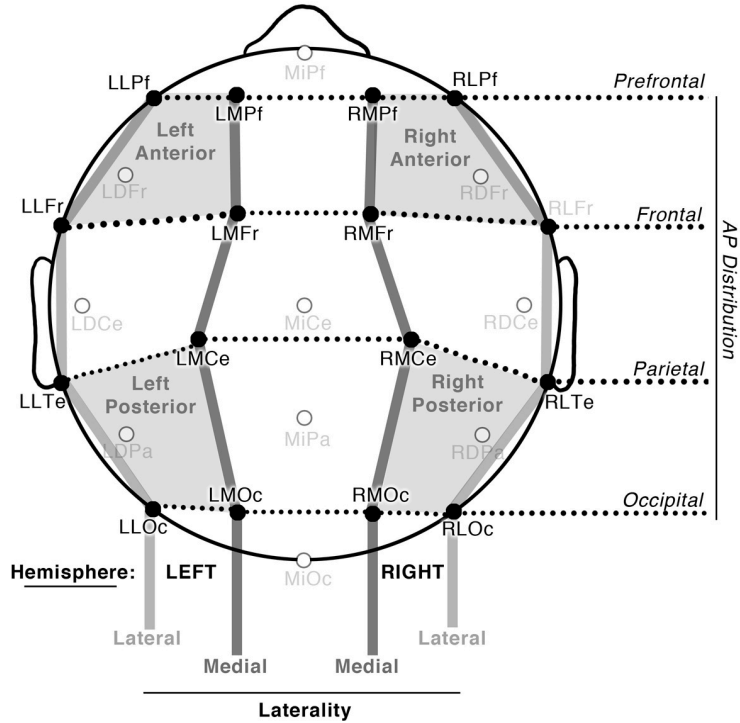
Figure 3. Electrode montage, illustrating 16 electrode sites analyzed across Hemisphere, Laterality, and Anterior-Posterior Distribution, as well as Quadrants used in follow up analyses.

3. Results

3.1. Comprehensibility judgment

Sequences with overt representation of events were rated higher than those demanding inference: Explicit sequences were judged as the most comprehensible (M=.83, SD=.12), followed by Expected sequences (M=.74, SD=.17), Implied sequences (M=.72, SD=.14), and Impoverished sequences (M=.71, SD=.15). Nevertheless, all critical sequences were rated significantly above chance (.5), meaning that participants judged all sequences as fairly comprehensible (all ts > 8.24, all ps < .001). This differed substantially from the low ratings to the filler sequences with sequence-final anomalies (M=.37, SD=.23), which were judged as significantly less comprehensible than chance, t(35)=-3.32, p<.005, and lower than all experimental sequence types (all ts > 9.69, all ps < .001).

Experimental sequences also differed from each other, F(3,105)=17.49, p<.001. Explicit sequences were more comprehensible than all other sequences (all ts < -3.9, all ps < .005). Expected sequences trended as more comprehensible than Impoverished sequences t(35)=-.1.88, p=.068, but no difference was found between Implied and Impoverished or Expected sequences (all ts > -.93, all ps < .346).

3.2. Event-related brain potentials

3.2.1. Critical panel

We found no significant differences between Sequence Types within the 200-400ms epoch at the first critical panel. However, in the 400-600ms and 600-900ms epochs we found

significant main effects of Sequence Type, along with interactions between Sequence Type and AP Distribution (see Table 1 for statistics). Follow up analyses in the 400-600ms epoch found significant main effects of Sequence Type in all posterior regions (Fs > 12.6, all ps < .001). Significant interactions between Sequence Type and AP Distribution appeared in all anterior regions (all Fs > 5.48, all ps < .01). In the 600-900ms epoch, significant main effects and interactions appeared at all regions (all Fs > 4.03, all ps < .01).

Impoverished critical panels were more negative than all other sequence types, followed by Implied, Explicit, and Expected panels (see Figure 4). Impoverished panels were more negative than Implied panels starting in the 400-600ms epoch only in the posterior regions, as evidenced by a main effect of Sequence Type in the right posterior region, F(1,35)=4.4, p<.05, and interactions approaching significance between Sequence Type and AP Distribution in both posterior regions (all Fs > 3.5, all ps < .069). This effect grew at the 600-900ms epoch with significant main effects in the right posterior and anterior regions (all Fs > 8.84, all ps < .01), and a trending main effect in the left posterior region, F(1,35)=3.46, p=.071.

Explicit panels were more positive than Implied panels, with a posterior distribution suggested by significant main effects and interactions in the anterior regions from 600 to 900ms (all Fs > 4.35, all ps < .05). In addition, the greater positivity shown to Expected than Explicit panels in the 400-600ms epoch was suggested by a significant main effect in the left posterior region, F(1,35)=4.9, p<.05.
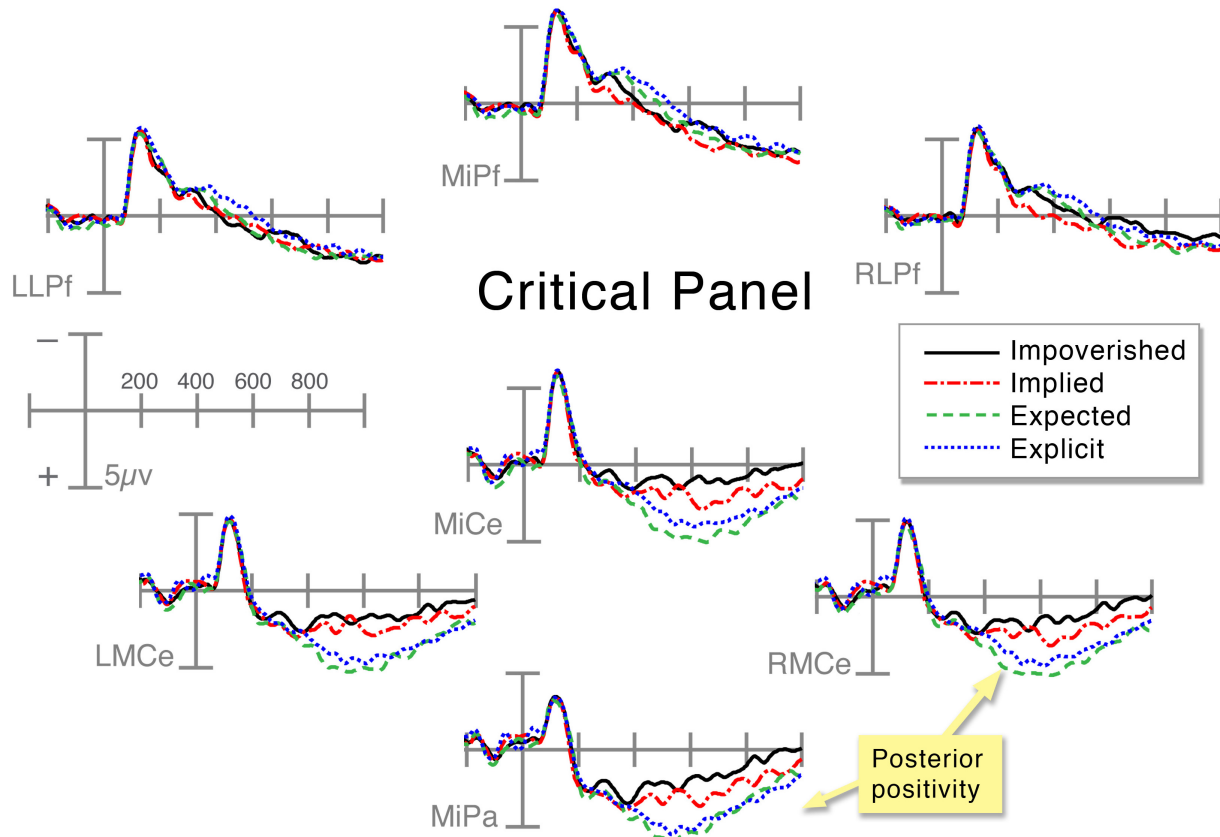

Figure 4. ERPs time-locked to the critical panel across all sequence types at frontal and centro-parietal electrode sites.

Table 1. Results of ANOVAs comparing each sequence type at the Critical Panel, Final Panel, and the difference between these panels. ST = Sequence Type, AP = Anterior-Posterior Distribution, H = Hemisphere. F-values are given.

	Critical Panel		*Final Panel*			*Critical Panel - Final Panel*		
All Sequence Types	400-600	600-900	200-400	400-600	600-900	200-400	400-600	600-900
ST	7.5***	9.9***	2.8*	3.68*	0.896	2.78*	9.6***	6.85***
ST x AP	3.65***	4.21***	6.43***	4.4***	2.68**	3.12**	6.5***	4.6***
ST x H	2.19^	2.14^	0.596	0.238	0.753	0.542	1.37	1.62
ST x AP x H	0.47	0.279	1.64	1.43	0.848	1.01	1.89	0.692
Impoverished - Implied								
ST	1.3	7.01*	5.59*	5.42*	2.39	7.45*	6.2*	10.1***
ST x AP	1.12	0.453	1.99	0.103	0.102	0.977	0.623	0.184
ST x H	3.75^	7.92**	1.47	0.114	0.496	0.321	1.98	4.33*
ST x AP x H	0.681	0.214	2.28^	2.28^	0.859	2.28	2.37^	0.854
Impoverished - Expected								
ST	21.7***	22.6***	6.24*	7.19*	1.14	3.86^	21.8***	12.8***
ST x AP	6.29**	5.98**	4.12**	6.7***	4.97**	2.04	9.5***	5.7**
ST x H	0.233	0.948	0.949	0.038	1.91	0.501	0.158	1.76
ST x AP x H	0.517	0.324	1.63	1.88	2.50^	1.31	1.2	1.08
Impoverished - Explicit								
ST	3.96^	11.5**	1.84	1.21	1.51	0.286	3.98^	8.39**
ST x AP	5.48**	3.74*	6.05**	6.12**	2.95*	2.51^	9.17	6.29**
ST x H	0.126	1.37	0.141	0.477	0.546	0.188	0.66	1.82
ST x AP x H	0.167	0.141	2.34^	1.1	0.981	0.554	0.882	0.782
Implied - Expected								
ST	19.1***	2.33	0.038	0.38	0.225	0.133	16.5***	4.39*
ST x AP	4.13**	5.45**	7.12***	5.5**	3.43*	4.05**	7.04***	5.76**
ST x H	5.3*	1.66	0.038	0.279	0.942	0.923	2.83	0.287
ST x AP x H	0.996	0.027	0.113	0.773	0.451	0.242	1.18	0.669
Implied - Explicit								
ST	1.88	11.5**	1.03	1.91	0.23	3.9^	0.144	0.65
ST x AP	3.88*	7.23***	14.6***	5.08**	2.3^	5.26**	7.93***	6.63***
ST x H	2.47	2.03	0.807	0.09	0.015	0.023	0.57	0.764
ST x AP x H	0.221	0.387	2.59^	0.927	0.465	0.748	0.447	0.157
Expected - Explicit								
ST	4.78*	1.72	1.62	4.2*	0.001	3.26^	11.15**	1.13
ST x AP	0.296	0.376	4.95**	1.61	1.8	2.71*	1.16	0.212
ST x H	0.624	0.077	0.314	0.82	0.421	1.01	1.7	0.083
ST x AP x H	0.421	0.606	1.59	1.91	0.273	0.731	1.41	0.318

All Sequence Types df = 3,93 except Midline ST x R = 12,372
All pairwise df = 1,31 except Midline ST x R = 4,124; ^p<.1, * p <.05, ** p<.01, ***p<.001

3.2.2. Final panel

At the final panel of the sequence, differences between sequence types appeared in the earlier, 200-400ms epoch, where we found main effects of Sequence Type and a Sequence Type by AP Distribution interaction (Table 1). Follow up analyses again found main effects and/or interactions at all regions across the scalp (all Fs > 2.8, all ps < .05). Similar trending or significant main effects and/or interactions appeared to all regions within the 400-600ms epoch (all Fs > 2.3, all ps < .08). However, in the 600-900ms region, only Sequence Type by AP Distribution interactions appeared for the omnibus analysis (see Table 1) and within each region (all Fs > 3.2, all ps < .05).

ERPs at the final panel appeared to have two distinct effects, depicted in Figure 5. First, a posterior positivity appeared between 400 and 900ms, maximal in the 400-600ms epoch. This effect was the most positive to panels in Impoverished sequences, intermediate to panels in Implied and Explicit sequences, and the least positive to those in Expected sequences. Impoverished panels were more positive than Implied panels in the left posterior region in 200-400ms epoch, suggested by a significant Sequence Type by AP Distribution interaction, F(1,35)=6.13, p<.05. In the 400-600ms epoch, main effects of Sequence Type appeared in both posterior regions (all Fs > 4.9, all ps < .05), but this difference disappeared by the 600-900ms epoch (all Fs < 2.8, all ps > .098).

Final panels in Implied and Explicit sequences differed only marginally, with interactions between Sequence Types and AP Distribution appearing between 200 and 400ms F(1,35)=9.7, p<.01, and then again between 600 and 900ms, F(1,35)=5.3, p<.05. These intermediate ERPs to Implied panels were only somewhat more positive than those in Expected sequences, suggested by near significant interactions between Sequence Type and AP Distribution in both posterior regions from 400 to 900ms (all Fs > 3.64, all ps < .065). Final panels in Explicit sequences were larger in amplitude than Expected panels in both posterior regions between 200 and 400ms (interactions in both regions, all Fs > 11.4, all ps < .001), though this interaction reduced to only the left posterior region between 400 and 600ms F(1,35)=13.1, p<.01, and disappeared fully between 600 and 900ms.
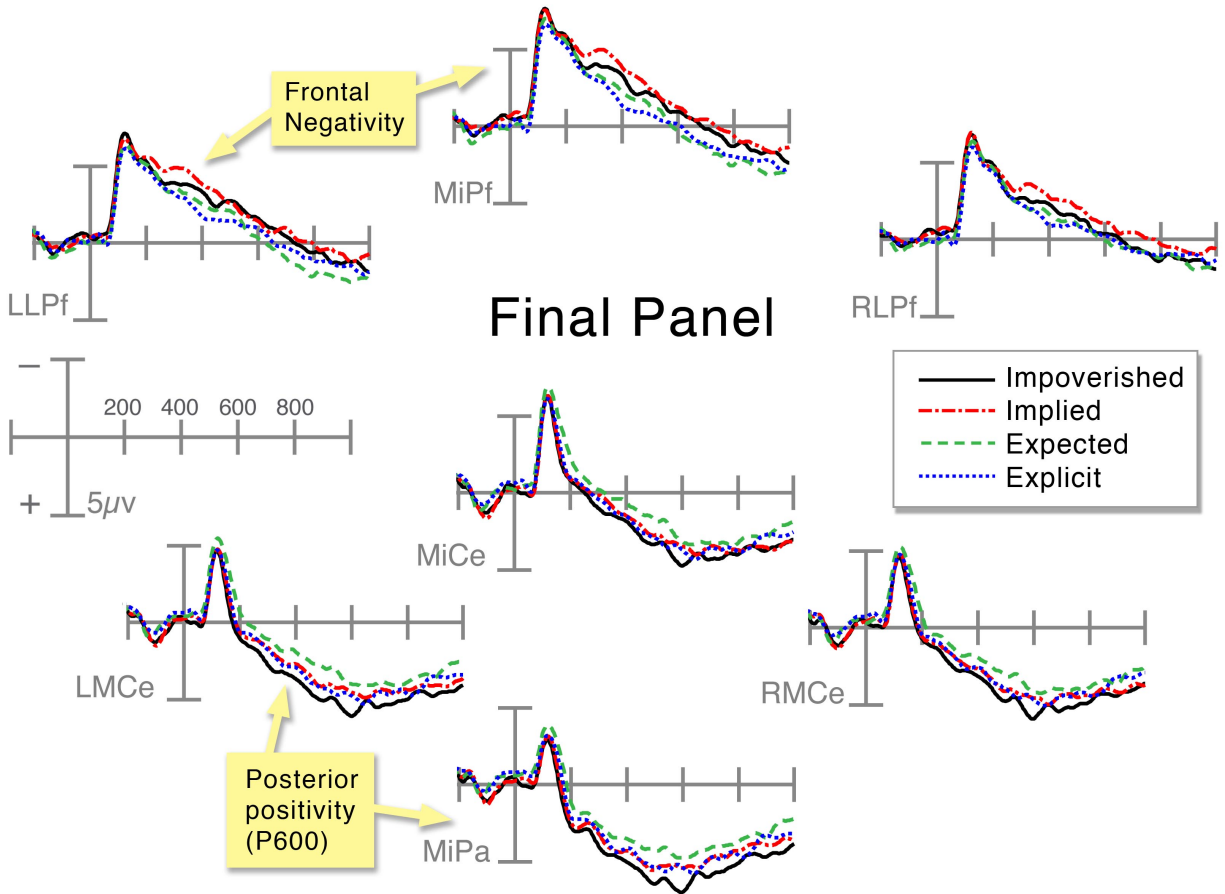
Figure 5. ERPs time-locked to the final panel across all sequence types at frontal and centro-parietal electrode sites.

In addition, a frontal negativity was shown to panels in Implied sequences, then Impoverished sequences, and finally to final panels in Expected and then Explicit sequences. Panels in Implied sequences were more negative than those in Impoverished sequences in an anterior and leftward distribution, with main effects of Sequence Type appearing in anterior regions within the 200-400ms epoch (all Fs > 4.3, all ps < .05), trending in the right anterior region between 400 and 600ms, F(1,35)=3.7, p=.064, but disappearing by 600 to 900ms (all Fs < 1.65, all ps > .206). Final panels in Implied sequences were more negative than those in Expected sequences, as evidenced by interactions between Sequence Type and AP Distribution in the anterior regions in the 200-400ms epoch (all Fs > 4.1, all ps < .05), the 400-600ms epoch (all Fs > 3.9, all ps < .057) and the 600-900ms epoch (all Fs > 8.1, all ps < .01). Finally, panels in Expected sequences were more negative than Explicit sequences, with main effects of Sequence Type in the left anterior region in the 200-400ms, F(1,35)=3.5, p=.069, and 400-600ms epochs, F(1,35)=4.99, p<.05, but no difference appeared in the right anterior region (all Fs < 1.9, all ps > .174).

15

3.2.3. Final panel – Critical panel

Finally, we assessed whether the changes in amplitude differed for the various sequence types between critical and final panels (Figure 6). This analysis was undertaken to examine the impact of the critical manipulation on the overall flow of processing from panel-to-panel. In other words, how did the processing of information in the critical panel affect the processing of the subsequent, final panel?

Similar to the ERPs at the final panel, difference waves varied between sequence types in distinct distributions posteriorly and anteriorly. In the centro-posterior region, differences between critical and final panels in both Impoverished and Implied sequences were more positive than those in Explicit or Expected panels. However, this positivity was larger to the difference between panels in Impoverished sequences than those in Implied sequences, as evidenced by significant main effects of Sequence Type in posterior regions across all time windows (all Fs > 4.25, all ps < .05). Explicit sequences showed relatively no amplitude change between critical and final panels, but this effect was less positive than the difference between Implied panels in posterior regions, as evidenced by main effects in the 400-600ms epoch (all Fs > 4.1, all ps < .05), and interactions in the 600-900ms epoch (all Fs > 15.05, all ps < .001). A widespread centrally distributed negativity appeared between critical panels in Expected sequences, which was more negative than differences in Explicit sequences from 200 to 600ms (interactions in both regions, all Fs > 5.89, all ps < .05).

This order of sequence types was mostly inverted in the anterior regions, where differences between panels in Implied sequences showed a larger negativity than differences between panels in Impoverished sequences, with main effects in the 200-400ms epoch in both anterior regions (all Fs > 4.07, all ps < .052), and main effects in the right anterior region from 400 to 900ms (all Fs > 6.04, all ps < .05). The difference between panels in Impoverished sequences was in turn larger than to those in Expected sequences with significant interactions appearing from 200 to 900ms in all anterior regions (all Fs > 6.8, all ps < .05). Finally, the difference between Expected panels was more negative than between Explicit panels only in the 400-600ms epoch, suggested by main effects in both anterior regions (all Fs > 5.35, all ps < .05).
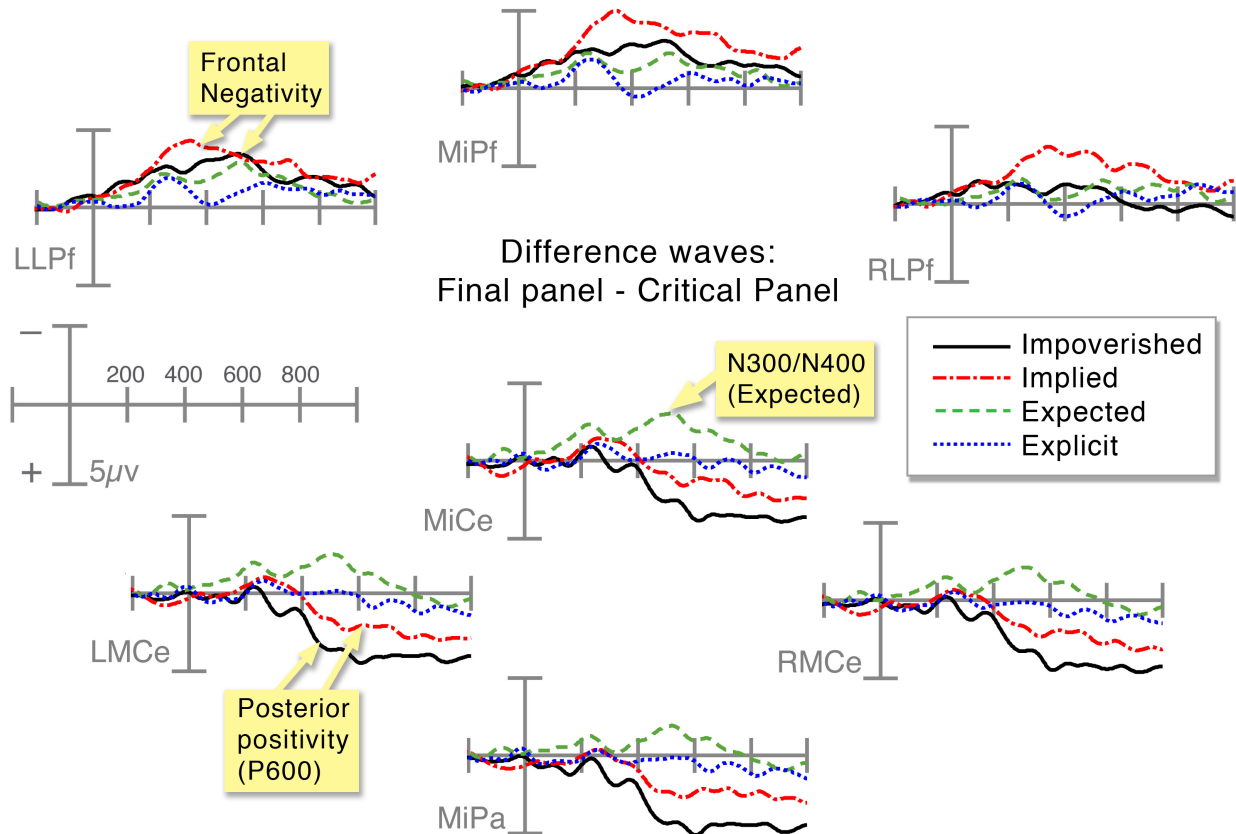
Figure 6. Difference waves between the critical and final panel across all sequence types at frontal and centro-parietal electrode sites.

3.3. Visual Language Fluency

A near significant positive correlation appeared between VLFI scores and the difference between critical Impoverished and Implied panels in the 600-900ms epoch, $r(34)=.322$, $p=.056$. At the final panel, a significant positive correlation appeared between VLFI scores and the difference between Implied and Expected panels in the 200-400ms epoch, $r(34)=.365$, $p<.05$. Finally, a positive correlation appeared between VLFI scores and the amplitude difference between critical panels and final panels for Implied sequences in the 200-400ms epoch, $r(34)=351$, $p<.05$. In all cases, these correlations implied that more experienced readers had a larger effect between sequence types.

4. Discussion

Our experiment asked two questions: 1) Can we find evidence of neural processing differences between visual narrative (comic strip) sequences that depict events overtly versus those where events are omitted (replaced by a passive onlooker)? and 2) Will processing vary between sequence types that are similar in that they both omit event information, but differ in their narrative structure? Overall, we hypothesized that visual sequences with absent critical information, presumed to rely on inference for their understanding, would incur greater costs than those that provide information overtly, whether expected or explicit.

At the critical panel, we observed a late positivity across sequence types between 400 and 900ms over posterior sites. In line with the proposal that late positivities (P600s) reflect the updating or revision of a mental model (Donchin & Coles, 1988)—be it in language (Hoeks & Brouwer, 2014; Kuperberg, 2013), visual events (Amoruso et al., 2013; Sitnikova et al., 2008b), or visual narratives (Cohn et al., 2014; Cohn & Maher, 2015)—larger positivities aligned with greater changes in event structure relative to their coherence with a prior context. The largest positivities occurred to Expected and Explicit critical panels, where event information was directly depicted and thereby progressed the narrative structure. However, images that were less contiguous (Explicit) with their prior context were more positive than those with greater continuity (Expected). Event information changed less between the prior context and Impoverished and Implied critical panels—which evinced the least positivity. These images depicted only a character passively watching off-panel, and thus presumably required less revision of the mental model. However, the visual cues in Implied critical panels suggested an off-panel event, and thereby provided more information to the narrative (a Peak) than Impoverished panels, where little changed in the situation (a Prolongation). This difference was associated with a larger positivity to Implied than Impoverished critical panels.

A similar posterior positivity appeared at the subsequent (and final) image, where Impoverished final panels evoked a greater positivity than Explicit and Implied panels, which did not differ, and were more positive than Expected final panels. These findings again accord with the view that the positivities index a revision of the mental model of the scene (Donchin & Coles, 1988; Kuperberg, 2007, 2013), here as a response to the prior, manipulated critical panel. The largest positivity was seen for Impoverished final panels, where no prior events had either depicted or implied what was to happen, and thus called for the greatest updating of structure when this panel appeared. Though Implied sequences also did not show the prior event, the positivity to their final panels were less positive than those in Impoverished sequences, equivalent to Explicit final panels, which were contiguous with the preceding fully-depicted events. It seems then that the implication of *unseen* event information in Implied sequences and the depiction of a full event have the same consequence in terms of information delivery and consequent mental model updating as indexed by the amplitude of the late positivity to the final panel. Taken at face value this suggests that both the Implied and Explicit critical panels may have played a similar structural role in revealing event information to a comprehender (i.e., as narrative Peaks), despite differing in their semantic content (i.e., explicit vs. inferred events).

Late positivities (P600s) to visual events have typically been elicited by violations of event representations (Amoruso et al., 2013; Cohn & Maher, 2015; Sitnikova et al., 2008b) or of narrative constituent structure (Cohn et al., 2014). However, our sequences do not contain overt incongruities; indeed, participants rated all sequences as fairly comprehensible. To the extent that this late positivity is consistent with those in prior studies, this response would seem to appear to events in various contexts, not just to violations.

In line with this view, recent neuroimaging work comparing coherent comic strips to scrambled strips showed that the temporal-parietal junction and superior temporal sulcus (STS) were activated at every image (Osaka, Yaoi, Minamoto, & Osaka, 2014). The STS has been speculated to be a neural generator for P600 effects in language (Kuperberg et al., 2003). Taken together, these positivities may reflect ongoing integration and/or (re)analysis of prior structure (Hoeks & Brouwer, 2014; Kuperberg, 2013) throughout processing of all sequential images. Such a process is consistent with discourse theories according to which a mental model is updated whenever it encounters changes across various semantic dimensions, including events,

characters, locations, etc. (Magliano & Zacks, 2011; Zwaan & Radvansky, 1998) as well as those according to which the P600 indexes subprocesses involved in building such mental representations (Brouwer, Fitz, & Hoeks, 2012). With regard to our sequence types, this process manifests when the comprehender encounters different event depictions and/or characters at the critical panel and their structural resolution at the final panel.

This interpretation is supported further by the amplitude differences between the critical and final panels. We undertook this analysis to further assess the narrative "flow" from panel-to-panel across our four sequence types: specifically to examine how acquiring different information at one panel affect the processing of the subsequent panel. Impoverished sequences showed the greatest positivity effect between panel positions, suggesting that the absence of information in the critical panel calls for a large revision of the mental model of the scene once further information is made available. The Explicit and Implied sequences did not maintain the same change in amplitude between panel positions, as might be suggested by their relatively equal amplitudes at the final panel alone. Explicit sequences showed relatively little amplitude change between critical and final panels, suggesting that the content of the final panel confirmed the expectations of continuity made at the prior, critical panel. By contrast, Implied sequences showed a distinct positive effect between panels—suggesting that some updating of context was indeed required between these panels, though not as much as when no event information was implied, as in Impoverished sequences.

Insofar as our interpretation of this positivity is correct, the pattern of results are not consistent with a view of P600s as strictly an index of syntactic or combinatorial processing (e.g., Bornkessel-Schlesewsky & Schlesewsky, 2008; Hagoort, Baggio, & Willems, 2009; Kos, Vosse, Van Den Brink, & Hagoort, 2010; Kuperberg, 2007), even if that syntax falls within the visual domain—i.e., a narrative grammar (Cohn et al., 2014). Rather, this pattern of results falls more in line with theories linking the P600 to mechanisms of mental model integration (Brouwer et al., 2012), prediction error (Kuperberg, 2013), and more generally the class of models that allow for domain-general processing (Kuperberg, 2013; Patel, 2003; Sitnikova et al., 2008a).

In addition, at the final panel, a larger frontal negativity appeared between 200 and 600ms to panels in Implied than Impoverished sequences, which in turn was larger than those in Expected and then Explicit sequences. Widespread negativities (N400s) in this time window have previously been seen to unexpected or incongruous semantic events in visual sequences (Amoruso et al., 2013; Cohn et al., 2012; Reid & Striano, 2008; West & Holcomb, 2002). If these negativities are N400s, then it may suggest that Implied and Impoverished sequences are more incongruous or unexpected than Expected or Explicit sequences. Alternatively, to the extent that N400s have been implicated in the inference of unstated causal information in verbal discourse (Burkhardt, 2006; Kuperberg et al., 2011; St. George et al., 1997; Yang et al., 2007), this pattern could index the inference of undepicted events, given subsequent information. This interpretation would be consistent with the bridging inferences required by the Implied and Impoverished but not the Expected or Explicit sequences, and evidence of such inference at the subsequent-to-critical panel (Cohn & Wittenberg, 2015).

Interpreting this negativity as an N400, however, is hard to reconcile with the pattern of amplitude differences observed between the critical and final panels. In particular, Expected final panels should be less semantically congruous with their prior sequence than either Implied or Impoverished final panels. The expected final panel should result in a semantic incongruity, and thus an N400 (West & Holcomb, 2002). Indeed, a widespread central negativity resembling an N400 appeared only between Expected critical and final panels (Figure 6), including an N300

(peaking near 250ms) that typically precedes the N400s to visual images (McPherson & Holcomb, 1999; West & Holcomb, 2002). By contrast, the anterior negativity to Implied and Impoverished sequences patterns differently than those to Expected sequences, and includes no N300.

On an alternative—and we think more likely—interpretation the frontal negativity does not index semantic processes (i.e., an N400), but rather relates to working memory demands created by omitting event information. Frontal negativities have been hypothesized to index working memory processes in sentence (e.g., Kluender & Kutas, 1993) and discourse processing (van Berkum, 2009; van Berkum et al., 2007), such as the costs of computing or maintaining inferred event information within or between sentences (Baggio et al., 2008; Bott, 2010; Coulson & Kutas, 2001; Paczynski et al., 2014; Wittenberg et al., 2014). When related to inference, such sustained frontal negativities have connected to the resolution of "deep ambiguities" regarding the mental model of the discourse (van Berkum, 2009). Frontal negativities in this time window also have been reported to index working memory search processes for referential information (the "Nref"), such as for anaphoric relations within discourse (Hoeks & Brouwer, 2014; van Berkum, 2009). A portion (30%) of our stimuli did indeed include a change in referential information (as in Figure 1), though it is unclear why such a search process would yield different amplitude negativities across the Impoverished, Implied, and Expected sequences, as they all featured them. Since these sequence types did differ in inference generation, we prefer that account. Anterior negativities with a similar leftward distribution also have been tied to the processing of narrative grammar in sequential images (Cohn et al., 2014; Cohn et al., 2012).

Under the interpretation that anterior negativities index memory processes related to inference generation, the absence of explicit event information in both Implied and Impoverished panels would require additional working memory processes to link the final panel to the previous context; this would not be required of Expected or Explicit final panels. Implied sequences would incur even more costs because the cues at the critical panel would activate event information yet maintain ambiguity, thereby requiring reactivation of these features at the final panel. On this view, the comparably attenuated frontal negativity to Impoverished sequences would reflect less working memory demands, because fewer cues would have activated event information, though this absence also would lead to greater revision of the mental model (i.e., P600) when the final panel is encountered. This account is consistent with discourse models that argue for a reactivation stage of information in working memory in maintaining coherence and generating inferences with respect to prior contexts (e.g., Kintsch, 1998; Klin & Myers, 1993; Myers & O'Brien, 1998; Trabasso & Suh, 1993). Whether or not such anterior negativities do indeed index a process related to inference generation and/or to referential information (i.e., an Nref) for sequential images, is an open empirical question.

Finally, positive correlations with Visual Language Fluency Index (VLFI) scores appeared at the critical panel between Impoverished and Implied sequences, at the final panel between Implied and Expected sequences, and between the critical and final panels for Implied sequences. In all cases, larger amplitude effects appeared for participants with higher fluency scores. If these effects are related to inferential processing, they are consistent with findings that experienced comic readers more accurately infer omitted content in comic strips (Nakazawa, 2005) as well as reported correlations between VLFI scores and viewing times to panels requiring inference generation (Cohn & Wittenberg, 2015). Prior work has shown that reading experience factors into discourse processing, including inference generation (e.g., Long, Oppy, & Seely, 1994, 1997). Given proposals for domain-general cognitive mechanisms operating

across verbal and visual modalities (Cohn, 2013a; Cohn et al., 2014; Sitnikova et al., 2008a), these within-domain expertise effects may tap into common underlying processes (e.g., Gernsbacher, Varner, & Faust, 1990). However, this remains an open question, as does the extent to which experience with one modality may aid in the comprehension of another.

Nevertheless, an alternative interpretation may not involve inferential processing. Given that these correlations with VLFI scores did not appear to all sequences omitting event information (Impoverished, Implied) against those that did not (Expected, Explicit), perhaps inferential processing is not the key. Correlations appeared between critical panels of Impoverished and Implied sequences where inferences may not yet have been forthcoming, and where the narrative structure was hypothesized to be the primary difference. Accordingly, these correlations may signal a familiarity with the narrative grammar, not inference generation alone. This interpretation aligns with previous findings that VLFI scores correlate with ERP effects to manipulations of narrative grammar (Cohn et al., 2012) and visual cues relevant to narrative grammar (Cohn & Maher, 2015). In any case, these findings overall provide further evidence that comprehension of sequential images is modulated by participants' "fluency" in a visual language (Cohn, 2013a).

Given our results, it is worth asking: why would Implied and Impoverished-type sequence constructions appear conventionally in visual narratives if they omit important event information, and thereby incur costs for updating a mental model (P600) and reactivating information in memory (frontal negativities)? By withholding relevant information, the author delays the "payoff" at the finale of the strip, and, where appropriate, may inject humor by resolving the prior ambiguity in an amusing way (Attardo & Raskin, 1991; Coulson, 2001). On this point, note that, although participants rated Explicit sequences as the most comprehensible, Impoverished or Implied sequences should be more humorous (as in Figure 2). Thus, building narrative tension through ambiguous events and the subsequent resolution of that uncertainty—as demonstrated by the waveforms at Implied and Impoverished final panels—may contribute to a more satisfying narrative for a reader. Many theorists have posited that storytelling is more engaging if it forces the reader to "interact" with the content (Herman, 2009; Zwaan, 2004). Thus, as has long been emphasized for visual narrative comprehension (McCloud, 1993), omitting information that subsequently requires context updating, as in our examined sequence types, may provide a way to engage and "immerse" readers into a visual narrative for a more rewarding reading experience.

5. Acknowledgements

6. References

Amoruso, L., Gelormini, C., Aboitiz, F., Alvarez González, M., Manes, F., Cardona, J., & Ibanez, A. (2013). N400 ERPs for actions: Building meaning in context. *Frontiers in Human Neuroscience, 7*. doi: 10.3389/fnhum.2013.00057

Attardo, S., & Raskin, V. (1991). Script theory revis(it)ed: joke similarity and joke representation model. *Humor - International Journal of Humor Research, 4*(3-4), 293-348. doi: 10.1515/humr.1991.4.3-4.293

Baggett, P. (1975). Memory for explicit and implicit information in picture stories. *Journal of Verbal Learning and Verbal Behavior, 14*(5), 538-548. doi: http://dx.doi.org/10.1016/S0022-5371(75)80031-4

Baggio, G., van Lambalgen, M., & Hagoort, P. (2008). Computing and recomputing discourse models: An ERP study. *Journal of Memory and Language, 59*(1), 36-53. doi: http://dx.doi.org/10.1016/j.jml.2008.02.005

Barrett, S. E., Rugg, M. D., & Perrett, D. I. (1988). Event-related potentials and the matching of familiar and unfamiliar faces. *Neuropsychologia, 26*(1), 105-117. doi: http://dx.doi.org/10.1016/0028-3932(88)90034-6

Bordwell, D. (1985). *Narration in the Fiction Film*. Madison: University of Wisconsin Press.

Bordwell, D. (2007). *Poetics of Cinema*. New York, NY: Routledge.

Bornkessel-Schlesewsky, I., & Schlesewsky, M. (2008). An alternative perspective on "semantic P600" effects in language comprehension. *Brain Research Reviews, 59*(1), 55-73. doi: http://dx.doi.org/10.1016/j.brainresrev.2008.05.003

Bott, O. (2010). *The processing of events* (Vol. 162): John Benjamins Publishing Company.

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

Brouwer, H., Fitz, H., & Hoeks, J. (2012). Getting real about Semantic Illusions: Rethinking the functional role of the P600 in language comprehension. *Brain Research, 1446*, 127-143. doi: http://dx.doi.org/10.1016/j.brainres.2012.01.055

Burkhardt, P. (2006). Inferential bridging relations reveal distinct neural mechanisms: Evidence from event-related brain potentials. *Brain and Language, 98*(2), 159-168. doi: http://dx.doi.org/10.1016/j.bandl.2006.04.005

Burkhardt, P. (2007). The P600 Reflects Cost of New Information in Discourse Memory. *NeuroReport, 18*(17), 1851-1854.

Chatman, S. (1978). *Story and Discourse*. Ithaca: Cornell University Press.

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

Cohn, N. (2013b). Visual narrative structure. *Cognitive Science, 37*(3), 413-452. doi: 10.1111/cogs.12016

Cohn, N. (2014). You're a good structure, Charlie Brown: The distribution of narrative categories in comic strips. *Cognitive Science, 38*(7), 1317-1359. doi: 10.1111/cogs.12116

Cohn, N., Jackendoff, R., Holcomb, P. J., & Kuperberg, G. R. (2014). The grammar of visual narrative: Neural evidence for constituent structure in sequential image comprehension. *Neuropsychologia, 64*, 63-70. doi: 10.1016/j.neuropsychologia.2014.09.018

Cohn, N., & Maher, S. (2015). The notion of the motion: The neurocognition of motion lines in visual narratives. *Brain Research, 1601*, 73-84. doi: 10.1016/j.brainres.2015.01.018

Cohn, N., & Paczynski, M. (2013). Prediction, events, and the advantage of Agents: The processing of semantic roles in visual narrative. *Cognitive Psychology, 67*(3), 73-97. doi: http://dx.doi.org/10.1016/j.cogpsych.2013.07.002

Cohn, N., Paczynski, M., Jackendoff, R., Holcomb, P. J., & Kuperberg, G. R. (2012). (Pea)nuts and bolts of visual narrative: Structure and meaning in sequential image comprehension. *Cognitive Psychology, 65*(1), 1-38. doi: 10.1016/j.cogpsych.2012.01.003

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

Coulson, S. (2001). *Semantic leaps: Frame-shifting and conceptual blending in meaning construction*: Cambridge University Press.

Coulson, S., & Kutas, M. (2001). Getting it: human event-related brain response to jokes in good and poor comprehenders. *Neuroscience Letters, 316*(2), 71-74. doi: http://dx.doi.org/10.1016/S0304-3940(01)02387-4

Donchin, E., & Coles, M. G. H. (1988). Is the P300 component a manifestation of context updating? *Behavioral and Brain Sciences, 11*(03), 357-374. doi: doi:10.1017/S0140525X00058027

Eisenstein, S. (1942). *Film Sense* (J. Leyda, Trans.). New York, NY: Harcourt, Brace World.

Ferretti, T. R., Rohde, H., Kehler, A., & Crutchley, M. (2009). Verb aspect, event structure, and coreferential processing. *Journal of Memory and Language, 61*(2), 191-205. doi: http://dx.doi.org/10.1016/j.jml.2009.04.001

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.

Hagoort, P., Baggio, G., & Willems, R. M. (2009). Semantic unification. In M. S. Gazzaniga (Ed.), *The cognitive neurosciences, 4th ed*. (pp. 819-836). Cambridge: MIT press.

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.

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.

Herman, D. (2009). *Basic Elements of Narrative*. West Sussex, UK: Wiley-Blackwell.

Hoeks, J. C. J., & Brouwer, H. (2014). Electrophysiological Research on Conversation and Discourse. In T. M. Holtgraves (Ed.), *The Oxford Handbook of Language and Social Psychology* (pp. 365-386). Oxford, UK: Oxford University Press.

Kintsch, W. (1998). *Comprehension: A paradigm for cognition*: Cambridge university press.

Klin, C. M., & Myers, J. L. (1993). Reinstatement of causal information during reading. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 19*(3), 554.

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.

Kos, M., Vosse, T. G., Van Den Brink, D., & Hagoort, P. (2010). About edible restaurants: conflicts between syntax and semantics as revealed by ERPs. *Frontiers in Psychology, 1*. doi: 10.3389/fpsyg.2010.00222

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

Kuperberg, G. R. (2013). The pro-active comprehender: What event-related potentials tell us about the dynamics of reading comprehension. In B. Miller, L. Cutting, & P. McCardle

(Eds.), *Unraveling the Behavioral, Neurobiological, and Genetic Components of Reading Comprehension*. Baltimore: Paul Brookes Publishing.

Kuperberg, G. R., Holcomb, P. J., Sitnikova, T., Greve, D., Dale, A. M., & Caplan, D. (2003). Distinct Patterns of Neural Modulation during the Processing of Conceptual and Syntactic Anomalies. *Journal of Cognitive Neuroscience, 15*(2), 272-293. doi: 10.1162/089892903321208204

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

Kutas, M., & Federmeier, K. D. (2011). Thirty years and counting: Finding meaning in the N400 component of the Event-Related Brain Potential (ERP). *Annual Review of Psychology, 62*(1), 621-647.

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

Long, D. L., Oppy, B. J., & Seely, M. R. (1994). Individual differences in the time course of inferential processing. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 20*(6), 1456. doi: 10.1037/0278-7393.20.6.1456

Long, D. L., Oppy, B. J., & Seely, M. R. (1997). Individual Differences in Readers' Sentence- and Text-Level Representations. *Journal of Memory and Language, 36*(1), 129-145. doi: http://dx.doi.org/10.1006/jmla.1996.2485

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

Magliano, J. P., & Zacks, J. M. (2011). The Impact of Continuity Editing in Narrative Film on Event Segmentation. *Cognitive Science, 35*(8), 1489-1517.

McCloud, S. (1993). *Understanding Comics: The Invisible Art*. New York, NY: Harper Collins.

McNamara, D. S., & Magliano, J. (2009). Toward a comprehensive model of comprehension. *Psychology of learning and motivation, 51*, 297-384.

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

Myers, J. L., & O'Brien, E. J. (1998). Accessing the discourse representation during reading. *Discourse Processes, 26*(2-3), 131-157. doi: 10.1080/01638539809545042

Nakazawa, J. (2005). Development of manga (comic book) literacy in children. In D. W. Shwalb, J. Nakazawa, & B. J. Shwalb (Eds.), *Applied Developmental Psychology: Theory, Practice, and Research from Japan* (pp. 23-42). Greenwich, CT: Information Age Publishing.

Nieuwland, M. S., & Van Berkum, J. J. A. (2005). Testing the limits of the semantic illusion phenomenon: ERPs reveal temporary semantic change deafness in discourse comprehension. *Cognitive Brain Research, 24*(3), 691-701. doi: 10.1016/j.cogbrainres.2005.04.003

Osaka, M., Yaoi, K., Minamoto, T., & Osaka, N. (2014). Serial changes of humor comprehension for four-frame comic Manga: an fMRI study. *Scientific Reports, 4*. doi: 10.1038/srep05828

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

Paczynski, M., Jackendoff, R., & Kuperberg, G. (2014). When Events Change Their Nature: The Neurocognitive Mechanisms Underlying Aspectual Coercion. *Journal of Cognitive Neuroscience, 26*(9), 1905-1917. doi: 10.1162/jocn_a_00638

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

Reid, V. M., & Striano, T. (2008). N400 involvement in the processing of action sequences. *Neuroscience Letters, 433*(2), 93-97. doi: http://dx.doi.org/10.1016/j.neulet.2007.12.066

Saraceni, M. (2001). Relatedness: Aspects of textual connectivity in comics. In J. Baetens (Ed.), *The Graphic Novel* (pp. 167-179). Leuven: Leuven University Press.

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

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

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.

Trabasso, T., & Suh, S. (1993). Understanding text: Achieving explanatory coherence through on-line inferences and mental operations in working memory. *Discourse Processes, 16*(1-2), 3-34. doi: 10.1080/01638539309544827

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

van Berkum, J. J. A., Koornneef, A. W., Otten, M., & Nieuwland, M. S. (2007). Establishing reference in language comprehension: An electrophysiological perspective. *Brain Research, 1146*(0), 158-171. doi: http://dx.doi.org/10.1016/j.brainres.2006.06.091

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

Võ, M. L.-H., & Wolfe, J. M. (2013). Differential Electrophysiological Signatures of Semantic and Syntactic Scene Processing. *Psychological Science*. doi: 10.1177/0956797613476955

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

Wilson, B., & Wilson, M. (1987). Pictorial Composition and Narrative Structure: Themes and Creation of Meaning in the Drawings of Egyptian and Japanese Children. *Visual Arts Research, 13*(2), 10-21.

Wittenberg, E., Paczynski, M., Wiese, H., Jackendoff, R., & Kuperberg, G. (2014). The difference between "giving a rose" and "giving a kiss": Sustained neural activity to the light verb construction. *Journal of Memory and Language, 73*(0), 31-42. doi: http://dx.doi.org/10.1016/j.jml.2014.02.002

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.

Yus, F. (2008). Inferring from comics: A multi-stage account. *Quaderns de Filologia. Estudis de Comunicacio, 3*, 223-249.

Zwaan, R. A. (2004). The immersed experiencer: Toward an embodied theory of language comprehension. In B. H. Ross (Ed.), *The psychology of learning and motivation* (Vol. 44, pp. 35-62). New York: Academic Press.

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