Brain and Cognition 146 (2020) 105634

Contents lists available at ScienceDirect

Brain and Cognition

journal homepage: www.elsevier.com/locate/b&c

Zooming in on the cognitive neuroscience of visual narrative

Neil Cohn [a,*], Tom Foulsham [b]

[a] Department of Communication and Cognition, Tilburg School of Humanities and Digital Sciences, Tilburg University, Netherlands
[b] Department of Psychology, University of Essex, UK

ARTICLE INFO

Keywords:
Visual language
N400
N300
P600
Comics
Film

ABSTRACT

Visual narratives like comics and films often shift between showing full scenes and close, zoomed-in viewpoints. These zooms are similar to the "spotlight of attention" cast across a visual scene in perception. We here measured ERPs to visual narratives (comic strips) that used zoomed-in and full-scene panels either throughout the whole sequence context or at specific critical panels. Zoomed-in panels were automatically generated on the basis of fixations from prior participants' eye movements to the crucial content of panels (Foulsham & Cohn, 2020). We found that these fixation panels evoked a smaller N300 than full-scenes, indicative of reduced cost for object identification, but that they also evoked a slightly larger amplitude N400 response, suggesting a greater cost for accessing semantic memory with constrained content. Panels in sequences where fixation panels persisted across all positions of the sequence also evoked larger posterior P600s, implying that constrained views required more updating or revision processes throughout the sequence. Altogether, these findings suggest that constraining a visual scene to its crucial parts triggers various processes related not only to the density of its information but also to its integration into a sequential context.

1. Introduction

Many theories of visual narratives like comics and film emphasize the ways in which comprehension overlaps with basic aspects of event and perceptual cognition (Loschky, Hutson, Smith, Smith, & Magliano, 2018). Some aspects of visual narratives may appear to depart from daily event perception, such as the way they can modulate the framing of content, where there may be contrasts between images with full and close-up viewpoints. However, this variation in framing can create a simulated "spotlight" of attention by using the frame to window specific information about a scene while filtering out other information (Cohn, 2013). That is, authors can use framing to guide the reader through an unfolding event structure in a way that simulates a perceptual experience of directing attention to different parts of a scene. We ask here: to what degree does altering such simulated attentional structure affect the online processing of visual narratives? Such work is informative both for research on attention and perception—given the visual nature of these narratives—but also for work on language and discourse—given their capacity for sequential meaning-making (Cohn & Magliano, 2020).

Recent models of sequential image comprehension have emphasized that processing passes through several stages (Cohn, 2020b; Loschky et al., 2018; Loschky, Magliano, Larson, & Smith, 2020). A comprehender will search a visual image to extract the relevant information (Loschky et al., 2020; Magliano, Loschky, Clinton, & Larson, 2013), which is then fed to comprehension processes in order to build a "situation model"—a mental model comprising the knowledge of the entities and events that unfold throughout the narrative (McNamara & Magliano, 2009; van Dijk & Kintsch, 1983). We can characterize this process broadly as starting with extracting cues from a visual signal, which allows for the *access* of semantic memory for the relevant information, which is then used to *update* the situation model based on the degree of contiguity with expectancies established from the prior context.

In the initial processes, when viewing a panel in a visual narrative, a comprehender will assess a picture for its relevant semantic cues (Hutson, Magliano, & Loschky, 2018; Loschky et al., 2018), particularly characters and their parts (Laubrock, Hohenstein, & Kümmerer, 2018). In visual narratives, search for such cues benefits from prior images in a sequence, and indeed fewer fixations appear to panels in coherently ordered visual narratives than panels in scrambled sequences where the sequence is uninformative (Foulsham, Wybrow, & Cohn, 2016). These visual cues provide signals to feed into the representations in semantic

* Corresponding author at: Tilburg University, Tilburg School of Humanities and Digital Sciences, Department of Communication and Cognition, P.O. Box 90153, 5000 LE Tilburg, Netherlands.
E-mail address: neilcohn@visuallanguagelab.com (N. Cohn).

https://doi.org/10.1016/j.bandc.2020.105634
Received 9 May 2020; Received in revised form 9 October 2020; Accepted 19 October 2020

memory. The access of semantic information is implicated in event-related brain potentials (ERPs) by the N400 response—a negative polarity deflection that peaks roughly 400 ms after the onset of a stimulus (Kutas & Federmeier, 2011). The N400 is thought to reflect a default neural process of accessing or retrieving information in semantic memory (Kutas & Federmeier, 2011). This brain response appears to be a domain-general index of semantic processing, and occurs consistently to words, pictures, sounds, and multimodal interactions, where the amplitude is modulated by the degree of semantic overlap of a stimulus item with its preceding context (Kutas & Federmeier, 2011), including in visual events (Sitnikova, Kuperberg, & Holcomb, 2003) and visual narratives (Cohn, Paczynski, Jackendoff, Holcomb, & Kuperberg, 2012; West & Holcomb, 2002).

Although the N400 itself is evoked by stimuli across modalities, N400s to images are also preceded by an N300 (McPherson & Holcomb, 1999), thought to reflect the rapid identification and/or categorization processes involved with accessing semantic *visual* information (Hamm, Johnson, & Kirk, 2002; Truman & Mudrik, 2018), or the structural mapping of visual features onto semantic representations (Schendan & Kutas, 2003). This latter view is buttressed by findings of N300s also to signs in sign languages (Meade, Lee, Midgley, Holcomb, & Emmorey, 2018). While some have posited that the N300 is a unique precursor to the N400 in visual stimuli (McPherson & Holcomb, 1999; Truman & Mudrik, 2018), other work has considered them as inseparable (Draschkow, Heikel, Võ, Fiebach, & Sassenhagen, 2018), particularly based on studies comparing congruous and incongruous objects within visual scenes (Draschkow et al., 2018; Hamm et al., 2002; Lauer, Cornelissen, Draschkow, Willenbockel, & Võ, 2018). In these studies, negative deflections consistent with both an N300 and an N400 are observed in response to incongruous objects (such as a football in a kitchen). However, multivariate pattern analysis suggests that these two components may come from the same source (Draschkow et al., 2018), and that scene context therefore has a general impact on both early and later processing (Draschkow et al., 2018; Mudrik, Lamy, & Deouell, 2010).

Studies of sequential image comprehension have suggested that semantic access is more difficult at the start of a sequence where no information has yet been established. This process of "laying a foundation" for the subsequent sequence (Gernsbacher, 1990) is supported by longer reading times and slower reaction times to target panels at the start of a sequence than at the end (Cohn & Paczynski, 2013; Cohn & Wittenberg, 2015; Foulsham et al., 2016). Some work has speculated that these reading times are motivated by greater demand for attentional search processes to explore an as-yet unfamiliar narrative (Loschky et al., 2018, 2020). While it is possible that attentional processes are engaged more at the start of a sequence, such perceptual search behavior may be motivated by later comprehension processes. Indeed, studies using ERPs have shown that larger N400 amplitudes appear at the first position of a visual narrative sequence and are attenuated across ordinal sequence position (Cohn et al., 2012). Attenuation of the N400 also appears across ordinal word position in sentence processing (Van Petten & Kutas, 1991), suggesting such processes are a feature of sequential comprehension more generally.

Throughout the reading of a visual narrative sequence, this semantic information then becomes incorporated into a growing situation model of the scene (Cohn, 2020b; Loschky et al., 2020). Shifts between images in dimensions of characters, spatial locations, or events incur a cost for incorporating this altered information into the ongoing understanding of the discourse (McNamara & Magliano, 2009; Zwaan & Radvansky, 1998). Given the prior context of a sequence, a situation model may use "mapping"—which involves only a cursory, incremental updating to a previous state—or it may require "shifting"—an updating process requiring a more significant revision to a whole new situation model (Huff, Meitz, & Papenmeier, 2014; Loschky et al., 2018, 2020). For example, minimal change between panels demand only mapping processes, but significant changes may require resolving more ambiguity through shifting between states of a situation model, like those demanded of inference generation. Thus, while updating processes may differ, they are overall viewed as an ongoing process at each unit of a visual narrative.

In ERPs to language and visual narratives, updating or revision processes have been implicated by positivities, such as the P600, a positive deflection peaking around 600 ms after the onset of a stimulus image (Baggio, 2018; Kuperberg, 2016; Leckey & Federmeier, 2020). In studies of visual narratives, greater P600s have been evoked in contexts involving unexpected changes in characters (Cohn & Kutas, 2015, 2017), or resolving inexplicit or incongruous actions (Cohn & Maher, 2015), a finding consistent with reanalysis of confounded expectations for visual events outside a narrative context (Amoruso et al., 2013; Sitnikova, Holcomb, & Kuperberg, 2008). However, P600s are not just evoked by incongruous information, but also occur to any shifts between characters or event states (Cohn & Kutas, 2015), which implies that updating is an ongoing process, not a surprisal response. However, no extant research has yet focused on this ongoing process specifically in the context of visual narratives.

Nevertheless, some research has suggested that the whole contents of a visual narrative image is not necessary for the sequential meaning, but rather that specific cues motivate the processing and updating of the sequence. For example, exploratory research on eye-movements in comics reports more fixations focused on characters than backgrounds, particularly first fixations (Laubrock et al., 2018). More experimental research has found that omission of focal cues that signal off-panel events (Cohn & Kutas, 2015) and those signaling movement (Cohn & Maher, 2015) lead to updating processes indexed by P600s. In addition, participants tend to agree on—and direct their attention to—the cues within images that are pertinent for drawing inferences (Hutson et al., 2018).

Thus, given that not all information in a panel is relevant for the sequencing of a visual narrative, focusing only on this specific information—as in a panel with a "zoomed-in" viewpoint—might convey the requisite information needed for a visual sequence. The question is then whether a panel that zooms in on relevant information would be sufficient compared to one that shows a full view of the scene. This logic is similar to studies using gaze-contingent moving window designs, where, in some cases, processing of a scene proceeds normally even when visual information away from fixation is removed (Loschky, McConkie, Yang, & Miller, 2005). If a zoom panel is sufficient, might then the access of semantic information be comparable between zoomed and full-scene panels, and would it require relatively little updating?

We first examined this role of focal information using images automatically generated from an eye-tracking study. In previous work, we tracked the fixations of participants' eye movements to visual narratives that were presented in either a coherent or scrambled order (Foulsham et al., 2016). To study which information in a panel was relevant for a sequence, we generated a heatmap of fixation data, and selected the area in a panel with the top 10% of fixations. As illustrated in Fig. 1, this region was then cropped and enlarged to form a new panel consisting only of information fixated by participants (Foulsham & Cohn, 2020). This process was entirely automatic and motivated solely by the data gathered in our prior study.

Using these automatically generated "fixation panels", we conducted a series of studies measuring participants' self-paced viewing times to each panel in a sequence (Foulsham & Cohn, 2020). We first showed that viewing times were longer for sequences where all the panels were presented in full than those zooming-in on fixated information, which in turn were longer than those where all panels zoomed-in on non-fixated information. However, across the ordinal position of the sequence, fixation panels were viewed for nearly the same duration as full panels after the first position of the sequence. We next compared sequences with full-scenes in each panel, but which manipulated only a single, critical panel. Here, panels showing only fixated information had the same viewing times as those showing a full-scene, and these panels were

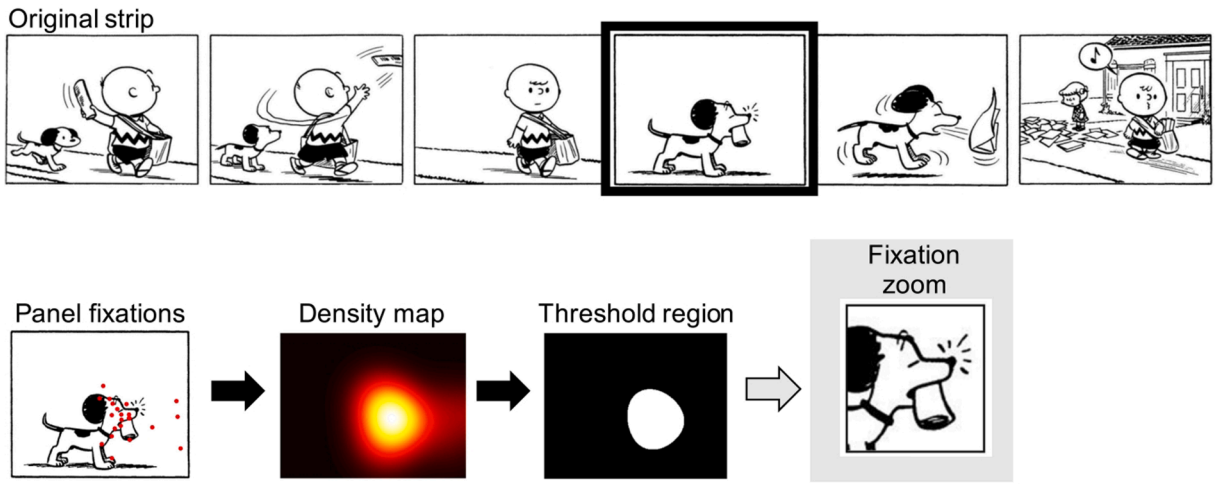

Fig. 1. Our method of automatically generating panels using the top 10% of fixations from a prior eye-tracking study (Foulsham et al., 2016), originally used in Foulsham and Cohn (2020). Peanuts artwork is © Peanuts Worldwide LLC.

both shorter than panels showing non-fixated information or fixated information that was incongruous to the sequence context. In addition, panels following the critical panel in the full-scene and fixation zoom conditions did not differ in their viewing times, though panels after non-fixation or anomalous panels were viewed longer. Overall, these results suggest that comprehenders face minimal costs for processing fixated information compared to full-scenes.

Nevertheless, behavioral measures like viewing times do not always reveal cognitive processes that appear with more sensitive measures, such as ERPs (e.g., Cohn & Maher, 2015). Here, we ask whether the minimal differences observed between fixation panels and full-scene panels would also manifest in neurocognition, or whether measuring ERPs would reveal more processes at work in the comprehension of these sequences. We thus measured ERPs to visual sequences that crossed fixation and full-scene panels throughout the whole sequence context or only at specific critical panels. As depicted in Fig. 2, this resulted in sequence contexts with full-scene panels that had critical panels with either a full or fixation panel, or zoom sequences with all fixation panels, where the critical panel had either a full or fixation panel.

If our neurocognitive findings are consistent with our prior behavioral results, we would expect that critical fixation panels and full panels would trigger relatively the same demands on semantic access, and therefore would differ minimally in the N400s that they generate. Such a result would suggest that focal information in a fixation panel is sufficient for providing the relevant semantic cues given the sequence context as a full-scene panel. Nevertheless, a second outcome could also be possible: If the increase of information in full-scene panels needs to be processed beyond that provided in the fixation panels, access should be *easier* for fixation panels because it requires less information to spread throughout a semantic network, thus resulting in an attenuated N400.

While fixated viewpoints may or may not influence the access of semantic processing, this constrained information could incur costs for updating of a mental model, particularly in sequences where all panels in a sequence depict zoomed-in framing. If each panel in a sequence depicts only focal information, additional contextual information may

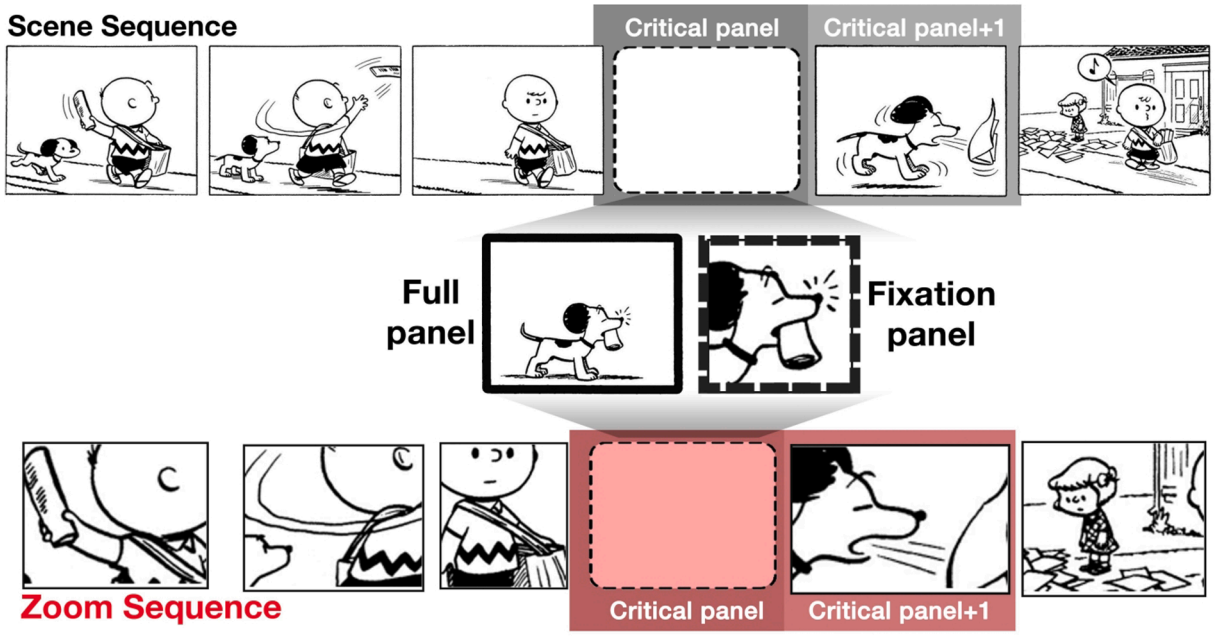

Fig. 2. Experimental sequence types crossing full-scene panels and zoomed-in panels for all the panels of a sequence and/or a specific critical panel. Peanuts artwork is © Peanuts Worldwide LLC.

be lost, thereby increasing the cost of integrating incoming information into a sequential context. Thus, in a zoom sequence context, the content of any critical panel should be harder to incorporate into a mental model, regardless of its own framing. Such updating would be suggested by greater P600s for panels in zoom sequences than in full-scene sequences. This would be consistent with self-reports from our behavioral study which showed that sequences where all the panels were zoomed-in were rated as more difficult to comprehend (Foulsham & Cohn, 2020).

An additional facet of this design is that it affords us the opportunity to compare ERPs across the ordinal position of the sequence using the different framing of information. Prior work has shown that the N400 is attenuated across the ordinal position of panels in coherent visual narratives, but not for sequences where panels lack semantic associative relationships and/or a narrative structure (Cohn et al., 2012). Here we ask, will constraining the viewpoint in each panel of a sequence affect its processing across the ordinal position of a sequence?

In our prior behavioral work measuring self-paced viewing times, minimal differences appeared between fixation and full-scene panels across the ordinal position of a sequence, with differences being most pronounced at the first and last panels of a sequence (Foulsham & Cohn, 2020). Here, we predicted that both fixated and full-scene panels will still attenuate N400s across the ordinal position of the sequence (Cohn et al., 2012), because the coherent sequence will still sufficiently allow for reactivation of information from the prior context (Kuperberg, 2016; Kutas & Federmeier, 2011). However, zoom sequences may lead to larger N400s, because repeatedly constraining incoming information will provide fewer bottom-up cues for reactivation. That is, panels depicting full scenes have a combination of 1) focal cues relevant for the changing sequential meaning, and 2) non-focal cues often persisting unchanged across each panel. Zoom sequences depicting only fixation panels thus will filter out the non-focal cues that passively reactivate aspects of semantic memory across the ordinal sequence position, leaving only changing focal cues. This should thus lead to comparatively larger N400s across the ordinal positions of the zoom sequences with fixation panels than for scene sequences with full panels, as fewer features in semantic memory become continuously reactivated, and may even lead to increasing N400s with each panel position.

In addition, such focal cues signal changes from panel to panel related to the sequential meaning, and thus should trigger updates to the situation model. Thus, repeatedly depicting only constrained zoomed-in framing may evoke greater updates to a situation model than depicting full-scenes, where such cues would be less salient. This contrast thus provides a way to examine whether such updating processes persist across the ordinal position of sequences with no incongruous situational changes. If such updating is present across ordinal sequence position, we expected to see greater P600s to fixation panels in zoom sequences than to full panels in scene sequences. In addition, if such updating has an additive cost, we might expect such positivities to be greater across each position of the sequence, in reverse of the type of attenuation observed to N400s.

2. Materials and methods

2.1. Stimuli

We used the 72 visual sequences constructed out of panels from *The Complete Peanuts* which appeared in our prior eye-tracking study of visual narrative comprehension (Foulsham et al., 2016) and in prior behavioral studies (Foulsham & Cohn, 2020). Sequences were 6-panels in length, with no text. We crossed full-scene panels and fixation panels across whole sequences and at a specific, critical panel. In "scene sequences", each non-critical panel in a sequence was shown with a "full panel" depicting a full scene. In "zoom sequences", all non-critical panels used "fixation panels" created using fixation data. The process for constructing these stimuli is described in the Introduction and in Foulsham and Cohn (2020). In brief, we automatically selected and then cropped sections from each panel, based on a heatmap distribution of the fixations made by an independent group of observers. The topmost 10% of each distribution was selected, reflecting the region that received the most frequent fixations from 14 representative observers (Foulsham et al., 2016). The closest fitting rectangular bounding box around this region was then cropped, enlarged to the same height as the full panels, and framed with a black border. The result was a series of zoom panels which highlight the focal details, based not on experimenter judgment but on the unconstrained attentional selection of naïve observers.

We then further manipulated a critical panel position to either have a full panel or a fixation panel, within the context of each sequence type. As in our prior work (Foulsham & Cohn, 2020), critical panels fell at the narrative "Initial"—i.e., a panel typically showing a preparatory action, preceding the "Peak" panel that contained the climactic events of the sequence (Cohn, 2013). Because Initial panels are posited to often have cues relevant for anticipating the subsequent primary actions (Peaks), we chose Initials to assess how framing of these cues might affect the processing of this downstream information. Altogether, as depicted in Fig. 2, this manipulation created a factorial design crossing the framing of panels within sequences (scene, zoom) and critical panels (full, fixation).

All sequences were counterbalanced in a Latin Square design into 4 lists such that each sequence appeared only once per list, but all conditions for a sequence appeared across lists. Thus, participants viewed each sequence only one time in one condition, but all conditions of a sequence were viewed an equal number of times across all participants. Experimental sequences were combined with 96 additional filler sequences varying in their comprehensibility to create added heterogeneity into the sequences. Lists were distributed evenly across participants such that all sequence types were viewed an equal number of times, though each participant viewed a unique order of sequences in their list, randomized using the PsychoPy2 (Peirce et al., 2019) experimental presentation software.

2.2. Participants

We recruited 26 participants from Tilburg University (14 male, 12 female; mean age: 22.3) to participate in the study, of which 2 were excluded for having unusable data due to excessive eye-movements and alpha. Although we did not calculate statistical power a priori, power analysis by simulation indicates that, with a fully within-subjects design, this sample size yields good power for our 2×2 experimental factors. All participants gave their informed written consent to participate. Participants were right-handed with normal or corrected-to-normal vision and no history of head trauma, and were on no psychoactive medication. Before participating in the study, participants filled out the *Visual Language Fluency Index* (VLFI) questionnaire which assessed their frequency of reading and drawing comics across several self-rated scales, in addition to their experience with movies and written books. *VLFI scores* calculated from this assessment have been shown to be predictive of individual differences in several measures of visual narrative comprehension, including ERP amplitudes (Cohn & Kutas, 2015; Cohn & Maher, 2015; Cohn et al., 2012). The 24 participants retained in the analysis had an average VLFI score of 17.8 (SD = 6.3; range: 7.25–27.5), which is a high average, where low is below 8, average is 12, and high is 22 and above.

2.3. Procedure

EEG was measured in a soundproofed chamber, where participants sat ~110 cm away from a computer screen with a keyboard on their lap. Trials were presented using PsychoPy2 (Peirce et al., 2019). A grey screen reading "Ready" in white letters began each trial. A red dot persisted in the center of the screen throughout the experiment in order to give participants a fixation point to reduce eye-movements. Participants then pressed a button on the keyboard to view panels which

appeared in the center of an otherwise grey screen for a duration of 1350 ms as in prior research (Cohn & Kutas, 2017; Cohn & Maher, 2015; Cohn, Jackendoff, Holcomb, & Kuperberg, 2014). Because of the automatic panel-making process, panels had slight differences in horizontal sizing, but each was approximately 10 × 8 cm, with a visual angle of ~5.2° horizontally and 4.2° vertically. A 300 ms ISI separated panels in order to prevent an effect of figures becoming animated like a flipbook. Following the last panel of a sequence, a question mark appeared on the screen where participants rated the comprehensibility of a sequence on a scale of 1 (=hard to understand) to 7 (=easy to understand). After the EEG experiment, participants filled in a questionnaire that asked them to describe any patterns or characteristics of the sequences that they may have noticed.

2.4. Data analysis

We measured EEG at a sampling rate of 250 Hz, using a Brain Products ActiChamp system, and recorded from the scalp using 32 channel Standard actiCAPs, which were referenced online to electrode Fz and re-referenced offline to the average of the mastoid channels (TP9, TP10). Eye-movements and blinks were measured with electrodes placed beneath the right eye and beside the left eye. All electrode impedances were kept below 10 kΩ. EEG data was analyzed using the ERPLAB plugin for EEGLAB in MATLAB (Lopez-Calderon & Luck, 2014). Data was refiltered offline with a bandpass filter of 0.1–30 Hz. We extracted epochs of 1500 ms with a 200 ms pre-stimulus baseline.

Our artifact rejection procedure removed trials due to eye-movements, blinks, muscle tension, and/or alpha using ICA. Rejection rates were kept below 10% of all trials, and participants were removed from analysis if they exceeded this threshold. This resulted in two participants being excluded from the final analyses (as mentioned above). An additional 0.1–15 Hz filter was applied for waveforms depicted in the figures, but this filtered data was not used in statistical analyses.

Our behavioral analysis compared participant's comprehensibility ratings for the whole sequences. We used a 2 × 2 repeated-measures ANOVA with factors of Sequence Type (2: Scene vs. Zoom), and Panel Type (2: Full vs. Fixation). Our analysis of ERPs focused on our critical, manipulated panels and the panels after them (critical panel +1). We focused on the epochs of 200–300, 300–500, 500–800, and 800–1100 ms, corresponding to expected ERP effects of the N300, N400, P600, and sustained effects, respectively. We additionally analyzed the earlier epoch of 100–200 ms to assess the influence of any stimulus differences at the critical panel. To analyze the scalp distribution of our ERP effects, as depicted in Fig. 3, our analysis also divided the scalp across 16 electrodes that allowed for contrasts of Hemisphere (2: left, right), Laterality (2: medial, lateral), and Anterior-Posterior distribution (4: prefrontal, fronto-central, centro-parietal, occipital). This method is consistent with analyses used in previous studies of language and visual narrative (Cohn & Kutas, 2015, 2017; Metusalem et al., 2012). We report findings for these factors only when interacting with our primary factors of Sequence Type and Panel Type to situate our ERP effects across the scalp. Our analysis used repeated-measures ANOVAs with factors of Sequence Type (2: Scene vs. Zoom), Panel Type (2: Full vs. Fixation), Hemisphere, Laterality, and AP Distribution.

An additional analysis examined whether zoom sequences differed from scene sequences in amplitude across the ordinal position of the sequence. We followed the methods of prior work (Coderre et al., 2018; Cohn et al., 2012) and averaged the ERP amplitudes of non-critical

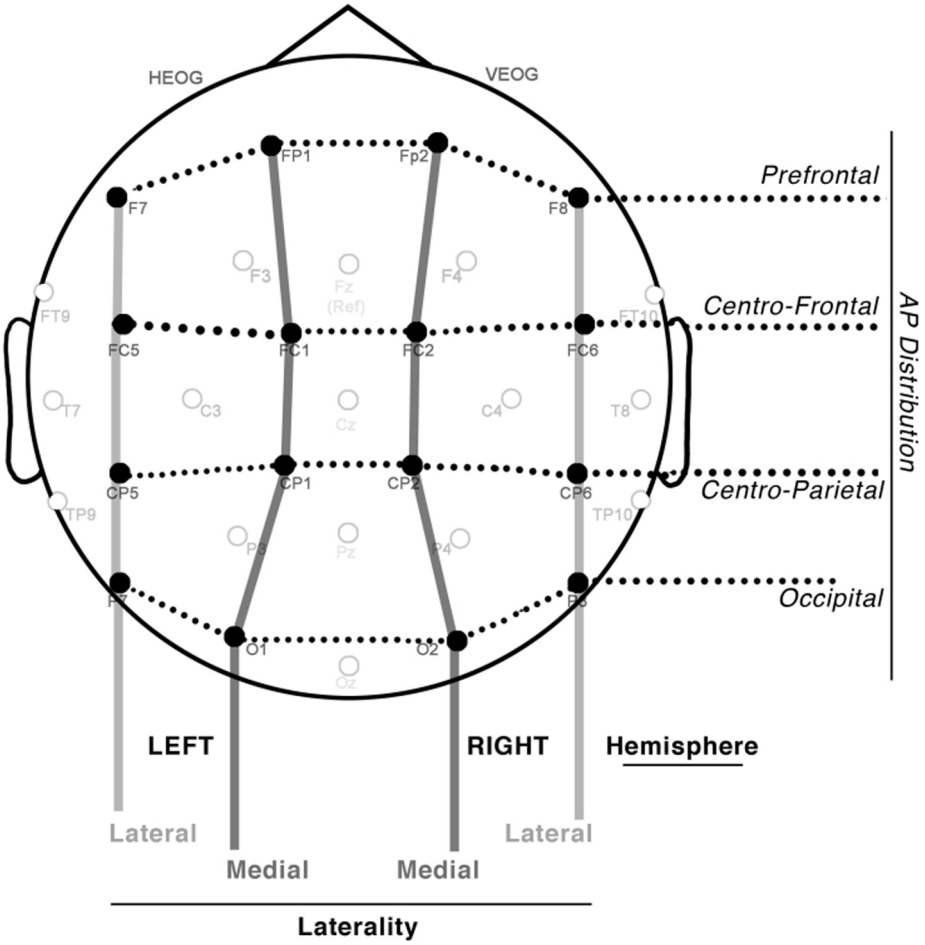

Fig. 3. Montage analysis dividing 16 electrodes across Hemisphere, Laterality, and Anterior-Posterior Distribution.

panels for each panel in a sequence across the frontal regions of the scalp (prefrontal, frontal, frontocentral), and in addition analyzed averages across posterior regions (centroparietal, parietal occipital). These regions sought to maximize the expected effects of the N400 and P600 respectively. Amplitudes in these regions were averaged for each epoch, and then analyzed using 2 (Sequence Type: Scene vs. Zoom) × 6 (Position) repeated-measures ANOVAs.

Finally, to investigate the influence of comic reading expertise on our results, Pearson's correlations with an alpha level set to 0.05 were used to compare VLFI scores with comprehension scores, and the mean amplitude differences between conditions, averaged across all electrode sites on the scalp.

3. Results

3.1. Behavioral results

Participants' ratings of sequences' comprehensibility revealed main effects of both Sequence Type and Panel Type (all Fs > 20.2, all ps < 0.001), but no interaction between them (p = .329). This arose because scene sequences with critical full panels (M = 5.29, SD = 0.14) were rated as more comprehensible than scene sequences with fixation panels (M = 5.0, SD = 0.15), and these were both rated as more comprehensible than zoom sequences with a full panel (M = 3.9, SD = 0.2) which was more comprehensible than zoom sequences with a critical fixation panel (M = 3.4, SD = 0.19). VLFI scores positively correlated with comprehensibility ratings of all sequences (all rs > 0.42, all ps < 0.05), except ratings for scene sequences with fixation panels, which only approached the threshold of significance, r(22) = 0.389, p = .06. These correlations suggested that participants with more experience reading comics found all sequence types to be easier to comprehend. Finally, participants were consciously aware of the zoom manipulation, with 74% (17 of 24) mentioning fixation panels without prompting in their post-experiment questionnaires.

3.2. Critical panel

ERPs at the critical panel in different conditions could indicate processing differences for that particular panel (full vs. fixation), the context of the sequence (scene vs. zoom) or both. Waveforms and topographic maps for ERPs at the critical panel are depicted in Fig. 4. Additional plots depicting the amplitude differences in each epoch for Panel Type (fixation minus full) and Sequence Type (zoom minus scene) are graphed in Fig. 5. A first difference between conditions was

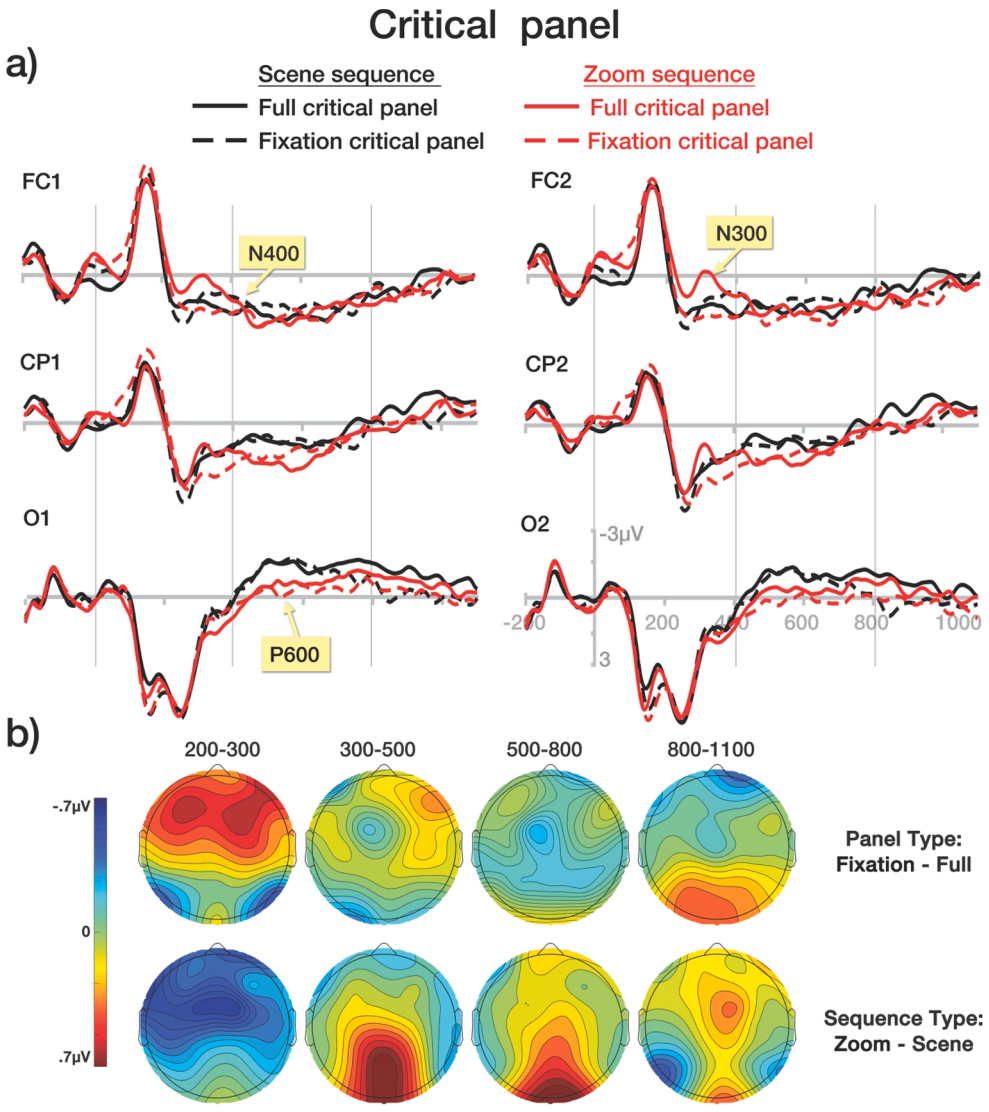

Fig. 4. a) Event-related potentials to full-scene and zoom panels placed within sequences with either full-scene or zoom panels, and b) topographic maps depicting the distribution of effects for panel types (fixation – full) or sequence types (zoom – scenes).

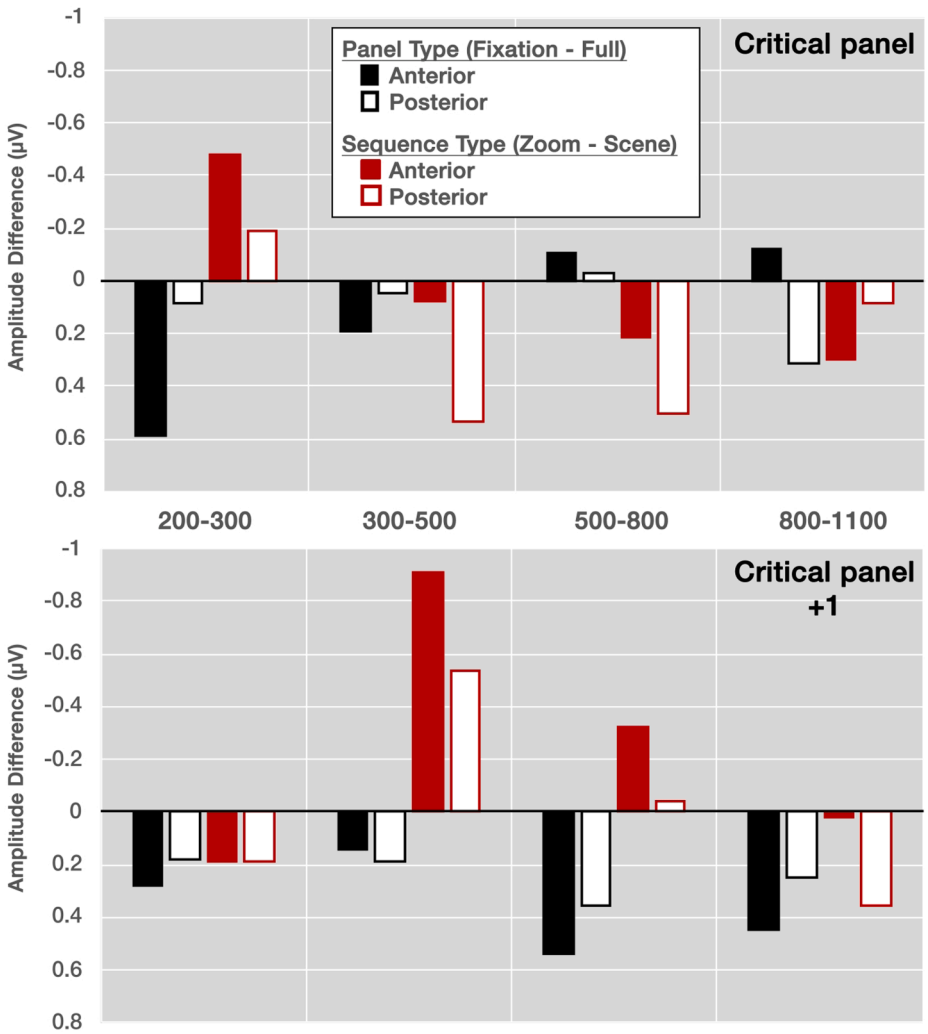

Fig. 5. Difference amplitudes for the factors of Panel Type (fixation – full) and Sequence Type (zoom – scene) averaged across both anterior and posterior electrodes. Zero means no difference between conditions, while bars above the x-axis indicate greater negative amplitude differences, while bars below represent greater positive differences.

implicated in the early, 100–200 ms epoch where we observed interactions of Panel Type × AP Distribution, and Sequence Type and Panel Type each interacting with Laterality × AP Distribution, and both of them with Hemisphere × Laterality × AP Distribution (all statistics are provided in Table 1). These interactions arose because critical panels in zoom sequences evinced a larger fronto-central negativity than those in scene sequences, but fixation panels in scene sequences had the least negative amplitude.

In the 200–300 ms epoch, a Panel Type × AP Distribution interaction suggested a greater frontal negativity (N300) to full panels than fixation panels, regardless of sequence type. This implied that fixation panels incurred less costs than full panels of processes associated with object identification (Draschkow et al., 2018; Hamm et al., 2002) or structural feature mapping (Schendan & Kutas, 2003). In contrast, in the 300–500 ms epoch, a frontal negativity (N400) appeared to fixation panels compared to full panels in scene sequences as did full panels in zoom sequences. Thus, despite their attenuation in the N300, fixation panels appeared to evoke greater costs of semantic processing (N400) compared to the full panels. This was suggested by interactions of Sequence Type × Panel Type × AP Distribution and Sequence Type × Panel Type × Laterality × AP Distribution. However, as shown in the graphs plotting amplitude differences for each factor in Fig. 5, no negativity effect was evident for Panel Type or Sequence Type on their own. Rather, these interactions suggested the start of a posterior positivity (P600) for critical panels in zoom sequences compared to those in scene sequences. This positivity was also suggested by Sequence Type × Laterality interactions that persisted throughout the 300–500 ms, 500–800 ms, and 800–1100 ms epochs. Such a positivity implies that panels from zoom sequences required greater updating or revision than those from scene sequences, regardless of the framing of the panel (Baggio, 2018; Cohn, 2020b; Kuperberg, 2016; Leckey & Federmeier, 2020).

Finally, in the 800–1100 ms epoch, an additional Sequence Type × Panel Type × Laterality × AP Distribution interaction arose because of a late frontal positivity to all panels other than the full panels in the scene sequences. As this effect occurred to all panels deviating from the canonical depiction of a sequence with all full panels, it implies that this late frontal positivity was sensitive to non-normative aspects of the framing in the sequence.

3.3. *Critical panel +1*

The manipulation at the critical panel persisted into the subsequent panel (Table 1, Fig. 6, Fig. 5). We first observed interactions in the 100–200 ms epoch between Sequence Type × AP Distribution and Sequence Type × Laterality × AP Distribution. Panels in zoom sequences evoked a greater frontal negativity and posterior positivity than those from scene sequences. In 200–300 ms epoch, interactions occurred

Table 1
F-values for results of ANOVAs comparing Sequence Types (S) and Panel Types (P) at the critical panel and critical panel + across Hemisphere (H), Laterality (L), Anterior-Posterior Distribution (AP). ^p < .1, * p < .05, ** p < .01, ***p < .001. df = 1,23, except those with AP Distribution: 3,69.

	Critical Panel					Critical Panel +1				
	100-200	200-300	300-500	500-800	800-1100	100-200	200-300	300-500	500-800	800-1100
Sequence (S)	.37	2.2	.46	1.6	.39	3.6^	.5	5.8*	.89	.83
Panel (P)	3.4^	1.8	.36	.03	.21	.72	1.2	.6	2.1	1.4
S*P	.07	.95	3.6^	.36	.03	.08	.01	2.0	.003	.54
S*H	1.8	.19	.3	.45	1.2	.19	1.1	.95	.06	.24
P*H	.61	.06	1.6	.01	.03	.09	.67	.31	1.0	.84
S*P*H	1.1	2.8	.1	2.7	1.6	.26	.25	.06	.01	.58
S*L	.68	.01	8.5**	11.0**	6.8*	.23	.17	8.2**	1.7	.06
P*L	.17	.63	.51	.51	.003	2.2	.88	.86	4.9*	3.7^
S*P*L	.1	.69	2.5	.62	.15	.22	.13	1.6	.003	.01
S*AP	1.8	.79	1.1	.4	1.1	3.8*	.8	1.1	1.4	1.4
P*AP	4.4*	2.9*	.52	.25	2.5^	.5	.28	.04	.97	.47
S*P*AP	1.8	2.0	4.3*	.65	1.0	.98	.23	.13	1.7	1.8
S*H*L	1.7	.74	.27	.08	.01	.01	.01	.002	.04	.003
P*H*L	.6	.03	.34	.32	2.2	.15	.03	.07	.31	.67
S*P*H*L	.85	.008	.08	3.2^	.5	.005	.58	.02	.1	.62
S*H*AP	.27	.159	.34	.2	.35	2.5	2.9*	.87	.71	.48
P*H*AP	.6	.24	.2	.36	.51	2.2	2.4^	.61	1.3	3.9*
S*P*H*AP	.35	.16	.65	.19	.03	1.6	.15	.43	2.1	.38
S*L*AP	4.3*	.64	2.3	2.2	1.9	3.7*	1.3	1.2	.3	2.5^
P*L*AP	4.6*	.91	.67	1.3	1.6	1.3	.82	1.5	4.9*	2.5^
S*P*L*AP	2.2	1.4	3.0*	.44	3.3*	1.9	2.9*	.78	1.1	4.3*
S*H*L*AP	.9	2.2	.78	1.1	.99	1.7	.11	1.2	.41	.88
P*H*L*AP	1.5	1.1	1.4	1.8	2.5^	2.4^	.9	1.2	1.2	2.17
S*P*H*L*AP	3.9*	1.7	1.4	2.5^	1.9	.29	.24	.59	2.1	1.3

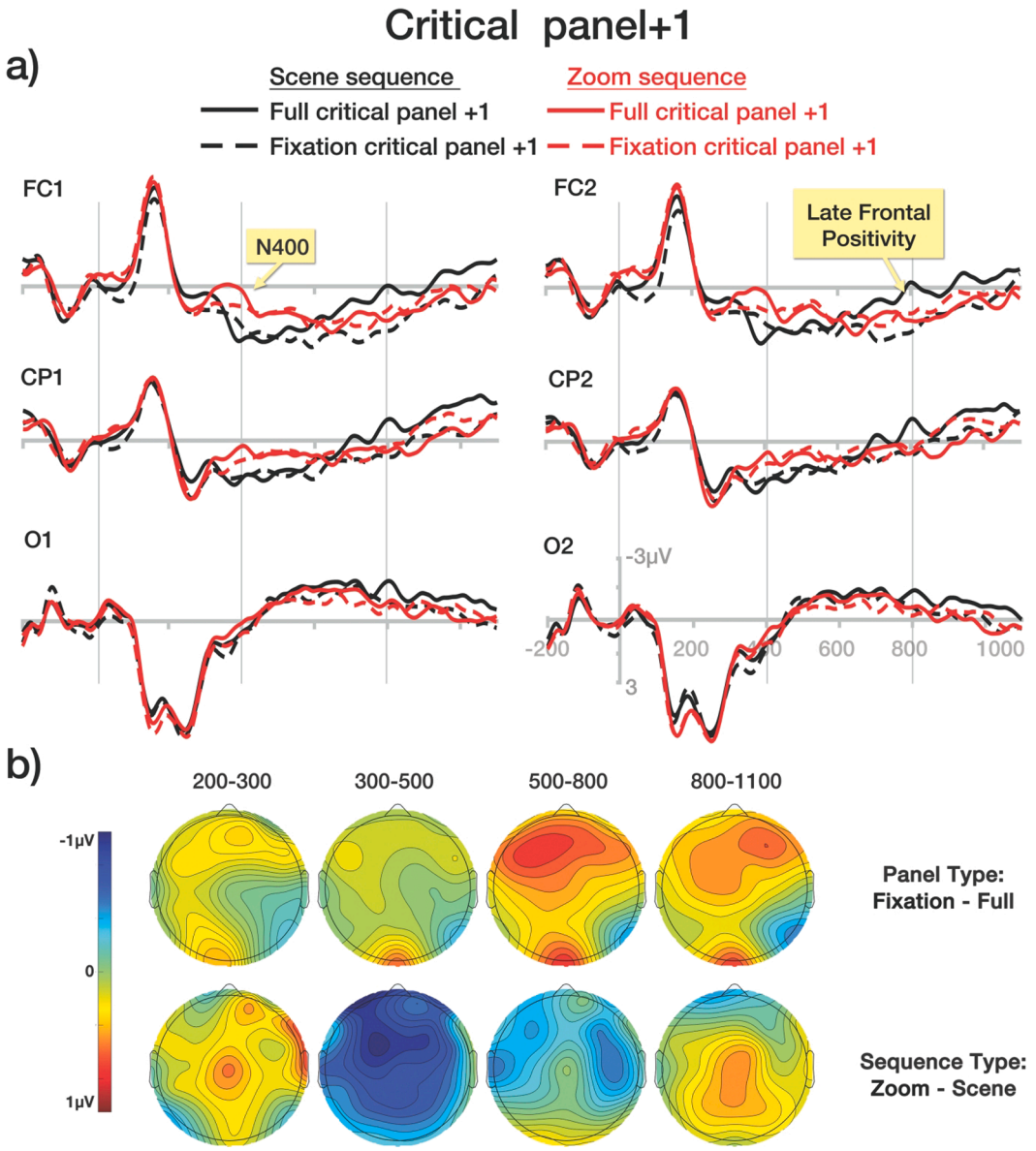

Fig. 6. a) Event-related potentials to panels after the critical full-scene and zoom panels placed within sequences with either full-scene or zoom panels, and b) topographic maps depicting the distribution of effects for panel types (fixation – full) or sequence types (zoom –scene).

between Sequence Type × Hemisphere × AP Distribution and Sequence Type × Panel Type × Laterality × AP Distribution. These interactions again suggested a prefrontal negativity (N300) to panels in scene sequences compared to zoom sequences. As in the preceding critical panel, this showed that full panels (scene sequences) evinced a greater cost of object identification (Draschkow et al., 2018; Hamm et al., 2002) or structural feature mapping (Schendan & Kutas, 2003) than fixation panels (zoom sequences). In the 300–500 ms epoch, a main effect of Sequence Type and a Sequence Type × Laterality interaction suggested a widespread central negativity (N400) that was greater to panels following critical panels in zoom sequences than those in scene sequences. This again suggested that, despite the larger N300s, full panels (scene sequences) evoked smaller N400s than fixation panels (zoom sequences).

An additional frontally distributed positivity emerged for panels following critical fixation panels, compared to those following critical full panels. This was first suggested by interactions of Panel Type with Laterality and with Laterality × AP Distribution in the 500–800 ms epoch. It continued in the 800–1100 ms epoch with interactions between Panel Type × Hemisphere × AP Distribution. This late frontal positivity implied that panels following a fixation panel—whether a full or fixation panel—may have been deemed as less probable or likely than those following full panels (Federmeier, Wlotko, De Ochoa-Dewald, & Kutas, 2007; Leckey & Federmeier, 2020; Van Petten & Luka, 2012). An additional Sequence Type × Panel Type × Laterality × AP Distribution interaction implied that, beyond this frontal positivity, an additional centro-parietal positivity appeared to panels following critical panels of all types other than those to full panels in scene sequences. Like the earlier P600 at the critical panel, this positivity may suggest that further revision or updating processes are required of panels in zoom sequences (here, both fixation panels) than those in scene sequences.

3.4. Ordinal sequence position

To examine the differences in ERPs across each panel of the sequence in non-critical panels, we averaged the amplitudes of electrodes across the anterior and posterior regions of the scalp. Statistics are provided in Table 2. As depicted in Fig. 7, in the 300–500 ms epoch a main effect of

Table 2
F-values from ANOVAs for ERP amplitudes of non-critical panels in scene versus zoom sequences across the ordinal position of the sequence. Sequence Type (ST), Position (P). degrees of freedom: ST = 1,24, P and ST*P = 5,120. ^p < .1, *p < .05, **p < .01, ***p < .001.

	200–300 ms		300–500 ms		500–800 ms		800–1100 ms	
	Anterior	Posterior	Anterior	Posterior	Anterior	Posterior	Anterior	Posterior
ST	0.056	0.95	4.5*	1.1	0.82	1.3	0.79	4.7*
P	4.5**	2.9*	3.6**	10.0***	0.61	2.3^	0.47	9.7***
ST*P	1.7	3.1*	0.32	3.5**	1.6	0.61	2.2^	0.42

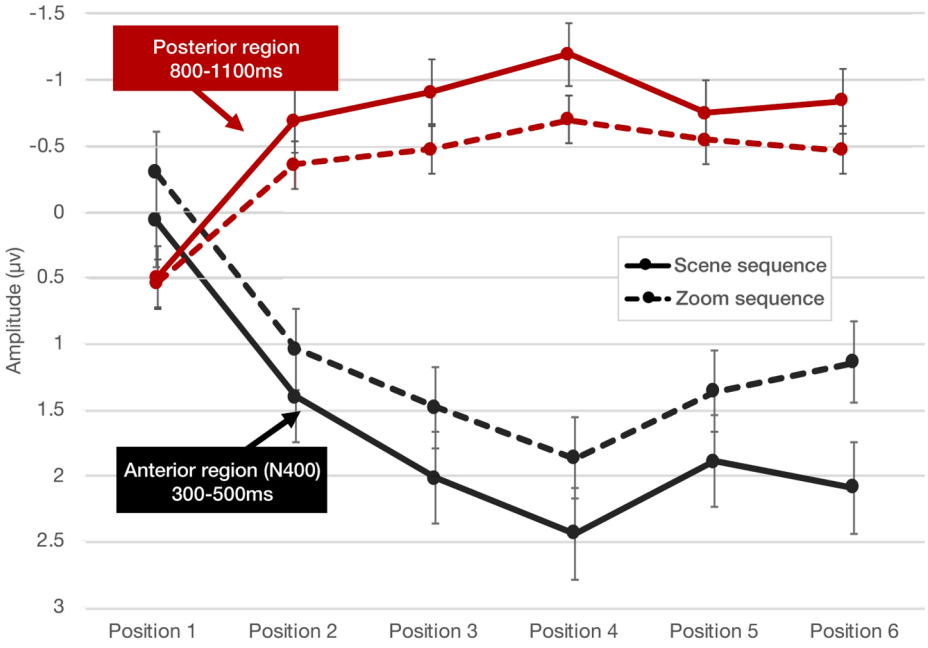

Fig. 7. Amplitudes at each ordinal position in the anterior region of the 300–500 ms epoch (N400) and in the posterior of the 800–1100 ms epoch (late positivity) for non-critical panels in full-scene and zoomed sequences.

Sequence Type appeared in the anterior region (Table 2) suggesting that the panels in zoom sequences elicited a larger amplitude negativity (N400) than the scene sequences. A main effect of Sequence Type in the posterior region of the 800–1100 ms epoch suggested a larger positivity to panels in zoom sequences than scene sequences. As at the critical panels, this N400 implied a greater cost for semantic processing of fixation panels (zoom sequences) than full panels (scene sequences), but sequences with only fixation panels evinced a greater need for subsequent updating or revision processes. Main effects of Position appeared in all epochs except both regions for the 500–800 ms epoch and in the anterior region for the 800–1100 ms epoch. Interactions between Sequence Type and Position appeared only in the posterior regions for the 200–300 and 300–500 ms epochs.

Follow up polynomial contrasts showed linear trends for position in the posterior region of the 300–500 ms and 800–1100 ms epochs (all Fs > 11.9, all ps < 0.005), and sequence type by position interactions appeared in the posterior regions of both the 200–300 ms and 300–500 ms epochs (all Fs > 74, all ps < 0.05). A quadratic trend was found in the anterior region of the 300–500 ms epoch, $F(1,24) = 6.3$, $p < .05$. Post-hoc pairwise comparisons suggested that these trends arose because the first position differed in amplitude from the other positions. The first position differed from all but the fourth position in the 300–500 ms epoch for the posterior region (all ps < 0.05), and from all other positions in the 800–1100 ms epoch for the posterior region (all ps < 0.05). Overall, amplitudes became more positive across position for the 200–300 ms and 300–500 ms epochs, and more negative across position in the 800–1100 ms epochs.

4. Discussion

This study examined the neurocognitive processing of visual narratives with framing of content showing either full scenes or only the most fixated information. We crossed these framing types for each panel in a sequence and/or those at specific panel positions. At the critical panel, we found larger negativities (N300s) to full panels than fixation panels, while later posterior positivities (P600s) were evoked by panels in zoom sequences compared to scene sequences. At the subsequent panel, larger N400s appeared to panels in zoom sequences than scene sequences, while a larger late frontal positivity (LFP) appeared to panels after critical fixation panels than after full panels. Finally, across sequence position, we observed larger N400s to panels in zoom sequences than scene sequences, though both were attenuated across ordinal position in the sequence. Similarly, greater posterior positivities were observed to each panel in a sequence after the first position in zoom sequences than scene sequences.

At the critical panel, we first observed an attenuated negativity to fixation panels compared to full panels. This negativity persisted throughout the 200–300 ms epoch, waning between 300 and 500 ms. This time course was suggestive of an N300, indexing processes posited to be related to object categorization or identification (Draschkow et al., 2018; Hamm et al., 2002) or structural feature mapping (Schendan & Kutas, 2003), rather than an N400 indexing the more general access of semantic memory (Kutas & Federmeier, 2011). This finding suggests that fixation panels attenuate the process of semantic categorization compared to full panels—presumably because they provide a focal viewpoint on specific visual features.

In contrast, between 300 and 500 ms an N400 was mostly evident to fixation panels compared to full panels in scene sequences, with the reverse relationship observed to those in zoom sequences. Unlike the main effect of panel type observed to the N300, no such main effect occurred for the N400. Thus, even though the content of the panel changed with the framing, the evoked N400 was consistent across different versions of the critical panel. These results are consistent with our prior findings that self-paced viewing times are largely the same for critical full panels and fixation panels, which were both shorter than non-fixated zooms and incongruous fixated zooms taken from other sequences (Foulsham & Cohn, 2020). Together, these negativities suggest that, while zoomed-in content required fewer categorization and identification processes than full-scenes (N300), the direct framing of this crucial information provided mostly similar access to semantic memory as would be needed for a whole scene (N400).

Our observations of an attenuated N300 with a *greater* N400 to fixation panels contrast with previous claims that these components index inseparable parts of semantic processing for visual stimuli (Draschkow et al., 2018), and support the possibility of functionally distinct components (McPherson & Holcomb, 1999; Truman & Mudrik, 2018), or at the least that they allow for contrasting attenuation between onsets and peaks. To our knowledge, this is the first observation of a change from an attenuated N300 effect to a subsequent increased N400 effect (or vice versa). This could be due to the nature of stimuli in prior studies of the N300/N400, which have largely contrasted changes in the viewpoint of scenes (Schendan & Kutas, 2003), semantically related or unrelated images (Hamm et al., 2002; McPherson & Holcomb, 1999), or visual scenes with congruous or incongruous elements (Draschkow et al., 2018; Hamm et al., 2002; Truman & Mudrik, 2018). In these cases, the amount of information to be identified remains fairly uniform, while categorization itself may be coupled with that of semantic memory, which only varies between levels of congruity. Here, congruity remained constant for the overall events being undertaken and who performed them. Thus, framing only the crucial information of a panel allowed for attenuation of categorization (N300), while creating only slight costs to accessing semantic memory (N400).

Following this, a posterior positivity consistent with a P600 was more clearly observed to the contrast between sequences from roughly 300 to 700 ms. This positivity was greater to panels from zoom sequences than those from scene sequences, regardless of the type of panel framing. Insofar as P600s index a backward-looking process of integrating incoming information with the expectancies of a preceding context (Baggio, 2018; Brouwer, Crocker, Venhuizen, & Hoeks, 2016; Kuperberg, 2016) this effect suggests that the content of any panel is harder to integrate with prior information that has been restricted in its framing throughout a sequence. This is consistent with previous work on visual narratives where P600s appeared to changes between characters and/or events across panels (Cohn & Kutas, 2015, 2017). However, here the framing changes do not shift the type of situational content mapping across panels, but rather the *amount* of information available for constructing a situation model. Thus, fixation panels constrain the information by which to build a situation model, thereby leading to greater costs of mapping incoming information to that content—regardless of the incoming panel's framing.

To summarize so far, a panel with fixated content appears to provide an easier access to object identification (N300) than framing of the full content, but with some costs to semantic access (N400). However, a panel of any framing type within the context of a zoom sequence may provide more difficulty for integrating this information together (P600).

Following this critical panel, we again observed an attenuated N300 to fixation panels compared to full panels in the 200–300 ms epoch, but an even larger fronto-central negativity in the 300–500 ms epoch consistent with the N400 (Kutas & Federmeier, 2011). This negativity was larger to fixation panels (zoom sequence) than full panels (scene sequence), regardless of the framing in the critical panel preceding it. At the subsequent 500–700 ms epoch, we again observed a posterior positivity consistent with the P600, though more constrained in its scalp distribution. Here, panels following critical fixation panels evoked a larger positivity than those following full panels. This suggested that, like the P600s at the critical panel, panels that follow a constrained framing require a greater updating of a situation model than those following a full-scene. In both cases, such positivities occurred regardless of the framing of the panel itself. This suggests that such processes are not simply about the bottom-up information provided by the incoming panel, as in the N300 or N400 effects, but about the integration of that information into a prior context (Baggio, 2018; Brouwer et al., 2016; Kuperberg, 2016).

In the 800–1100 ms epoch, panels following critical fixation panels also evoked a late frontal positivity (LFP) compared to those following critical full panels, regardless of sequence framing. A similar LFP also occurred in critical panels following fixation panels compared to those following full panels. This LFP appeared between 500 and 1100 ms across frontal regions of the scalp, consistent with late frontal positivities appearing in sentence processing (Van Petten & Luka, 2012). In these sentence contexts, LFPs typically appear to words that are congruent, but unlikely given the context (Federmeier et al., 2007; Leckey & Federmeier, 2020; Van Petten & Luka, 2012). In visual narratives, frontal positivities have appeared to both congruent and incongruent contexts, such as panels with motion lines that have been reversed to depict a path that is "backwards" from the expected direction (Cohn & Maher, 2015), and to unexpected character changes across panels (Cohn & Kutas, 2017). They have also appeared to descriptive words (*Punch!*) substituted for climactic actions compared to onomatopoeic (*Pow!*) substitutions (Manfredi, Cohn, & Kutas, 2017), where such descriptive words have a lower frequency of appearance in comics (Pratha, Avunjian, & Cohn, 2016). In this context, the LFP to any panel following a fixation panel perhaps supports them as being a low likelihood, regardless of their framing.

Some research has connected the LFP in language to the more general processes of the P300 (Leckey & Federmeier, 2020; Van Petten & Luka, 2012), which in a frontal distribution (P3a) has been associated with demands of attentional processing (Polich, 2007). As our fixation panels are direct representations of fixated information, they represent the attentional focus of prior participants. In visual narratives, panels have been described as "attention units," with variation in framing providing a possible "simulation of attention" for a reader's eyes on a scene (Cohn, 2013). In that we observed an LFP to any panel following a zoom, it could reflect the "simulated" attentional demands on this framed content.

Another interpretation is worth considering though. The frontal distribution of this effect makes it possible that this deflection is not a late frontal positivity to fixation panels, but rather a sustained frontal *negativity* to full panels relative to those zooms. Indeed, late sustained anterior negativities (SAN/Nref) have been interpreted as reflecting a stage of interpretive semantic processing (Baggio, 2018), such as maintaining referential entities in a mental model or making connections between anaphoric relations in a discourse (Baggio, 2018; van Berkum, 2012). It is possible that this effect is similar, demanding greater processing of referential information to any panel following a full-scene relative to a fixation panel, because the prior full panels will activate more referential information than fixation panels, which will then be reactivated in subsequent panels. Clarity for whether this later frontal effect is a positivity or negativity can be further provided by future studies.

This issue of changing viewpoints in the scene raises interesting questions about the degree to which comprehenders make expectations about the types of framing involved in a sequence. While participants may have habituated to the potential for zoomed-in panels in the context of this experiment, such framing is not typical for *Peanuts* strips, which were our source stimuli. Furthermore, corpus analyses have shown that the proportion of framing types differ across cultures, with Asian comics typically using more zoomed-in panels than comics from the United

States or Europe (Cohn, 2013, 2020a). On this basis, we ran exploratory Pearson's correlations between ERPs averaged across electrodes on the scalp for the primary factors (Sequence Type, Panel Type) in each epoch with participants' self-assessed ratings of how often they read Japanese manga currently and while growing up (from the VLFI questionnaire). At the critical panel +1, current manga reading negatively correlated with effects of Panel Type (fixation minus full, collapsed across sequence type) in both the 200–300 ms epoch, ($r = -0.48$, $p < .05$) and 300–500 ms epoch ($r = -0.54$, $p < .01$). These correlations thus suggested N300 and N400 effects were smaller between panels following full and fixation panels with greater manga reading frequency. While such exploratory findings should be taken cautiously, they may suggest varying processing strategies on the basis of exposure, as has also been observed to narrative patterns (Cohn & Kutas, 2017). Further exploration of the processing of such framing variation in manga directly—or for readers of these other types of comics—would be an interesting cross-cultural extension of these comparisons.

Finally, in addition to the ERPs at or after our manipulated critical panels, we also analyzed the neural response at each ordinal position between our two sequence types. Consistent with the N400 at the subsequent-to-critical panel, panels in zoom sequences evoked consistently larger N400s than scene sequences across each ordinal panel position. These findings again imply that panels using a constrained view of the scene made the meaning more difficult to access. This N400 effect remained consistent across ordinal sequence positions, suggesting the bottom-up cost for fixation panels relative to full panels remains constant, despite the attenuation across sequence position. This decreasing amplitude of the N400 across each ordinal position for both sequence types aligns with prior observations of attenuated N400s across positions for coherent narrative sequences, but not for sequences lacking the combination of narrative structure and semantic associations (Cohn et al., 2012). That the N400s to fixation panels across ordinal position were still attenuated, despite being larger than those to full panels, further supports their congruity as a visual narrative sequence even with the constrained framing.

It is also noteworthy that the greater amplitude N400s to fixation panels appeared even at the first panel of the sequence. For both sequence types, this first position had larger N400 amplitudes (>1μv) than all other positions, consistent with prior observations of ERPs for coherent narrative sequences (Cohn et al., 2012), and with slower viewing times at the first position of visual narratives (Cohn & Wittenberg, 2015; Foulsham et al., 2016), including for fixation zoom panels (Foulsham & Cohn, 2020). This cost at the start of the sequence represents a process of "laying a foundation" of information for the subsequent discourse (Cohn & Paczynski, 2013; Gernsbacher, 1990; Loschky et al., 2020), and it occurs also across verbal discourse and even at the sentence level (Gernsbacher, 1990; Haberlandt, 1980; Van Petten & Kutas, 1991). Some have speculated that such costs in viewing times for visual narratives are motivated by perceptual visual search processes (Loschky et al., 2018). Our findings here suggest that laying a foundation is not necessarily motivated by attentional processes, as the cost is clearly visible in "back end" semantic processing, where semantic access is more strained when panels depict only the constrained fixation information. However, like the scene sequences, the N400 attenuation for zoom sequences across positions suggests a benefit from the prior context, even with the constrained viewpoints.

In addition to the change in N400 amplitude, we also observed larger posterior positivities to fixation panels than full panels, sustained across each sequence position. While this epoch is somewhat late for the typical P600, the polarity (positive) and distribution (posterior) remain consistent, possibly suggesting a similar, albeit delayed, response. If so, this suggests that updating processes are ongoing throughout a narrative sequence (Cohn, 2020b; Loschky et al., 2020; Zwaan & Radvansky, 1998). Insofar as this P600 interacts with the preceding N400, it is consistent with models positing connections between retrieval (N400) and integration (P600) processes indexed by these ERP components (Baggio, 2018; Brouwer et al., 2016; Cohn, 2020b; Tanner, Goldshtein, & Weissman, 2018).

Nevertheless, this positivity did not differ between fixation and full panels at the first panel of the sequence. At the second panel, zoom panels became substantially more positive in amplitude, and then this difference in amplitudes maintained across the sequence. This suggests that fixation and full panels engage a similar updating process at the start of a sequence, but zooms require more updating as a sequence progresses. This is unlike the N400, which differed even at the first position, and then became more attenuated across the sequence. Thus, under this interpretation, updating involves a fairly uniform, ongoing process, perhaps here reflecting the fairly uniform character of these sequences (i.e., either full-scene or zoom panels across the whole sequence).

Taken together, our findings here suggest that constrained framing within a visual narrative sequence facilitates identification processes (N300), while demanding slightly greater access to semantic memory (N400). A sequence of such constrained views may require additional updating costs (P600), while such changes in framing may evoke additional demands (LFP/SAN). The evocation of these components in sequences that remain fairly congruous, manipulating only the density of presented information, implies that such processes are active throughout comprehension of all visual narratives. Finally, the consistency of these brain responses to visual narratives with those evoked in other domains like visual scenes and language reinforce the similarity (or connections) between comprehension systems across domains.

Funding sources

This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

CRediT authorship contribution statement

Neil Cohn: Conceptualization, Methodology, Formal analysis, Investigation, Writing - original draft. **Tom Foulsham:** Conceptualization, Methodology, Writing - review & editing.

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*. https://doi.org/10.3389/fnhum.2013.00057.

Baggio, G. (2018). *Meaning in the Brain*. Cambridge, MA: MIT Press.

Brouwer, H., Crocker, M. W., Venhuizen, N. J., & Hoeks, J. C. J. (2016). A Neurocomputational Model of the N400 and the P600 in Language Processing. *Cognitive Science, 41*(S6), 1318–1352. https://doi.org/10.1111/cogs.12461.

Coderre, E. L., Cohn, N., Slipher, S. K., Chernenok, M., Ledoux, K., & Gordon, B. (2018). Visual and linguistic narrative comprehension in autism spectrum disorders: Neural evidence for modality-independent impairments. *Brain and Language, 186*, 44–59.

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

Cohn, N. (2020a). *Who understands comics? Questioning the universality of visual language comprehension*. London: Bloomsbury.

Cohn, N. (2020b). Your brain on comics: A cognitive model of visual narrative comprehension. *Topics in Cognitive Science, 12*(1), 352–386. https://doi.org/10.1111/tops.12421.

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. https://doi.org/10.1016/j.neuropsychologia.2014.09.018.

Cohn, N., & Kutas, M. (2015). Getting a cue before getting a clue: Event-related potentials to inference in visual narrative comprehension. *Neuropsychologia, 77*, 267–278. https://doi.org/10.1016/j.neuropsychologia.2015.08.026.

Cohn, N., & Kutas, M. (2017). What's your neural function, visual narrative conjunction? Grammar, meaning, and fluency in sequential image processing. *Cognitive Research: Principles and Implications, 2*(27), 1–13. https://doi.org/10.1186/s41235-017-0064-5.

Cohn, N., & Magliano, J. P. (2020). Editors' Introduction and Review: Visual Narrative Research: An Emerging Field in Cognitive Science. *Topics in Cognitive Science, 12*(1), 197–223. https://doi.org/10.1111/tops.12473.

Cohn, N., & Maher, S. (2015). The notion of the motion: The neurocognition of motion lines in visual narratives. *Brain Research, 1601*, 73–84. https://doi.org/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. https://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. https://doi.org/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, 27*(7), 812–828. https://doi.org/10.1080/20445911.2015.1051535.

Draschkow, D., Heikel, E., Võ, M. L. H., Fiebach, C. J., & Sassenhagen, J. (2018). No evidence from MVPA for different processes underlying the N300 and N400 incongruity effects in object-scene processing. *Neuropsychologia, 120*, 9–17. https://doi.org/10.1016/j.neuropsychologia.2018.09.016.

Federmeier, K. D., Wlotko, E. W., De Ochoa-Dewald, E., & Kutas, M. (2007). Multiple effects of sentential constraint on word processing. *Brain Research, 1146*, 75–84. https://doi.org/10.1016/j.brainres.2006.06.101.

Foulsham, T., & Cohn, N. (2020). Zooming in on visual narrative comprehension. *Memory & Cognition*. https://doi.org/10.3758/s13421-020-01101-w.

Foulsham, T., Wybrow, D., & Cohn, N. (2016). Reading without words: Eye movements in the comprehension of comic strips. *Applied Cognitive Psychology, 30*, 566–579. https://doi.org/10.1002/acp.3229.

Gernsbacher, M. A. (1990). *Language Comprehension as Structure Building*. Hillsdale, NJ: Lawrence Earlbaum.

Haberlandt, K. (1980). Story grammar and reading time of story constituents. Retrieved from *Poetics, 9*(1–3), 99–118 http://www.sciencedirect.com/science/article/B6VC3-469742G-F/2/8866d40fbf60e04950af36b4b0015851.

Hamm, J. P., Johnson, B. W., & Kirk, I. J. (2002). Comparison of the N300 and N400 ERPs to picture stimuli in congruent and incongruent contexts. *Clinical Neurophysiology, 113*(8), 1339–1350. https://doi.org/10.1016/S1388-2457(02)00161-X.

Huff, M., Meitz, T. G. K., & Papenmeier, F. (2014). Changes in situation models modulate processes of event perception in audiovisual narratives. *Journal of Experimental Psychology: Learning, Memory, & Cognition, 40*(5), 1377–1388. https://doi.org/10.1037/0033-2909.133.2.273.

Hutson, J. P., Magliano, J., & Loschky, L. C. (2018). Understanding moment-to-moment processing of visual narratives. *Cognitive Science, 42*(8), 2999–3033. https://doi.org/10.1111/cogs.12699.

Kuperberg, G. R. (2016). Separate streams or probabilistic inference? What the N400 can tell us about the comprehension of events. *Language, Cognition and Neuroscience, 31*(5), 602–616. https://doi.org/10.1080/23273798.2015.1130233.

Kutas, M., & Federmeier, K. D. (2011). Thirty years and counting: Finding meaning in the N400 component of the Event-Related Brain Potential (ERP). Retrieved from *Annual Review of Psychology, 62*(1), 621–647 http://www.annualreviews.org/doi/abs/10.1146/annurev.psych.093008.131123.

Laubrock, J., Hohenstein, S., & Kümmerer, M. (2018). Attention to comics: Cognitive processing during the reading of graphic literature. In A. Dunst, J. Laubrock, & J. Wildfeuer (Eds.), *Empirical Comics Research: Digital, Multimodal, and Cognitive Methods* (pp. 239–263). New York: Routledge.

Lauer, T., Cornelissen, T. H., Draschkow, D., Willenbockel, V., & Võ, M.-L.-H. (2018). The role of scene summary statistics in object recognition. *Scientific Reports, 8*(14666), 1–12.

Leckey, M., & Federmeier, K. D. (2019). The P3b and P600(s): Positive contributions to language comprehension. *Psychophysiology*, e13351. https://doi.org/10.1111/psyp.13351.

Lopez-Calderon, J., & Luck, S. J. (2014). ERPLAB: An open-source toolbox for the analysis of event-related potentials. Retrieved from *Frontiers in Human Neuroscience, 8*, 213 https://www.frontiersin.org/article/10.3389/fnhum.2014.00213.

Loschky, L. C., Hutson, J. P., Smith, M. E., Smith, T. J., & Magliano, J. (2018). Viewing static visual narratives through the lens of the scene perception and event comprehension theory (SPECT). In A. Dunst, J. Laubrock, & J. Wildfeuer (Eds.), *Empirical Comics Research: Digital, Multimodal, and Cognitive Methods* (pp. 217–238). London: Routledge.

Loschky, L. C., Magliano, J., Larson, A. M., & Smith, T. J. (2020). The scene perception & event comprehension theory (SPECT) applied to visual narratives. *Topics in Cognitive Science, 12*(1), 311–351. https://doi.org/10.1111/tops.12455.

Loschky, L. C., McConkie, G., Yang, J., & Miller, M. (2005). The limits of visual resolution in natural scene viewing. *Visual Cognition, 12*(6), 1057–1092. https://doi.org/10.1080/13506280444000652.

Magliano, J. P., Loschky, L. C., Clinton, J. A., & Larson, A. M. (2013). Is reading the same as viewing? An Exploration of the similarities and differences between processing text- and visually based narratives. In B. Miller, L. Cutting, & P. McCardle (Eds.), *Unraveling the Behavioral, Neurobiological, and Genetic Components of Reading Comprehension* (pp. 78–90). Baltimore, MD: Brookes Publishing Co.

Manfredi, M., Cohn, N., & Kutas, M. (2017). When a hit sounds like a kiss: An electrophysiological exploration of semantic processing in visual narrative. *Brain and Language, 169*, 28–38. https://doi.org/10.1016/j.bandl.2017.02.001.

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.

Meade, G., Lee, B., Midgley, K. J., Holcomb, P. J., & Emmorey, K. (2018). Phonological and semantic priming in American Sign Language: N300 and N400 effects. *Language, Cognition and Neuroscience, 33*(9), 1092–1106. https://doi.org/10.1080/23273798.2018.1446543.

Metusalem, R., Kutas, M., Urbach, T. P., Hare, M., McRae, K., & Elman, J. L. (2012). Generalized event knowledge activation during online sentence comprehension. *Journal of Memory and Language, 66*(4), 545–567. https://doi.org/10.1016/j.jml.2012.01.001.

Mudrik, L., Lamy, D., & Deouell, L. Y. (2010). ERP evidence for context congruity effects during simultaneous object-scene processing. *Neuropsychologia, 48*(2), 507–517. https://doi.org/10.1016/j.neuropsychologia.2009.10.011.

Peirce, J., Gray, J. R., Simpson, S., MacAskill, M., Höchenberger, R., Sogo, H., … Lindeløv, J. K. (2019). PsychoPy2: Experiments in behavior made easy. *Behavior Research Methods, 51*(1), 195–203. https://doi.org/10.3758/s13428-018-01193-y.

Polich, J. (2007). Updating P300: An integrative theory of P3a and P3b. *Clinical Neurophysiology, 118*(10), 2128–2148.

Pratha, N. K., Avunjian, N., & Cohn, N. (2016). Pow, punch, pika, and chu: The structure of sound effects in genres of American comics and Japanese manga. *Multimodal Communication, 5*(2), 93–109.

Schendan, H. E., & Kutas, M. (2003). Time course of processes and representations supporting visual object identification and memory. *Journal of Cognitive Neuroscience, 15*(1), 111–135. https://doi.org/10.1162/089892903321107864.

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

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

Tanner, D., Goldshtein, M., & Weissman, B. (2018). Individual differences in the real-time neural dynamics of language comprehension. In K. D. Federmeier, & D. G. Watson (Eds.), *Psychology of learning and motivation* (Vol. 68, pp. 299–335). Academic Press.

Truman, A., & Mudrik, L. (2018). Are incongruent objects harder to identify? The functional significance of the N300 component. *Neuropsychologia, 117*, 222–232. https://doi.org/10.1016/j.neuropsychologia.2018.06.004.

van Berkum, J. J. A. (2012). The electrophysiology of discourse and conversation. In M. J. Spivey, M. Joanisse, & K. McRae (Eds.), *The Cambridge handbook of psycholinguistics*. Cambridge: Cambridge University Press.

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

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

Van Petten, C., & Luka, B. J. (2012). Prediction during language comprehension: Benefits, costs, and ERP components. *International Journal of Psychophysiology, 83*(2), 176–190. https://doi.org/10.1016/j.ijpsycho.2011.09.015.

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.