

Action starring narratives and events: Structure and inference in visual narrative comprehension

Neil Cohn^{1,2} & Eva Wittenberg³

¹Department of Psychology, Tufts University, Medford, MA.

²Department of Cognitive Science, UC San Diego, La Jolla, CA

³Department of Linguistics, UC San Diego, La Jolla, CA

Correspondence:

neilcohn@visuallanguagelab.com

www.visuallanguagelab.com

Uncorrected accepted proof copy. For full publication:

Cohn, Neil, and Eva Wittenberg. 2015. "Action starring narratives and events: Structure and inference in visual narrative comprehension." *Journal of Cognitive Psychology* 27 (7):812-828. doi: 10.1080/20445911.2015.1051535.

Abstract

Studies of discourse have long placed focus on the inference generated by information that is not overtly expressed, and theories of visual narrative comprehension similarly focused on the inference generated between juxtaposed panels. Within the visual language of comics, star-shaped “flashes” commonly signify impacts, but can be enlarged to the size of a whole panel that can omit all other representational information. These “action star” panels depict a narrative culmination (a “Peak”), but have content which readers must infer, thereby posing a challenge to theories of inference generation in visual narratives that focus only on the semantic changes between juxtaposed images. This paper shows that action stars demand more inference than depicted events, and that they are more coherent in narrative sequences than scrambled sequences (Experiment 1). In addition, action stars play a felicitous narrative role in the sequence (Experiment 2). Together, these results suggest that visual narratives use conventionalized depictions that demand the generation of inferences while retaining narrative coherence of a visual sequence.

Keywords: discourse; narrative; inference; comics; visual language

Introduction

How do we make sense of information in a narrative that is not overtly provided? The generation of inferences—the information that a reader understands despite being unstated in a discourse—has long been a primary focus in the study of discourse (Graesser, Millis, & Zwaan, 1997; Keenan, Potts, Golding, & Jennings, 1990; van den Broek, 1994; Zwaan & Rapp, 2006). Because inferences allow a reader to make sense of unexpressed material, they contribute towards building a “situation model” of the discourse in memory (van Dijk & Kintsch, 1983). This emphasis on inference generation has also been a hallmark of film theory (Bordwell & Thompson, 1997; Eisenstein, 1942; Kuleshov, 1974), and studies of film comprehension support that viewers are consciously able to identify changes in time, characters, and spatial locations (Magliano, Miller, & Zwaan, 2001; Magliano & Zacks, 2011; Zacks, Speer, & Reynolds, 2009).

Theories of visual narrative comprehension have also emphasized inference (Bordwell, 1985, 2007; Branigan, 1992; Chatman, 1978; Eisenstein, 1942; Magliano, Dijkstra, & Zwaan, 1996; McCloud, 1993; Saraceni, 2001; Yus, 2008), especially the bridging inferences where readers “fill in” the information left unstated between “panels”—the encapsulated image units of a static visual narrative sequence. Similar to the linear coherence relationships between sentences (Halliday & Hasan, 1976; Hobbs, 1985; Kehler, 2002; Mann & Thompson, 1987; Zwaan & Radvansky, 1998), theories of visual narratives emphasize the linear semantic changes between panels across dimensions of time, causation, characters, environments, and scenes (McCloud, 1993; Saraceni, 2000, 2001; Stainbrook, 2003). More inference is hypothesized to be demanded by greater discontinuities between panels, such as when incoming panels do not repeat information in prior panels or do not share elements related to a broader semantic field (Saraceni, 2000, 2001).

To create these bridging inferences, two panels must provide bottom-up content for a reader to infer the link between them. However, some panels in the visual language of comics have such impoverished semantic content that inference is necessary to understand what they mean, let alone how they unite with other information. Consider Figure 1a, where a dog curiously chases a ball until unexpectedly getting scared off by people playing soccer: We never actually see the players interacting with the dog, though we know this event occurs in panel 3. This panel only depicts an “action star,” a “visual morpheme” used to represent impacts (Cohn, 2013a; Potsch & Williams, 2012; Walker, 1980). In panel 3, which only shows the action star, the events are inferred via the preceding and final panels. This inference does not occur between panels 2 and 3 or between panels 3 and 4. Rather, inference is necessary to comprehend the events omitted *within* the action star itself, not just to understand the relations between panels.

Action starring narrative and events

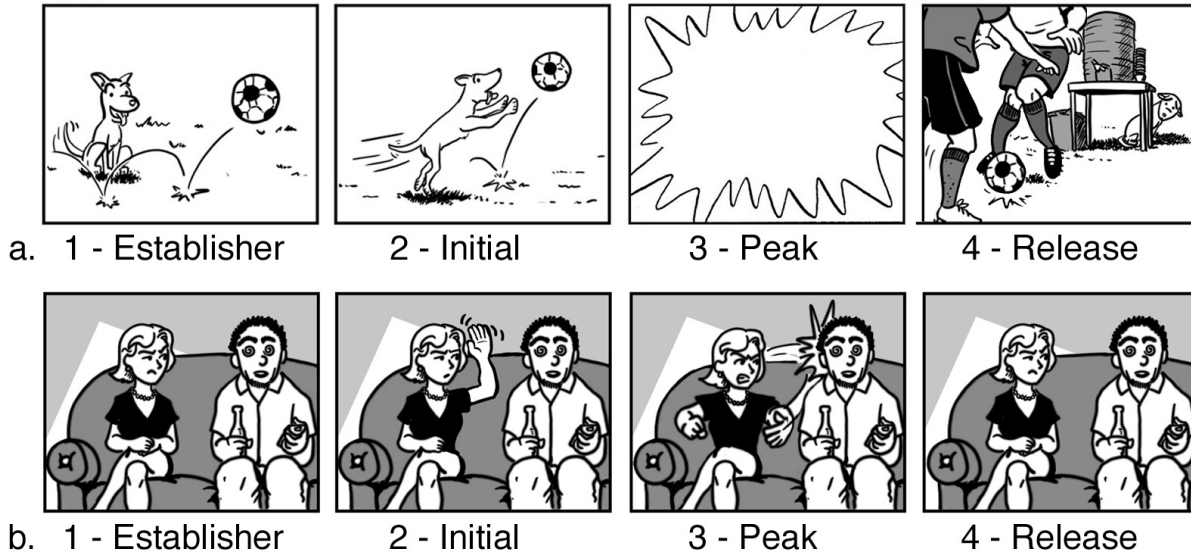


Figure 1. Simple visual narrative sequences. (a) shows a sequence where an “action star” omits crucial event information in the culminating Peak of the sequence, while (b) shows a sequence with a canonical mapping of event structure to narrative structure.

The structure of visual narratives

Action stars are notable not only because they require inferences, but because they seem to play a narrative function in visual sequences. We can understand this role by drawing on the theory of *Visual Narrative Grammar* (VNG), which posits that individual panels play categorical roles in a narrative sequence, which then become structured into hierarchic constituents (Cohn, 2013b) analogous to the way that words play categorical roles in the hierarchic structure of sentences. This comparison between syntax and visual narrative is one of function—the units of sentences (words) and visual narratives (images) convey information in different ways and levels of meaning. Functionally though, a narrative grammar packages meaning into a sequence using similar architectural constraints (categories, hierarchy, etc.) as how syntax packages meaning in sentences, only operating at a discourse level of information. While VNG has some similarities with previous “grammatical” approaches to narrative (e.g., Mandler & Johnson, 1977; Rumelhart, 1975; Stein & Glenn, 1979; Thorndyke, 1977), VNG uses simpler structures (Cohn, 2013b), makes an explicit separation of structure and meaning (Cohn, Paczynski, Jackendoff, Holcomb, & Kuperberg, 2012), and incorporates modifiers beyond a canonical narrative arc (Cohn, 2013a, 2013b). Also, VNG is not incompatible with most models of discourse, which tend to focus on semantic aspects of comprehension like coherence relationships and inference generation (for review, see McNamara & Magliano, 2009), while VNG outlines the “grammatical” relationships that interface with those semantic processes. For example, although VNG extends beyond linear coherence relations (e.g., McCloud, 1993; Zwaan & Radvansky, 1998), such semantic changes should interface with the narrative grammar in predictable ways, like linear coherence relationships correlating with breaks between constituents (Cohn, 2013b).

Narrative categories in VNG are assigned through an interaction of the bottom-up semantic content of panels and their top-down context in the broader narrative (Cohn, 2013b, 2014). Consider Figure 1b, which progresses in a canonical narrative arc. An “Establisher” opens

the sequence with a woman sitting angrily next to a man, which functions to set up the characters and situations of a sequence without acting upon them. Next, an Initial begins the events of the sequence, prototypically with a preparatory action, like the woman reaching back to smack the man. A sequence climaxes at a Peak, where completed events or actions typically occur (like smacking the man). The aftermath occurs in the Release, as in the final panel where the man humorously is not affected by the woman's actions. Figure 1b shows a prototypical interface between structure and semantics, where the narrative categories directly correspond to the event structure. In addition, though it will not be dealt with here, these narrative categories apply both to individual panels and to whole constituents, recursively extending to visual narratives of greater lengths (see Cohn, 2013b; Cohn, Jackendoff, Holcomb, & Kuperberg, 2014). Such constituents also allow for surface patterns to violate a canonical arc, though the individual constituents that make up that sequence may not (for example, a sequence with the surface structure *Initial-Peak-Initial-Peak* could be felicitous if segmented [[_II-P]-[_PI-P]]).

Now, reconsider Figure 1a. The Establisher here starts with a lot of action: the soccer ball flies into the frame and the dog is excited by it. Despite not depicting a passive state, this panel still introduces the reader to the characters involved in the situation (reinforced by the characters also meeting each other). The Initial shows the dog chasing the ball, but no preparatory action (the dog is already engaged in this action). The action star does have a culminating Peak, but without depicting any completed actions (a point returned to below), nor is a completed action inferred (the dog's action is interrupted). Finally, the Release shows the aftermath of the (unseen and now inferred) prior event. Thus, prototypical and non-prototypical mappings can occur between narrative structure and meaning, and their assignment involves both bottom-up semantic features and top-down context in a sequence.

Let us now return to action stars, which appear within American comic books and strips to stand in for events, both related and unrelated to impacts. Action stars vary in appearance both between and within authors, sometimes appearing just as a star or sometimes with text that disambiguates the actions (like “Pow!” or “Zap!”). Thus, as a conventionalized aspect of a broader “lexicon” of visual narratives, action stars have allomorphic representations (Cohn, 2013a); for additional examples of action stars, see supplementary material available at http://www.visuallanguagelab.com/A/AS_Supplement.pdf.

Though action stars show minimal event information, thereby demanding inference for their meaning, their morpho-semantics implies a “culminating event”, which provides enough information for them to act as Peaks of the sequence. They thereby provide a way to elide information about events but retain narrative felicity (Cohn, 2013b). This would be *functionally* analogous to a “pro-form” in the syntactic structure of sentences, which plays a grammatical role as a noun (*he, she, it*) or preposition (*there, here, then*), yet provides fairly minimal semantic information. Similarly, action stars act as narrative Peaks, but physically convey only an unspecified event, with no properties about what that event is or who is involved. Granted, visual narratives contain far more information per unit than individual words, but again this analogy between action stars and pro-forms is made purely at the *functional* level related to structure, not the level of conceptualized information (just like the broader analogy between syntax and narrative structure in VNG).

We therefore have two hypotheses about action stars: First, they should require inference to be understood because of their impoverished semantic structure (e.g., McKoon & Ratcliff, 1992), not only for inferences between panels, but to understand action stars themselves. Second, they should play a role in the narrative grammar as Peaks. This is what makes them interesting

and unique: they are an impoverished conventionalized depiction that demands inference, yet maintains a felicitous role in a narrative structure.

Processing of visual narratives

Despite many theories stressing the importance of inference in visual narrative comprehension, thus far no studies have explicitly examined the image-by-image processing of bridging inferences in visual narratives. Yet, some work has shown that accuracy for inferring omitted panels from visual narratives correlates with age and experience reading comics (Nakazawa, 2005). In addition, narrative categories seem to differ in their inferential demands. In a previous study, participants were more accurate at recognizing the ellipsis of Peaks from sequences than other elided categories, and strips with missing Peaks were rated lower than those omitting other categories (Cohn, 2014). Given that Peaks contain the apex of many causal relations, these findings are consistent with research emphasizing that the locus of causal relations in a discourse may be more important than units with more peripheral information (Trabasso, Secco, & van den Broek, 1984; Trabasso & Sperry, 1985).

Additional research using sequential images has begun to construct a view of image-by-image processing of visual narrative sequences. First, comprehenders use both bottom-up content and top-down context to make expectations about subsequent information in a sequence. For example, comprehenders may assume that bottom-up semantic referential information like characters, locations, and/or semantic associative fields will repeat across images (Cohn et al., 2012; Magliano & Zacks, 2011; Saraceni, 2001). Semantic information also involves expectations about events (Reid & Striano, 2008; Sitnikova, Holcomb, & Kuperberg, 2008), like that a completed action will be presumed to follow a preparatory action (Cohn & Paczynski, 2013). Interfacing with this semantic information, comprehenders may also anticipate top-down narrative structural information, such as that a Peak will follow an Initial (Cohn et al., 2014). Disconfirmation of these structural and semantic predictions incurs processing costs both at an unexpected or anomalous image itself (Cohn, 2014; Cohn et al., 2014; Cohn et al., 2012; West & Holcomb, 2002), and potentially at subsequent panels where further context leads to (re)assessment of the prior information (Cohn et al., 2014; Cohn & Paczynski, 2013). However, the coherent combination of both structure and meaning allows for a facilitation of semantic comprehension with each subsequent image in a sequence (Cohn et al., 2012).

Given this framework, an action star would thus satisfy a structural prediction that a Peak would follow a preceding Initial. However, its content would remain ambiguous, and the bottom-up content would only inform that an “event” takes place. While no continuity would be maintained for low-level referential information (e.g., Magliano & Zacks, 2011; Saraceni, 2001), event information may influence inferences about its content given the semantic constraints of the prior panel (McKoon & Ratcliff, 1986, 1992), allowing for at least some sense of causal cohesion (Magliano, Baggett, Johnson, & Graesser, 1993; Singer, Halldorson, Lear, & Andrusiak, 1992; Trabasso et al., 1984). For example, a preparatory action (e.g., the runner going towards the catcher in Figure 2) may allow for the predicted inference of the subsequent action star containing a completion (i.e., the collision) since preparations generate expectations about subsequent actions (Cohn & Paczynski, 2013). No matter the preceding content, full inference may only be totally accessible once the image after the action star is reached, where reanalysis and/or confirmation can be made given the subsequent context of the sequence. Bridging inferences related to action stars must therefore involve information in both the prior and subsequent panels (e.g., Kintsch, 1988, 1998). Thus, while structural felicity may be assessed at

the action star panel itself, evidence of inference should appear at the subsequent panel, where the contents of an action star must be integrated/analyzed in light of additional context.

We investigated these semantic and structural traits of action star comprehension using two experiments in a “self-paced viewing” paradigm (Cohn, 2012, 2014; Cohn & Paczynski, 2013). Experiment 1 compared action stars and normal Peaks in the context of coherent and scrambled narrative sequences, while Experiment 2 compared the comprehension of action stars to other types of panels in coherent sequence frames.

Experiment 1: Scrambling

If action stars play a narrative role that requires the generation of inferences through their relationships with surrounding panels, then such effects should disappear in sequences lacking a coherent narrative grammar, such as when discourse units are rearranged to not make sense. Participants are better able to recall verbal narratives that follow a canonical structure than those where temporal order is changed (Mandler & Johnson, 1977), where sentences are inverted (Mandler, 1978, 1984; Mandler & DeForest, 1979), or where sentences are fully scrambled (Mandler, 1984). Cross-modal comparison of narratives have supported that scrambling the order of discourse units inhibits comprehension across domains, be it in the verbal, written, or visual-graphic modality (Gernsbacher, Varner, & Faust, 1990; Robertson, 2000). Consistent with this, target panels from random sequences of images elicit slower response times than panels from sequences with narrative grammar and/or semantic associations between panels (Cohn et al., 2012). Furthermore, panels in scrambled sequences evoked larger amplitude N400 effects than those in normal sequences with coherent narrative structure (Cohn et al., 2012)—the N400 effect being a neural response elicited by both words (Kutas & Hillyard, 1980) and images (Barrett & Rugg, 1990; Barrett, Rugg, & Perrett, 1988) and modulated by the degree to which the semantic features of an input matches or mismatches with its prior context (Kutas & Federmeier, 2011; Kutas & Hillyard, 1980).

Action stars should demand little inference within scrambled sequences that do not allow for a coherent narrative structure, because relations between panels would lack the narrative and causal information necessary to comprehend sequential events. In Experiment 1, we used a “self-paced viewing” paradigm to explore the hypothesis that action stars would evoke inferences at the subsequent panel by comparing normal Peaks with action stars within both coherent and scrambled visual narratives. Here, participants controlled the pace of viewing each panel in a sequence while we measured how long each panel stayed on the screen. Self-paced viewing paradigms have long been used in the study of inference in verbal discourse comprehension (Haviland & Clark, 1974; Keenan et al., 1990; McKoon & Ratcliff, 1986), and have proven to be a successful technique for measuring comprehension in visual narratives (Cohn, 2012, 2014; Cohn & Paczynski, 2013).

In discourse studies, longer viewing times have typically appeared to sentences that require inference to understand prior information relative to non-inference generating controls (Haviland & Clark, 1974; Keenan et al., 1990; Sanford & Garrod, 1981; van den Broek, 1994). Analogously, if action stars force a reader to infer the unseen content, we reasoned that slower viewing times should appear to the panel following action stars than to corresponding panels following normal Peak panels. However, little or no difference should appear between panels following normal Peaks and action stars in scrambled sequences, because the context would create little demand for inference generation. Additionally, we would expect that action stars

Action starring narrative and events

playing a structural narrative role would be viewed shorter in coherent narratives than scrambled sequences, just as we would expect normal Peaks to be viewed shorter in coherent narratives than in scrambled sequences.

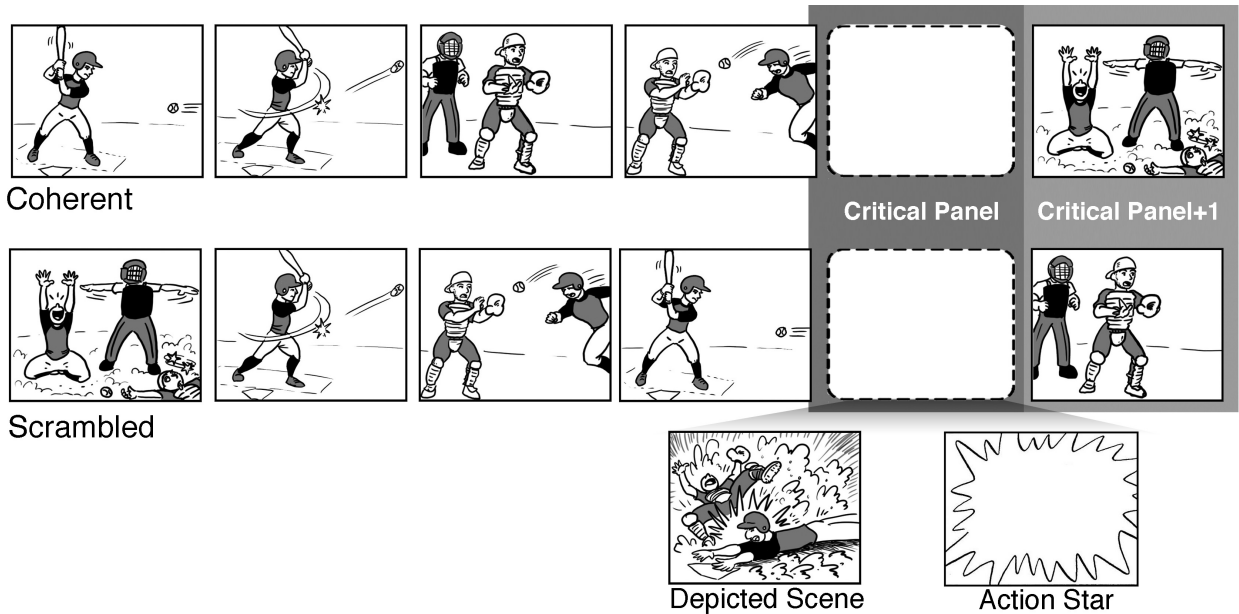


Figure 2. Experimental stimuli contrasting Coherent visual sequences with a felicitous narrative and Scrambled sequences, where the order of panels were rearranged. Within these sequences, critical panels either maintained the original depicted scene of a Peak panel—the culmination of the narrative—or substituted it with an “action star.” Critical panel position was placed anywhere from the third to the sixth position in the sequence.

Methods

Stimuli

We used 60 coherent 6-panel long visual narrative sequences from an existing corpus with coherency confirmed in a prior rating study (Cohn et al., 2012). Sequences maintained panels of a similar size and had no text to eliminate any influence of written language on comprehension. These sequences were then manipulated in two ways.

First, “scrambled” sequences rearranged panels into an incomprehensible order that would contrast with the “coherent” normal sequences. Panels were rearranged such that their order would not create alternate coherent sequences (e.g., moving an Establisher such that it would act as a Release (Cohn, 2014)), particularly by reversing the order of Initials and Peaks, both locally and across constituents, among other rearrangements. Critical Peak panels remained in the same position for both Coherent and Scrambled versions of a given sequence, distributed throughout ordinal sequence positions 3 through 6. Peak panels were able to fall between positions 3 and 6 because these Coherent 6-panel long sequences often consisted of multiple constituents, where Peak panels could vary from the penultimate position.

Second, within these sequence types (Coherent, Scrambled) critical panels used either the original Peak panel of the sequence, or were replaced by an action star. This yielded a 2 (Sequence Type: Coherent/Scrambled) x 2 (Peak Type: Depicted Scene/Action Star) design, as

Action starring narrative and events

in Figure 2. These sequences were divided into four counterbalanced lists such that lists included each strip only once, and no sequence appeared in the same list twice. 30 fillers of coherent sequences were added to balance the number of coherent scenes (45 total: 15 experimental, 30 fillers) with scrambled and action star sequences (45 total: 15 Coherent Action Stars, 15 Scrambled Depicted Scenes, 15 Scrambled Action Stars) viewed by each participant. Fillers also added variability in the length of sequences (fillers ranged from 6 to 12 panels long), such that not all sequences ended after 6 panels. Each list presented sequences in a randomized order.

Procedure

Participants viewed each strip frame-by-frame on a computer screen with a pace under their own control. Viewing times were measured to each button press for how long each frame stayed on the screen. Trials began with a screen reading READY, followed by a fixation cross (+). Each panel then appeared one at a time centered on an otherwise black screen. A 300ms ISI prevented panels from overlapping to appear like a flipbook style animation. A question mark appeared after each sequence, where participants rated how easy the strip was to understand (1=difficult, 7=easy). A practice list with ten stimuli oriented participants to the procedure.

Participants

Twenty-eight comic readers from the Tufts University population (19 male, 9 female, mean age: 21.04) were compensated for their participation in the study. All participants gave their informed written consent according to Tufts University Human Subjects Review Board guidelines.

Previous studies have shown that comprehension of sequential images differs based on comic reading ability (Cohn et al., 2012), including inferences drawn from sequences with omitted information (Nakazawa, 2005), so fluent comic readers were recruited to ensure fluency in this “visual language.” This expertise was assessed using the “Visual Language Fluency Index” (VLFI) questionnaire (Cohn et al., 2012) asking how often participants read various types of comics on a scale of 1 (never) to 7 (always) including comic books, comic strips, graphic novels, and Japanese comics. These ratings assessed both current reading habits as well as when they were growing up. A “VLFI score” was then computed using the following formula:

$$\left(\frac{\text{Mean Comic Reading Freq.}}{x \text{ Comic reading expertise}} \right) + \left(\frac{\text{Comic Drawing Freq.} \times \text{Drawing Ability}}{2} \right)$$

This formula weights fluency towards comic reading comprehension, while giving an additional “bonus” for fluency in comic production. Previous research has shown that the score derived from this metric provides a strong predictor of both behavioral and neurophysiological effects in online comprehension of visual narratives (Cohn & Maher, 2015; Cohn et al., 2012). Within this metric, an idealized average would be a score of 12, with low being below 7 and high above 20. Participants had an “average” fluency, with a mean of 15.13 (SD = 8.48; range = 1.5 - 38.12).

Data Analysis

Outlier viewing times for each participant were discarded if they fell below a threshold of 300ms, or above 8000ms. This lower limit was set below half the fastest mean panel viewing times seen in our previous studies of visual narrative (Cohn, 2012, 2014; Cohn & Paczynski,

2013), while the upper limit was roughly four times the longest viewing times. This amounted to few rejected trials, with 99% (SD = 1.4%) of trials retained across all participants.

We analyzed all data using mixed-effects regression models (Baayen, Davidson, & Bates, 2008) with maximal random effects structure, including Peak Type (Scene or Star) and Sequence Type (Coherent or Scrambled) and their interaction as fixed effects, and random slopes for both participants and items (Barr, Levy, Scheepers, & Tily, 2013), using the lme4 package (Bates, Maechler, Bolker, & Walker, 2014). Viewing times were log-transformed and analyzed at the critical panel (CP) and the immediately subsequent panel (CP+1), as well as non-critical panels. Finally, correlations were used to compare the VLFI fluency scores with viewing times.

Results

Ratings

Figure 3 shows how participants rated the strips, according to condition. Coherent Scenes were rated highest (5.99, SD: 1.57), followed by Coherent Stars (5.14, SD: 1.95), Scrambled Scenes (3.98, SD: 1.88), and finally, Scrambled Stars (3.78, SD: 2.06). Peak type had a significant influence on ratings ($\beta = -.83$, $t = -5.73$, $p < .0001$), indicating that sequences with Depicted Scenes received a significantly higher rating than those with action stars. Sequence Type also significantly influenced ratings ($\beta = -2.01$, $t = -8.67$, $p < .0001$), with Coherent sequences receiving better ratings than Scrambled sequences. There was also a significant interaction between Peak Type and Sequence Type ($\beta = .63$, $t = 3.09$, $p < .001$).

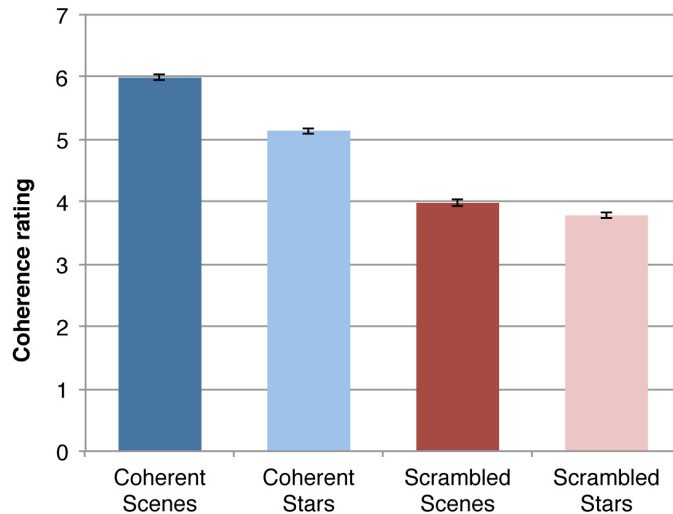


Figure 3. Ratings on a 1 to 7 scale (1=hard, 7=easy) for how easy strips were to understand for Scrambled and Coherent sequences with Depicted Scenes or Action Stars. Error bars depict standard error.

Viewing times

Viewing times to non-critical panels across ordinal sequence position showed that panels in Scrambled sequences were consistently viewed slower than those in Coherent sequences. While the particular panel position had no influence on viewing times overall ($\beta = .01$, $t = -0.5$, $p > .61$), Sequence Type did ($\beta = 0.21$, $t = 4.33$, $p < .0001$), and the interaction was also significant

Action starring narrative and events

($\beta=-.04$, $t=-2.87$, $p<.005$). In addition, as depicted in Figure 4, final panels of Coherent sequences were slower than those in Scrambled sequences, $t(27)=4.4$, $p<.001$, $d=.28$, while the first panel of each sequence was viewed longer than other ordinal positions in both sequence types (all t s > 4.4 , all p s $< .001$, all d s $> .65$), while the final panel of only Coherent sequences was viewed longer than the preceding panels in positions 2 through 5 (all t s > 3.3 , all p s $< .005$, all d s $> .41$).

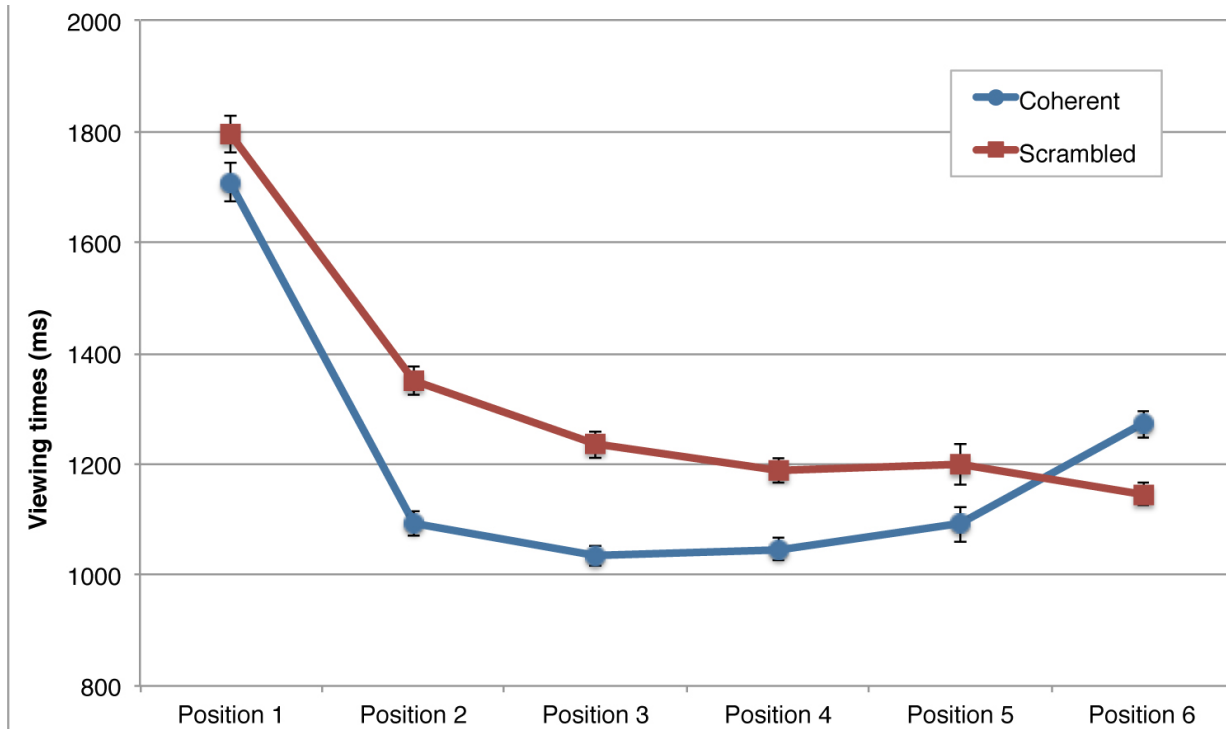


Figure 4. Viewing times across ordinal panel position for non-critical panels collapsed across Coherent and Scrambled sequences. Error bars depict standard error

Viewing times for critical panels are depicted in Figure 5, and listed along with standard deviation and standard error in Table 1. At the critical panel (CP), Peak Type had a significant influence on viewing times ($\beta=.3$, $t=-4.2$, $p<.0001$), with Depicted Scenes being viewed slower than Action Stars (Figure 5). Sequence Type also significantly influenced how long participants viewed a panel ($\beta=.18$, $t=2.28$, $p<.02$). There was also a significant interaction between Peak Type and Sequence Type ($\beta=-.09$, $t=-2.16$, $p<.02$). At the panel following the critical panel (CP+1), again Peak Type had a significant influence on viewing times ($\beta=.36$, $t=4.38$, $p<.0001$), with Depicted Scenes being viewed faster than Action Stars (Figure 5). Sequence Type also significantly influenced how long participants viewed a panel ($\beta=.18$, $t=2.24$, $p<.03$). Again, a significant interaction appeared between Peak Type and Sequence Type ($\beta=-.16$, $t=-3.13$, $p<.001$).

Action starring narrative and events

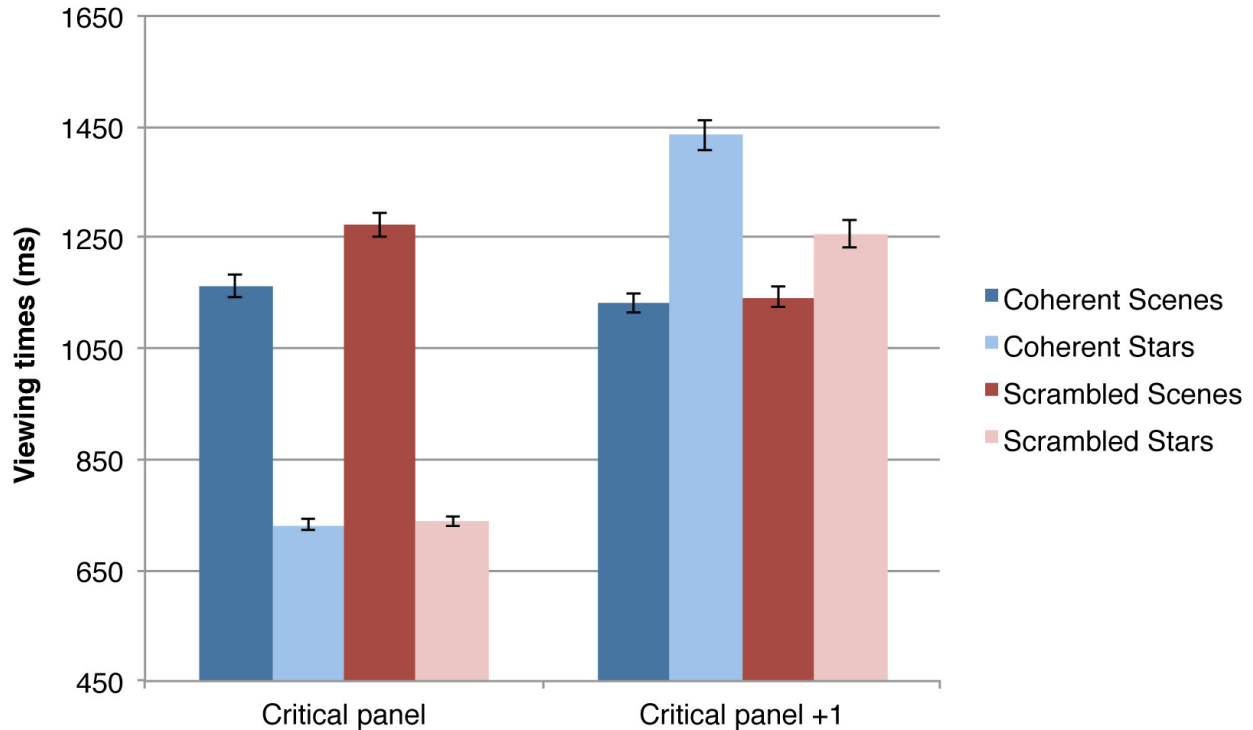


Figure 5. Viewing times at the critical Peak panel and the panel following it (CP+1) for Scrambled and Coherent sequences with either Depicted Scenes or Action Stars. Error bars depict standard error.

Discussion

This experiment compared the viewing times of depicted scenes and action stars in coherent narrative sequences and scrambled sequences. Coherent sequences were easier to understand than scrambled sequences, with slower viewing times appearing across the ordinal position of non-critical panels in sequences for scrambled sequences, and coherent sequences rated as more comprehensible than scrambled ones, regardless of Peak type. These results replicate established findings across domains that scrambling a narrative impairs comprehension (Cohn et al., 2012; Gernsbacher et al., 1990; Mandler, 1978, 1984; Mandler & DeForest, 1979; Stein & Nezworski, 1978), and are consistent with findings that narrative events with functional relations and/or continuity with preceding information are read faster than those that are not causally related (Radvansky & Copeland, 2000; Zwaan, Magliano, & Graesser, 1995).

In addition, slower viewing times appeared for starting panels for both sequence types, again consistent with previous findings in verbal discourse (Glanzer, Fischer, & Dorfman, 1984; Haberlandt, 1984) and visual narratives (Cohn, 2014; Cohn & Paczynski, 2013; Gernsbacher, 1983) where the starting unit “lays a foundation” of information for the subsequent narrative (Gernsbacher, 1990). That viewing times were slightly longer for starting panels of scrambled sequences, above and beyond this process, implies that these panels were not prototypical for beginning a sequence (Cohn, 2014), since no prior context would have impacted their processing. In contrast, the slowing of viewing times to the final panel of the coherent sequences suggests a wrap-up effect (Cohn, 2014) consistent with those observed at the end of sentences (e.g., Rayner, Kambe, & Duffy, 2000). However, this slowing appeared only for coherent

sequences and not for scrambled sequences, suggesting that participants responded to a feature of the narrative (ex. a Release panel) rather than ordinal position alone. This interpretation is further supported by the fact that filler sequences varied the length of the stimuli, making it harder for participants to anticipate that the sixth panel would be the final panel in these experimental sequences.

Like non-critical panels, critical panels with depicted scenes appeared to be slower in scrambled than coherent sequences. However, action stars in coherent and scrambled sequences were viewed at nearly the same pace, despite having very different contexts. We therefore cannot confirm that action stars play a narrative role in coherent sequences that is ameliorated in scrambled sequences. However, action stars were viewed almost twice as fast as depicted scenes. This rapid viewing can perhaps be attributed to the physical differences between these panels: Without representational information (i.e., characters, objects) action stars contain far less visual information than normal Peaks, and thus can be viewed more rapidly because they lack the need to process basic scenes (e.g., Oliva, 2005). The significantly shorter viewing times to action stars than depicted scenes may thus reflect a ceiling for action stars, which may have been reached regardless of sequence context because of their impoverished representation.

At the panel following the critical panel, longer viewing times appeared to panels following action stars than to the same panels following depicted scenes within each of the sequence types. This supports the idea that comprehenders generate inferences at a panel following an action star, because comprehenders must infer the contents of the prior panel, unlike when such information is provided overtly in a depicted Peak. While viewing times alone cannot confirm that inferences may be made—because no exploration of participants' actual inferences were tested (such as with a think-aloud task)—these results are consistent with studies of discourse that have interpreted longer viewing times to sentences eliciting inferences than controls as evidence of inference generation (Haviland & Clark, 1974; Keenan et al., 1990; Sanford & Garrod, 1981; van den Broek, 1994).

Despite not finding evidence for a narrative role of action stars, these results may be indicative of action stars maintaining the causal coherence of a sequence (Magliano et al., 1993; Singer et al., 1992; Trabasso et al., 1984), possibly through inference (Magliano et al., 1993). While processing is generally faster for elements that retain strong causal relations than those with weaker causal connections (Keenan, Baillet, & Brown, 1984; Myers, Shinjo, & Duffy, 1987; Radvansky & Copeland, 2000), the shorter viewing times to panels after action stars in scrambled sequences relative to coherent sequences may provide the reverse interpretation: The lack of an influence of action stars in scrambled sequences for generating inferences may indicate a lack of causal structure in which action stars are embedded. Action stars therefore maintain causal coherence in the sequence, despite not sustaining referential continuity with the prior context (e.g., Magliano & Zacks, 2011; Saraceni, 2001). Although we find this explanation appealing, we cannot be certain that viewing times across sequence types reflects an amelioration of inference generation, given that CP+1 panels were likely different images between coherent and scrambled sequences. Thus, we offer this analysis only tentatively.

An alternative interpretation may attribute the slowing caused by action stars not to inference generation, but to them being “surprising” panels given the context. Under this view, action stars may not play a narrative role at all, and the slowing to subsequent panels is a reaction to their disruption of the semantic and/or narrative structure of a sequence (Zwaan et al., 1995)—with or without inference. This slowing would be consistent with prior research showing that longer viewing times appear to panels following narrative and/or semantic violations than those

following normal Peak panels (Cohn, 2012). Indeed, slightly slower viewing times also arose to panels following action stars than following depicted scenes in scrambled sequences, where no inference should be possible. Under this view, action stars are an unexpected panel that requires a “recovery” following their appearance, no matter the sequence’s context or the inference required.

The longer times appearing to panels after coherent action stars than scrambled action stars could be accounted for by the “surprise” interpretation: Since coherent narratives are disrupted more by an incongruity than a scrambled sequence (as evident in viewing times at and after normal Peaks), longer viewing times appear after action stars in coherent sequences than those in scrambled sequences. An additional alternative may attribute the relative difference between viewing times following action stars to inference, with action stars in coherent sequences sponsoring the generation of inference above and beyond a reaction to their incongruity.

Experiment 2: Action stars and violations

Altogether, Experiment 1 suggested that action stars require comprehenders to generate inferences. However, such results do not provide evidence that action stars play a narrative role in the sequence, and, further, leave open the possibility that action stars may be incongruous to a sequence—regardless of inference generation. If action stars are surprising, they should evoke the same increase in viewing times at a subsequent panel as fully anomalous panels, such as a Peak from another unrelated sequence. Nevertheless, the impoverished graphic structure of action stars should still be viewed faster than panels with more graphic content. Thus, an additional contrast would be to compare them to an empty, blank panel lacking content entirely, which would be closer in physical appearance. An empty panel, devoid of content, should indeed be anomalous to a sequence, but should also evoke inferences for missing information, similar to action stars. Experiment 2 therefore contrasted viewing times and ratings of normal, coherent Peaks with action stars, blank panels, and anomalous Peaks.

Methods

Stimuli

40 strips were chosen from the same corpus of sequences as in Experiment 1 and expanded into four experimental sequence types (Figure 6). Coherent Peaks used the Peak panels from the original strips, depicting the full representation of events. These panels were replaced with Action Stars or Blank Panels, which were both expected to force a reader to create inferences about their contents. Blank Panels were preferred over omitting Peak panels in order to maintain the same number of panels in a sequence, and to provide a panel to compare in viewing times to the Action Star. Finally, Anomalies replaced the Coherent Peak with another Peak panel that did not make sense in the context of the sequence. These panels used Peaks from other Coherent sequences, counterbalanced such that each participant viewed these critical panels only once.

Action starring narrative and events

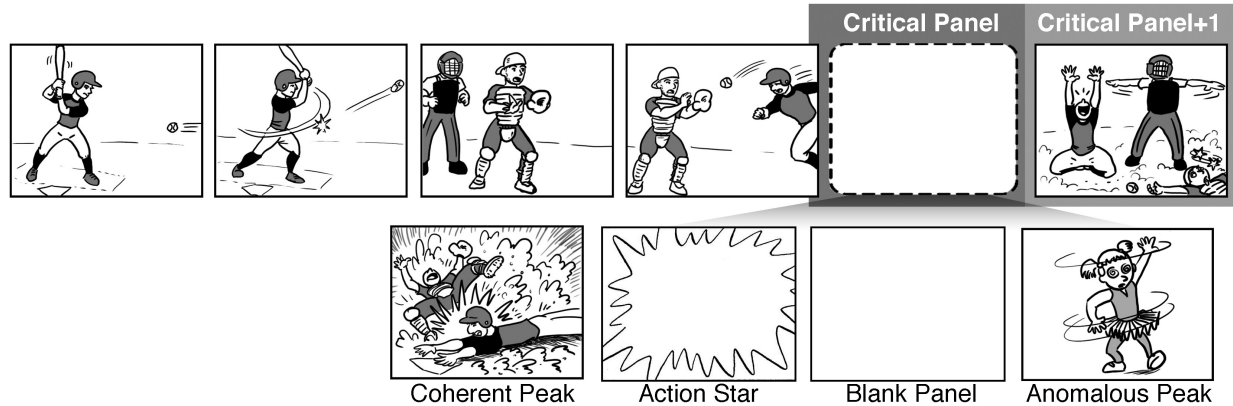


Figure 6. Example stimuli from sequences where the original Coherent Peak was replaced with either Action Stars, Blank Panels, or Anomalous Peaks.

Sequences from each quadruplet were distributed into four counterbalanced lists such that no sequences or panels were repeated for a participant (except for Action Stars and Blank Panels), and each participant viewed 10 of each sequence type. 10 additional filler strips inserted action stars into non-Peak positions either towards the beginning or end of the sequence, distributed evenly into each list. This control ensured that participants also viewed action stars at non-Peak positions. Finally, 30 sequences provided additional coherent strips to increase the number of coherent sequences without action stars. Each list presented participants with sequences in a randomized order.

Participants

Twenty-eight Tufts University undergraduates with experience reading comics (15 male, 13 female, mean age: 19.8,) participated in the experiment for compensation and gave their informed written consent according to the guidelines of the Tufts Human Subjects Review Board. Participants' comic reading fluency was assessed as "average," with a mean of 13.6 (SD=7.1; range = 2.6 - 31).

Procedure

The same procedure was used in Experiment 2 as Experiment 1.

Data Analysis

We again analyzed viewing times and ratings using mixed-effects regression models (Baayen et al., 2008) with maximal random effects structure, this time with Peak Type only as fixed effect, and random slopes for both participants and items. In addition to the overall regression, we analyzed pairs of interest separately (see below). Data from one item was discarded due to recording errors, leaving 39 sequences. As in Experiment 1, sequences were rated for how easy they were understood (1=hard to understand, 7=easy to understand). Again, outlier viewing times for each participant were discarded if they fell below 300ms or above 8000ms.

Results

Ratings

Figure 7 shows how participants rated the strips. Coherence ratings differed between sequences for all Peak Types. Sequences with Coherent Peaks were the most understandable (6.24, SD: 1.2), followed by those with Action Stars (5.26, SD: 1.9), then those with Blank Panels (4.81, SD: 1.8), and those with Anomalous Peaks (3.66, SD: 1.8). Differences between ratings were significant (all β s>|.98|, all t s>|5.5|, all p s<.00001).

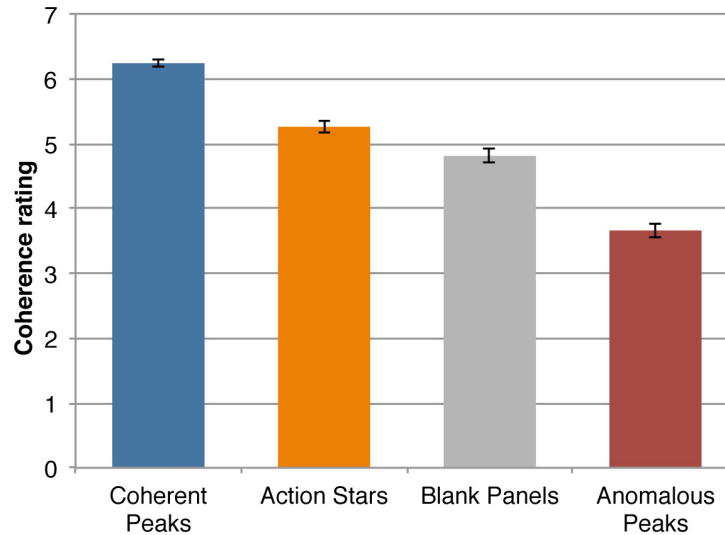


Figure 7. Mean ratings for the coherence for sequences with various Peak Types on a 1 to 7 scale (1=incoherent, 7=coherent). Error bars depict standard error.

Viewing times

Figure 8 shows mean viewing times on the critical panel (CP) and subsequent panel (CP+1) (see also Table 1). At the critical panel (CP), Peak Type had a significant influence on viewing times (β =.12, t =6.44, p <.0001), with Coherent Peaks and Anomalous Peaks viewed longer than Action Stars or Blank Panels. Critically, the difference in viewing times between Action Stars and Blank Panels was significant (β =0.17, t =4.44, p <.0001). At the panel following the critical panel (CP+1), Peak Type again influenced viewing times significantly (β =.07, t =3.28, p <.002), with Coherent Peaks being viewed faster than Action Stars, Blank Panels, or Anomalous Peaks (Figure 5). There was no significant difference between Action Stars or Blank Panels (β =.02, t =.33, p >.73).

Action starring narrative and events

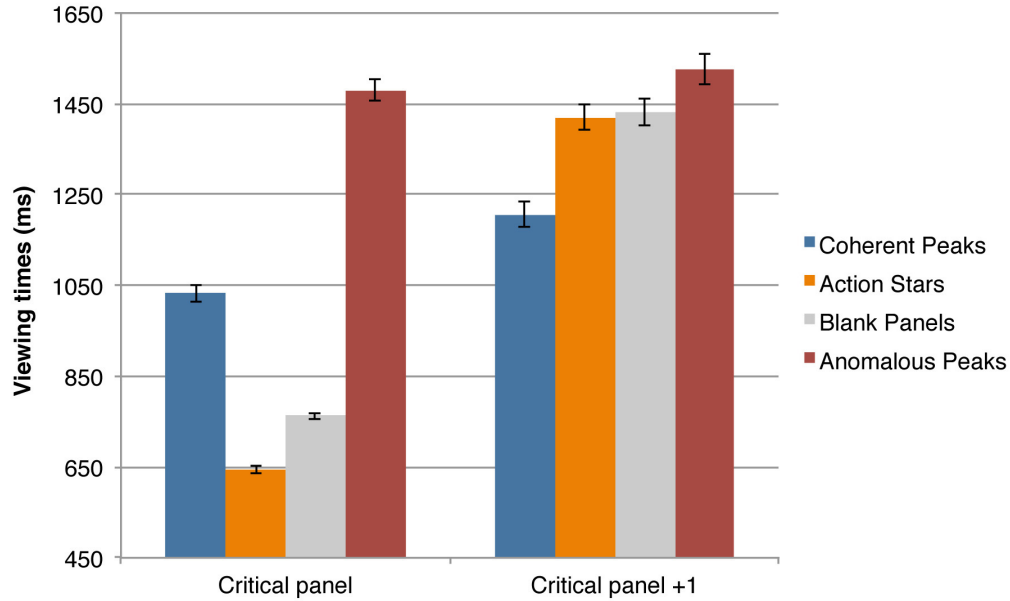


Figure 8. Viewing times to the critical Peak panel and the subsequent panel (Critical Peak+1) in sequences where the Peak panel was substituted for Action Stars, Blank Panels, or Anomalous Peaks. Error bars depict standard error.

VLFI score

A significant negative correlation between VLFI scores and panels following Blank Panels suggested that faster viewing times by participants with greater experience reading comics, $r(54) = -.27, p < .05$. No other significant correlations were found.

Discussion

This experiment examined action stars' structural and inferential roles in a visual narrative sequence by comparing viewing times between sequences with coherent Peaks, action stars, blank panels, and anomalous Peaks. Overall, action stars were viewed faster than all other critical panels, and panels following them were viewed comparably to panels following blank panels and anomalous Peaks. These results are consistent with the idea that action stars play a structural role as Peaks in the narrative grammar and require inference to be understood.

Action stars as Peaks

Our primary question was whether action stars function as Peaks in the sequence, or if they were simply incongruous, as implied by the results of Experiment 1. Coherent Peaks were always viewed faster than anomalous Peak, while action stars and blank panels were viewed considerably faster than both coherent and anomalous Peaks. As in Experiment 1, this gross difference was likely due to the amount of visual information between these panels: action stars and blank panels have far less visual information than coherent and anomalous Peaks, and this lack of needing basic scene perception processing should lead to faster viewing times than panels depicting actual referential and event information.

The crucial distinction compared action stars and blank panels. If quantity of visual information alone guided the comprehension of these panels, action stars should be viewed

longer than blank panels. Alternatively, if action stars were incongruous to a sequence, they should be viewed at comparable times to the incongruous blank panels. In fact, action stars were viewed significantly faster than blank panels at the critical panel position, despite being more visually rich. These shorter viewing times suggest that action stars do indeed play a felicitous functional role in the narrative at the Peak position, implying that this conventionalized visual morpheme satisfied the structural expectations that a Peak panel should follow the prior Initial panels (Cohn et al., 2014; Cohn & Paczynski, 2013). Participants' ratings further support this: sequences with action stars were judged as more coherent than those with blank panels.

The generation of inference

In assessing the generation of inference, structurally congruous sequences were rated higher than structurally incongruous sequences: Ratings descended from sequences with coherent Peaks to action stars to blank panels and finally anomalous Peaks. These ratings are consistent with previous findings of low coherence judgments to sequences with omitted Peak panels (Cohn, 2014). The higher ratings to sequences with action stars than those with blank panels or anomalous panels supports their felicity. However, their omission of information kept them as less coherent than strips with normal Peaks, suggesting that less informative sequences are rated lower, regardless of felicity.

Viewing times for panels following action stars, blank panels, and anomalous Peaks were longer than to panels following coherent Peaks. The slower times to panels after anomalous Peaks is consistent with longer viewing times shown previously to panels following violations of narrative structure and/or semantic associations to Peak panels (Cohn, 2012), and is consistent with findings of faster reading times to congruous discourse information than information that does not fit a situation (e.g., Zwaan et al., 1995). This suggests that some “recovery” or “reanalysis” of the sequence is necessary following an incongruous panel. Such slowing cannot necessarily be attributed to inference, given the lack of inference able to be drawn from a contextually incongruous preceding image.

Panels following action stars or blank panels did not differ from each other in viewing times, and were both longer than panels following coherent Peaks. Like the comparable times to panels after anomalous Peaks, these slower viewing times could be interpreted as a reaction to action stars or blank panels being incongruous to the sequence. However, action stars are *not* viewed as incongruous as blank or anomalous panels, as evidenced by viewing times at the critical panel and by differences in coherence ratings (not to mention that they are an explicit part of created visual narratives). This means that viewing times after action stars may not be motivated by recovery or reanalysis due to incongruity, but rather may be indicative of inference to understand the omitted event information. In addition, participants viewed panels following blank panels faster when they had greater comic reading expertise. This suggests that inference generation following a narrative incongruity benefits from frequency of reading visual narratives, as in previous findings that inference of omitted content from visual narratives correlates with experience reading comics (Nakazawa, 2005).

Nevertheless, the similarity in viewing times between panels following action stars and blank panels implies that action stars do not provided added benefit to inference generation, and thus that any semantically depleted content would promote inferences (e.g., Myers & O'Brien, 1998). Despite action stars being more congruous than blank panels or anomalous panels, the similarity in viewing times after these panels suggests that action stars are not “facilitating” the processing of inference: Action stars do require inference, but they do not aid in their

computation. The absence of a difference at this panel may occur because simple narratives like these allow for inference motivated by bottom-up information (Long & Lea, 2005) that would be insensitive to the panel omitting information, and such narratives may not allow for a strong relationship between the evaluation of coherence (as found in the ratings) and processing time (Long & Lea, 2005; McCrudden, Magliano, & Schraw, 2011; Rapp & Mensink, 2011). Yet, this interpretation potentially implies similar processing mechanisms after congruous (action stars) and incongruous panels (blank panels, anomalous panels) along with similar processing between those that require inference (action stars, blank panels) and those that do not (anomalies). It may be the case that viewing times cannot detect these types of functional differences, assessment of which would require a more sensitive measurement technique.

General Discussion

These experiments looked at the interaction between narrative structure and inference generation in the comprehension of visual narratives by examining a semantically impoverished conventionalized unit in visual narrative: action stars. First, we were interested in the inference that “filled in” the information about the missing content. Second, we were interested in whether action stars play a functional role as Peaks of a visual narrative sequence.

The generation of inference

A primary goal of this study was to confirm whether action stars force a reader to infer unspecified information. Previous research found that participants were better at recognizing omitted Initial or Peak panels, which typically depict more active events, than omitted Establishers or Releases, which typically depict more passive events (Cohn, 2014). We asked here: Do panels that contain minimal semantic information force comprehenders to infer the contents?

In Experiment 1, longer viewing times appeared to panels after action stars than after normal Peaks, implying that participants generated inferences for the unseen information in the action stars. This is consistent with research showing longer viewing times to sentences that elicit bridging inferences than to non-inference generating controls (Haviland & Clark, 1974; Keenan et al., 1990; Sanford & Garrod, 1981; van den Broek, 1994). Nevertheless, the possibility still existed that action stars were simply surprising to the sequence, and not requiring inferences. Coherence ratings partially supported this interpretation, because sequences with action stars were rated as less coherent than sequences with normal Peaks, whether in coherent or scrambled sequence contexts.

Experiment 2 further explored whether action stars were incongruous to the sequence by contrasting them with coherent Peaks, blank panels, and anomalous Peaks. First, the longer viewing times to blank panels than action stars, despite having less physical information, suggested that action stars were more congruous to the sequence than blank panels, which clearly were incongruous. Second, panels after action stars were again viewed longer than those following coherent Peaks and equal to panels following blank panels and anomalous Peaks. This suggested that participants elicited inferences to understand the omitted information in action stars. Together, these results imply that action stars require participants to generate inferences in order to make sense of unseen information.

Despite these results, viewing times and coherence ratings alone do not directly test what types of inferences may be generated by action stars, or whether inferences may indeed be

generated at the action star itself. For example, we would expect that, despite both requiring inference, images following blank panels may involve a different type of processing than those after action stars, given their structural incongruity. Such differences would require measurements such as event-related potentials, which can detect functional differences that may not be distinguishable in viewing times (e.g., Cohn et al., 2014; Cohn et al., 2012; Sitnikova et al., 2008).

Action stars as “lexical” items

Finally, these experiments suggest that action stars, a convention of the visual language of comics, require inference because of an impoverished semantic structure, yet they also play a specific functional role in narrative as a Peak panel. Evidence for this came from Experiment 2, where action stars were viewed shorter than blank panels, despite having more physical graphic structure and semantic information. This suggested that action stars fulfilled the prior expectations for an incoming Peak panel (Cohn et al., 2014) more than a fully incongruous and semantically impoverished blank panel, and thus they play a more felicitous role in the narrative structure. Such a narrative role is also interesting semantically because action stars sustain the semantic causal coherence in the sequence (Magliano et al., 1993; Singer et al., 1992; Trabasso et al., 1984), despite maintaining no referential continuity.

Because action stars play this role as Peaks while still maintaining very little semantic information, they have been likened as structurally analogous to “pro-forms” in sentences, such as *there*, *she*, *it*, *them*, etc. (Cohn, 2013a, 2013b). Like pro-forms, we might expect that action stars could function as a diagnostic tool for Peak panels. For example, pronouns can grammatically be substituted for nouns or noun phrases (*The mailman delivered a package to her*), but not for any other grammatical category, like a prepositional phrase (**The mailman delivered a package her*). Similarly, we might expect that action stars can felicitously be substituted for other Peaks, but not for any other narrative category (Cohn, 2013a, 2013b), though the present study provides no explicit evidence of this hypothesis. This incongruity of action stars outside their context as Peaks was potentially suggested by the lower ratings to scrambled sequences with action stars than coherent sequences with action stars in Experiment 1. In a fully scrambled sequence, where action stars can play no structural role, they lead to an even more incongruous interpretation. Nevertheless, how the contrast of grammatical or ungrammatical substitutions would impact the comprehension of felicitous visual narrative can be addressed by future experiments.

Acknowledgements

Gina Kuperberg and Phillip Holcomb are thanked for funding this research, and along with Ray Jackendoff provided helpful suggestions in research design and analysis. Martin Paczynski gave insightful feedback on previous drafts. Fantagraphics Books is thanked for their generous donation of *The Complete Peanuts*.

Funding

This work was supported by NIH grants awarded to Gina Kuperberg and Phillip Holcomb under Grant NIH HD25889 and NIMH (R01 MH071635) respectively; and NARSAD (with the Sidney Baer Trust) to Gina Kuperberg.

References

- Baayen, R. H., Davidson, D. J., & Bates, D. M. (2008). Mixed-effects modeling with crossed random effects for subjects and items. *Journal of Memory and Language*, 59(4), 390-412. doi: <http://dx.doi.org/10.1016/j.jml.2007.12.005>
- Barr, D. J., Levy, R., Scheepers, C., & Tily, H. J. (2013). Random effects structure for confirmatory hypothesis testing: Keep it maximal. *Journal of Memory and Language*, 68(3), 255-278. doi: <http://dx.doi.org/10.1016/j.jml.2012.11.001>
- Barrett, S. E., & Rugg, M. D. (1990). Event-related potentials and the semantic matching of pictures. *Brain and Cognition*, 14(2), 201-212. doi: [http://dx.doi.org/10.1016/0278-2626\(90\)90029-N](http://dx.doi.org/10.1016/0278-2626(90)90029-N)
- 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](http://dx.doi.org/10.1016/0028-3932(88)90034-6)
- Bates, D., Maechler, M., Bolker, B., & Walker, S. (2014). lme 4: Linear mixed-effects models using Eigen and S4. R package v. 1.1-5.
- Bordwell, D. (1985). *Narration in the Fiction Film*. Madison: University of Wisconsin Press.
- Bordwell, D. (2007). *Poetics of Cinema*. New York, NY: Routledge.
- Bordwell, D., & Thompson, K. (1997). *Film Art: An Introduction* (5th Edition ed.). New York, NY: McGraw-Hill.
- Branigan, E. (1992). *Narrative Comprehension and Film*. London, UK: Routledge.
- Chatman, S. (1978). *Story and Discourse*. Ithaca: Cornell University Press.
- Cohn, N. (2012). *Structure, meaning, and constituency in visual narrative comprehension*. (Doctoral Dissertation), Tufts University, Medford, MA.
- 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
- Eisenstein, S. (1942). *Film Sense* (J. Leyda, Trans.). New York, NY: Harcourt, Brace World.

- Gernsbacher, M. A. (1983). *Memory for surface information in non-verbal stories: Parallels and insights to language processes*. (Doctoral Dissertation), University of Texas at Austin, Austin, Texas.
- Gernsbacher, M. A. (1990). *Language Comprehension as Structure Building*. Hillsdale, NJ: Lawrence Erlbaum.
- Gernsbacher, M. A., Varner, K. R., & Faust, M. (1990). Investigating differences in general comprehension skill. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *16*, 430-445.
- Glanzer, M., Fischer, B., & Dorfman, D. (1984). Short-term storage in reading. *Journal of Verbal Learning and Verbal Behavior*, *17*, 467-486.
- Graesser, A. C., Millis, K. K., & Zwaan, R. A. (1997). Discourse Comprehension. *Annual Review of Psychology*, *48*, 163-189.
- Haberlandt, K. (1984). Components of sentence and word reading times. In D. E. Kieras & M. A. Just (Eds.), *New methods in reading comprehension research* (pp. 219-251). Hillsdale, NJ: Erlbaum.
- Halliday, M. A. K., & Hasan, R. (1976). *Cohesion in English*. London: Longman.
- Haviland, S. E., & Clark, H. H. (1974). What's new? Acquiring new information as a process in comprehension. *Journal of Verbal Learning and Verbal Behavior*, *13*, 512-521.
- Hobbs, J. R. (1985). *On the coherence and structure of discourse*. Stanford, CA: CSLI Technical Report 85-37.
- Keenan, J. M., Baillet, S. D., & Brown, P. (1984). The effect of causal cohesion on comprehension and memory. *Journal of Verbal Learning and Verbal Behavior*, *23*, 115-126.
- Keenan, J. M., Potts, G. R., Golding, J. M., & Jennings, T. M. (1990). Which elaborative inferences are drawn during reading? A question of methodologies. In D. A. Balota, G. B. Flores d'Arcais, & K. Rayner (Eds.), *Comprehension Processes in Reading* (pp. 377-402). Hillsdale, NJ: Lawrence Erlbaum Associates.
- Kehler, A. (2002). *Coherence, Reference, and the Theory of Grammar*. Stanford: CSLI Publications.
- Kintsch, W. (1988). The role of knowledge in discourse comprehension: A construction-integration model. *Psychological Review*, *95*(2), 163-182.
- Kintsch, W. (1998). *Comprehension: A paradigm for cognition*: Cambridge university press.
- Kuleshov, L. (1974). *Kuleshov on Film: Writings of Lev Kuleshov* (R. Levaco, Trans.). Berkeley: University of California Press.
- 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., & Lea, R. B. (2005). Have We Been Searching for Meaning in All the Wrong Places? Defining the "Search After Meaning" Principle in Comprehension. *Discourse Processes*, *39*(2-3), 279-298. doi: 10.1080/0163853X.2005.9651684
- Magliano, J. P., Baggett, W. B., Johnson, B. K., & Graesser, A. C. (1993). The time course of generating causal antecedent and causal consequence inferences. *Discourse Processes*, *16*(1-2), 35-53. doi: 10.1080/01638539309544828

- Magliano, J. P., Dijkstra, K., & Zwaan, R. A. (1996). Generating predictive inferences while viewing a movie. *Discourse Processes*, 22, 199-224.
- Magliano, J. P., Miller, J., & Zwaan, R. A. (2001). Indexing space and time in film understanding. *Applied Cognitive Psychology*, 15, 533-545.
- Magliano, J. P., & Zacks, J. M. (2011). The Impact of Continuity Editing in Narrative Film on Event Segmentation. *Cognitive Science*, 35(8), 1489-1517.
- Mandler, J. M. (1978). A code in the node: The use of story schema in retrieval. *Discourse Processes*, 1, 14-35.
- Mandler, J. M. (1984). *Stories, Scripts, and Scenes: Aspects of Schema Theory*. Hillsdale, NJ: Lawrence Earlbaum Associates.
- Mandler, J. M., & DeForest, M. (1979). Is there more than one way to recall a story? *Child Development*, 50, 886-889.
- Mandler, J. M., & Johnson, N. S. (1977). Remembrance of things parsed: Story structure and recall. *Cognitive Psychology*, 9, 111-151.
- Mann, W. C., & Thompson, S. A. (1987). *Rhetorical structure theory: A theory of text organization*. Marina del Rey, CA: Information Sciences Institute.
- McCloud, S. (1993). *Understanding Comics: The Invisible Art*. New York, NY: Harper Collins.
- McCrudden, M. T., Magliano, J. P., & Schraw, G. J. (2011). Relevance in text comprehension. In M. T. McCrudden, J. P. Magliano, & G. J. Schraw (Eds.), *Text relevance and learning from text* (pp. 1-18). Greenwich, CT: Information Age.
- McKoon, G., & Ratcliff, R. (1986). Inferences about predictable events. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 12(1), 82.
- McKoon, G., & Ratcliff, R. (1992). Inference during reading. *Psychological Review*, 99(3), 440-466.
- McNamara, D. S., & Magliano, J. (2009). Toward a comprehensive model of comprehension. *Psychology of learning and motivation*, 51, 297-384.
- 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
- Myers, J. L., Shinjo, M., & Duffy, S. A. (1987). Degree of causal relatedness and memory. *Journal of Memory and Language*, 26(4), 453-465. doi: 10.1016/0749-596x(87)90101-x
- 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.
- Oliva, A. (2005). Gist of the scene. *Neurobiology of attention*, 696(64), 251-258.
- Potsch, E., & Williams, R. F. (2012). Image Schemas and Conceptual Metaphor in Action Comics. In F. Bramlett (Ed.), *Linguistics and the Study of Comics*. New York: Palgrave MacMillan.
- Radvansky, G. A., & Copeland, D. E. (2000). Functionality and spatial relations in memory and language. *Memory & Cognition*, 28(6), 987-992. doi: 10.3758/BF03209346
- Rapp, D. N., & Mensink, M. C. (2011). Focusing effects from online and offline reading tasks. In M. T. McCrudden, J. P. Magliano, & G. J. Schraw (Eds.), *Text relevance and learning from text* (pp. 141-164). Greenwich, CT: Information Age.
- Rayner, K., Kambe, G., & Duffy, S. A. (2000). The effect of clause wrap-up on eye movements during reading. *The Quarterly Journal of Experimental Psychology: Section A*, 53(4), 1061-1080.

- 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>
- Robertson, D. A. (2000). *Functional neuroanatomy of narrative comprehension*. (Doctoral Dissertation), University of Wisconsin, Madison, Madison, WI.
- Rumelhart, D. E. (1975). Notes on a schema for stories. In D. Bobrow & A. Collins (Eds.), *Representation and understanding* (pp. 211-236). New York, NY: Academic Press.
- Sanford, A., & Garrod, S. (1981). *Understanding written language: Exploration of comprehension beyond the sentence*. Chichester: Wiley.
- Saraceni, M. (2000). *Language Beyond Language: Comics as Verbo-Visual Texts*. (Doctoral Dissertation), University of Nottingham, Nottingham.
- Saraceni, M. (2001). Relatedness: Aspects of textual connectivity in comics. In J. Baetens (Ed.), *The Graphic Novel* (pp. 167-179). Leuven: Leuven University Press.
- Singer, M., Halldorson, M., Lear, J. C., & Andrusiak, P. (1992). Validation of causal bridging inferences in discourse understanding. *Journal of Memory and Language*, 31(4), 507-524. doi: 10.1016/0749-596x(92)90026-t
- 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.
- Stainbrook, E. J. (2003). *Reading Comics: A Theoretical Analysis of Textuality and Discourse in the Comics Medium*. (Doctoral Dissertation), Indiana University of Pennsylvania, Indiana, PA.
- Stein, N. L., & Glenn, C. G. (1979). An analysis of story comprehension in elementary school children. In R. Freedle (Ed.), *New Directions in Discourse Processing* (pp. 53-119). Norwood, NJ: Ablex.
- Stein, N. L., & Nezworski, T. (1978). The effects of organization and instructional set on story memory. *Discourse Processes*, 1(2), 177-193.
- Thorndyke, P. (1977). Cognitive structures in comprehension and memory of narrative discourse. *Cognitive Psychology*, 9, 77-110.
- Trabasso, T., Secco, T., & van den Broek, P. (1984). Causal cohesion and story coherence. In H. Mandl, N. L. Stein, & T. Trabasso (Eds.), *Learning and comprehension of text* (pp. 83-111). Hillsdale, NJ: Erlbaum.
- Trabasso, T., & Sperry, L. L. (1985). Causal relatedness and importance of story events. *Journal of Memory and Language*, 24, 595-611.
- van den Broek, P. (1994). *Comprehension and Memory of Narrative Texts: Inferences and Coherence*. New York: Academic Press.
- van Dijk, T., & Kintsch, W. (1983). *Strategies of Discourse Comprehension*. New York: Academic Press.
- Walker, M. (1980). *The Lexicon of Comicana*. Port Chester, NY: Comicana, Inc.
- West, W. C., & Holcomb, P. (2002). Event-related potentials during discourse-level semantic integration of complex pictures. *Cognitive Brain Research*, 13, 363-375.
- Yus, F. (2008). Inferring from comics: A multi-stage account. *Quaderns de Filologia. Estudis de Comunicacio*, 3, 223-249.
- Zacks, J. M., Speer, N. K., & Reynolds, J. R. (2009). Segmentation in reading and film comprehension. *Journal of Experimental Psychology: General*, 138(2), 307-327.

Action starring narrative and events

- Zwaan, R. A., Magliano, J. P., & Graesser, A. C. (1995). Dimensions of situation model construction in narrative comprehension. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, *21*, 386-397.
- Zwaan, R. A., & Radvansky, G. A. (1998). Situation models in language comprehension and memory. *Psychological Bulletin*, *123*(2), 162-185.
- Zwaan, R. A., & Rapp, D. N. (2006). Discourse Comprehension. In M. J. Traxler & M. A. Gernsbacher (Eds.), *Handbook of Psycholinguistics* (2nd ed., pp. 641-674). New York, NY: Academic Press.