Drawing the line between constituent structure and coherence relations in visual narratives

Neil Cohn[1,2] and Patrick Bender[1]

[1] Psychology Department, Tufts University, Medford, MA 02155
[2]Center for Research in Language, UC San Diego, La Jolla, CA 92093-0526

Email: neilcohn@visuallanguagelab.com

Short title: Visual narrative constituents

For final draft version, please consult the published paper at:

Cohn, Neil, and Patrick Bender. 2017. "Drawing the line between constituent structure and coherence relations in visual narratives." ***Journal of Experimental Psychology: Learning, Memory, & Cognition*** **43 (2):289-301. Doi: http://dx.doi.org/10.1037/xlm0000290**

Abstract

Theories of visual narrative understanding have often focused on the changes in meaning across a sequence, like shifts in characters, spatial location, and causation, as cues for breaks in the structure of a discourse. In contrast, the theory of Visual Narrative Grammar posits that hierarchic "grammatical" structures operate at the discourse level using categorical roles for images, which may or may not co-occur with shifts in coherence. We therefore examined the relationship between narrative structure and coherence shifts in the segmentation of visual narrative sequences using a "segmentation task" where participants drew lines between images in order to divide them into sub-episodes. We used regressions to analyze the influence of the expected constituent structure boundary, narrative categories, and semantic coherence relationships on the segmentation of visual narrative sequences. Narrative categories were a stronger predictor of segmentation than linear coherence relationships between panels, though both influenced participants' divisions. Altogether, these results support the theory that meaningful sequential images use a narrative grammar that extends above and beyond linear semantic shifts between discourse units.

Keywords: Narrative; Visual Narrative Grammar; event-indexing model; discourse; comics; visual language

1. Introduction

Research on language has long distinguished between the linear connections of units and their organization into a hierarchic constituent structure. At the discourse level, theories have argued that linear changes in meaning index changes in a broader segmental structure (Asher & Lascarides, 2003; Mann & Thompson, 1987), and such claims have been extended to the non-verbal domain regarding the comprehension of visual narratives (Gernsbacher, 1990; Zacks, Speer, & Reynolds, 2009). Recent work looking specifically at visual narratives has argued that sequences of images (like in comics) are organized by a narrative "grammar" using constituent structures that go beyond linear coherence relationships between individual images (Cohn, 2013b; Cohn, Jackendoff, Holcomb, & Kuperberg, 2014). In this theory, linear changes in meaning may correlate with constituent boundaries, but are not exclusively relied upon to signal such structures. Here, we examine this relationship between "visual narrative grammar" and linear coherence relationships in the segmentation of drawn sequential images. We hypothesized that coherence relations would predict the boundaries between constituents, but not as well as structural aspects of the narrative grammar.

Predominant theories of visual narrative comprehension have focused on the linear relationships between panels—the encapsulated image units of a visual narrative. These linear relationships have often focused on the degree of change that occurs between images with regard to dimensions of characters, spatial locations, causation, and connections to a broader semantic associative network (Magliano & Zacks, 2011; McCloud, 1993; Saraceni, 2001). Similar semantic changes have also been prominent in theories of verbal discourse, exemplified by the event-indexing model (Zwaan, Langston, & Graesser, 1995; Zwaan & Radvansky, 1998), which argues that these coherence changes incur costs in comprehension, as the mental model for understanding a discourse must be updated to incorporate new information. Research with film narratives has confirmed that viewers intuit changes in characters, spatial location, and time between individual film shots (Magliano, Miller, & Zwaan, 2001; Magliano & Zacks, 2011; Zacks et al., 2009).

In contrast to this emphasis on meaning, ***Visual Narrative Grammar*** (VNG) argues that full comprehension extends above and beyond the semantic shifts between units. VNG draws an analogy between the structure of sequential images and the structure of sentences, in that panels take on functional "grammatical" roles that can be organized into hierarchic constituents (Cohn, 2013b). Insofar as it proposes a hierarchic structure for narrative, it may appear similar to previous "grammars" for verbal stories (e.g., Mandler & Johnson, 1977; Rumelhart, 1975; Stein & Nezworski, 1978; Thorndyke, 1977) and film (e.g., Carroll, 1980), which grouped sentences into constituents based on characters' goal-directed events. However, VNG differs from these models in that it uses simpler structures (Cohn, 2013b, 2015b) based on contemporary linguistic models of construction grammar (Culicover & Jackendoff, 2005; Jackendoff, 2002), and uses modifiers beyond a canonical narrative arc (Cohn, 2013a, 2013b, 2015b). In addition, VNG posits an unambiguous separation between structure and meaning (Cohn, Paczynski, Jackendoff, Holcomb, & Kuperberg, 2012), which evoke different neural responses when violated (Cohn et al., 2014; Cohn & Kutas, 2015; Cohn et al., 2012), consistent with the neural responses shown to violations of syntax and semantics in sentences (Friederici, 2002; Hagoort, 2003; Kuperberg, 2007).

In VNG, a narrative schema outlines the canonical order of categorical roles. A narrative sequence may begin with an Establisher, which sets up a situation, often with a passive action.

Initials then set the interactions in motion, which climax in a Peak, concluding with a Release that dissolves this narrative tension. Although other categories and sequencing constructions may elaborate or modify a narrative (Cohn, 2015b), this *Establisher-Initial-Peak-Release* schema characterizes the canonical arc as a constructional pattern stored in memory (Cohn, 2014b; Mandler & Johnson, 1977). In addition, these narrative categories characterize both individual panels and *groupings* of panels containing these narrative sequences. This can better be understood by an example.

Figure 1 illustrates how VNG would describe the narrative structure of a short visual sequence. This sequence shows Charlie Brown and Snoopy playing in the snow: Charlie Brown throws a snowball, which Snoopy chases, only to have it roll down the hill after him turning into a giant snow-boulder. The first panel is an *Initial* since it begins the events of the sequence, here depicting Charlie reaching back with a snowball. A *Peak* then shows a "mini-climax" with the completion of this action: Charlie throwing the snowball. The next panel, an *Establisher*, sets up a new interaction between Snoopy and the snowball. Another Initial then starts a new event, with Snoopy noticing the snowball rolling towards him. Another climax then occurs in the Peak, as Snoopy runs away from the snowball, which has grown to a frightening size. The final panel is a *Release*, a panel showing a resolution, aftermath, or coda of an action. In this panel, the Release shows Snoopy's reaction to the snowball, as he hides behind a tree.

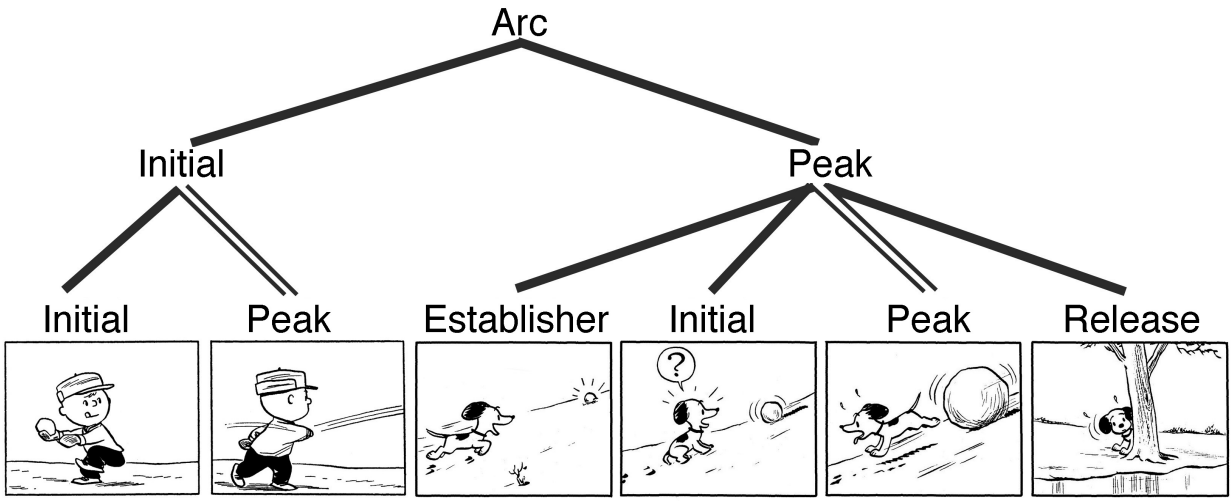

Figure 1. Structure of a novel visual sequence with narrative categories and constituents. Note that the major constituent boundary also has a coherence change in characters and spatial location. Within our 6-panel experimental sequences, this was coded as a "2-4" strip pattern, because the first constituent has two panels and the second constituent contains four panels. *Peanuts* is © Peanuts Worldwide LLC.

An important feature of VNG is that narrative categories do not just apply to panels, but also to whole constituents. In Figure 1, the first two panels do not just precede the final four panels linearly, but rather the first two panels together form a constituent within a larger structure, as do the final four panels. As a result, the first two panels form their own constituent (an Initial) that, as a whole, set in motion the entire second constituent (a Peak) at a higher level of structure. This constituent break does feature a surface semantic change between characters, but structurally it marks an "illegal" surface string: a *Peak-Establisher* panel bigram does not

follow the canonical narrative schema (*E-I-P-R*), and thus should mark the change between constituents. The double bar lines in Figure 1 denote the "heads" of each constituent—the panels within a constituent that motivate their broader clause (usually Peaks). In this way, narrative categories can recursively characterize both individual panels and whole groupings of panels.

Although meaningful cues within panels may influence a panel's role, the narrative categories in VNG are not solely determined by semantic content. Narrative categories are determined both by a panel's bottom-up semantic content and its top-down context in a global sequence. These contextual constraints are determined by distributional tendencies throughout a narrative sequence (Cohn, 2014b), which may prototypically correspond to semantic aspects of event structure (such as preparatory actions corresponding with Initials). This is again analogous to the way that syntactic categories (nouns, verbs) are determined by distributional trends, but prototypically correspond to the semantic content (objects, events) of words (Jackendoff, 1990). Yet, they do not take grammatical roles until appearing in a sentence. For example, the sound string "hit" can play several grammatical roles that are only disambiguated in context: *He hit the wall* (verb); *The song was a hit* (noun); *It was a hit song* (adjective). In a similar way, the context of a sequence may influence the narrative roles played by various images, with some content being more flexible than others (Cohn, 2014b).

Constituent structures are not a unique feature of VNG, and substantial evidence has suggested their presence in visual narrative comprehension both within the VNG paradigm. Participants are highly consistent in where they choose to divide a picture story into sub-episodes (Gernsbacher, 1985), and they are more accurate at remembering altered film shots or picture story images when they precede, rather than follow, a constituent boundary (Carroll & Bever, 1976; Gernsbacher, 1985). Although these findings support the idea that comprehenders group information into segments, such effects could maintain a view of linear coherence relationships. For example, Gernsbacher's (1990) Structure Building Framework posits that comprehenders may simply build a structure until a break between episodes occurs, at which time a new structure begins (Gernsbacher, 1990; Zacks et al., 2009). Such a view does not necessarily require categorical roles that build an internally hierarchic constituent structure for a sequence. This view has been backed by findings that participants' chosen boundaries between discourse structures highly correlate with shifts in linear coherence (Speer & Zacks, 2005; Zacks & Magliano, 2011; Zacks, Speer, Swallow, & Maley, 2010). These observations have been extended to claim that, not only do coherence shifts align with the boundaries of discourse segments, but they provide the signals for such constituents to a comprehender (Gernsbacher, 1990; Zacks & Magliano, 2011; Zacks et al., 2009). Because comprehenders incrementally update their mental models of a situation both within and between segments (Huff, Meitz, & Papenmeier, 2014; Kurby & Zacks, 2012), greater dimensional change at segmentation boundaries results in prediction error that signals a constituent break (Huff et al., 2014; Magliano & Zacks, 2011; Zacks, Speer, Swallow, Braver, & Reynolds, 2007). Thus, in this view, linear coherence changes play an integral role in defining the boundaries between constituents.

It is important to note that VNG is not incompatible with views of linear changes in semantic coherence, nor their correlation with linear coherence relations. VNG hypothesizes that major coherence shifts operate within a *semantic* processing stream that is separate from the narrative grammar (Cohn, 2013b, 2014a; Cohn et al., 2012), and these shifts may indeed inform a reader about the boundaries between *narrative* constituents (as is the case in Figure 1). However, not all breaks in constituent structure align with coherence shifts. For example, in Figure 2b, the first constituent shows Schroeder oppressed by the sun while playing in the sand,

so he builds a sand mound to hide behind in the second constituent. Here, no shift in characters or location characterizes the constituent break. In addition, not all coherence shifts signal boundaries between constituent structures, contrary to other theories of discourse (Gernsbacher, 1990; Zacks & Magliano, 2011; Zacks et al., 2009). For example, some character changes result in two panels that belong to the same constituent (Cohn, 2015b). Thus, in VNG semantic coherence relationships correlate with constituent structures, but do not exclusively motivate breaks between structures. This correlative relationship is made explicit because of the unambiguous separation of the narrative grammar and semantics in VNG (Cohn, 2013b; Cohn et al., 2012).

Recent research has provided evidence that comprehension of constituent structures does not exclusively rely on linear coherence relationships. We followed the logic of the classic "click experiments" from psycholinguistics, which found greater costs to recall and comprehension for disruptions placed within syntactic constituents of sentences than those placed between constituents (Fodor & Bever, 1965; Garrett & Bever, 1974). Similarly, we measured participants' event-related brain potentials to visual narratives in which blank white disruption panels were inserted either between narrative constituents or within the first or second constituent (Cohn et al., 2014). A left-lateralized anterior negativity was greater to disruptions within constituents than between constituents, consistent with anterior negativities shown previously to violations of syntax in language and music (Hagoort, 2003; Neville, Nicol, Barss, Forster, & Garrett, 1991; Patel, 2003; Patel, Gibson, Ratner, Besson, & Holcomb, 1998).

In this experiment, high proportions of shifts in characters and spatial location did indeed fall at narrative constituent boundaries (reported in Cohn, 2012). If such situational changes cued the break between constituents, as predicted by theories focusing on coherence shifts, then a comprehender would need to reach the panel *after* the constituent break, where that situational change would manifest. Yet, we observed larger amplitude left anterior negativities to disruptions within the first constituent compared to those between constituents—and these disruptions occurred *prior* to crossing the boundary where a coherence shift would be made. This suggests that participants predicted the upcoming constituent structure based on the content of panels preceding the disruptions, and did not rely on changes in coherence as a signal for them. Indeed, such semantic shifts had not yet been reached.

Given these findings, it is important to clarify just what type of "hierarchy" or "structure" is emphasized in theories of visual narrative (and discourse) comprehension. The assumption in many models (stated or unstated) has been that "structure" is a uniform phenomenon. However, as emphasized by Jackendoff (2002), all components of the linguistic system may use combinatorial (i.e., hierarchic) structures. Thus, when discourse theories emphasize the "build up of structure" in terms of coherence shifts (e.g., Gernsbacher, 1990), it may reflect a hierarchy intrinsic to semantics and event structures (e.g., Asher & Lascarides, 2003; Bateman & Wildfeuer, 2014; Cohn, 2015b; Jackendoff, 2007; Kintsch, 1988, 1998; Radvansky & Zacks, 2014) rather than to the constituent structure of a narrative grammar. This would be consistent with the finding that the amplitude of the N400 effect—a brainwave response thought to index the activation state of an incoming stimulus in *semantic* memory (Kutas & Federmeier, 2011)—is attenuated across the ordinal position of coherent sequential images (Cohn et al., 2012). However, the N400 is not sensitive to the presence of the narrative grammar (Cohn et al., 2012), and our study on constituent structure observed neurocognitive responses to the violation of the narrative constituents in visual sequences (Cohn et al., 2014) typically seen to violations of

syntax—i.e., left anterior negativities and P600s (e.g., Hagoort, 2003; Patel, 2003)—reinforcing that these are separate systems.

With two hierarchic systems, we should expect to find mutual interfaces between them in predictable ways, with coherence shifts marking a surface structure of the semantics (i.e., events) that maps to particular constructs in the constituency of the narrative grammar (Cohn, 2015b). Such a mapping would be consistent with the interface between semantics and syntax at the sentence level, which optimally—but not always—maintains an isomorphic relationship (Culicover & Jackendoff, 2005; Jackendoff, 2002). This relationship thus predicts that coherence shifts would align with breaks in narrative constituent structure, though would not determine such boundaries alone.

Even though our prior work has provided evidence that constituent structures do not solely rely on breaks in linear coherence, the explicit relationship between coherence relations and this narrative grammar remains unexplored. Prior studies of the relation between coherence shifts and "structure" in discourse have often relied on "segmentation tasks" first used by Newtson and colleagues (Newtson, 1973; Newtson & Engquist, 1976) to study event comprehension. This methodology has generally presented event sequences or visual narratives (drawn or filmed) to participants and asked them to segment such representations where one event ends and another begins. Subsequent segmentation tasks have been deployed using both "offline" and "online" methods (Mura, Petersen, Huff, & Ghose, 2013), which differ based on whether stimuli are presented as static or temporally successive representations.

"Offline" segmentation tasks often present participants with whole visual or verbal narratives, and then are asked to locate the breaks in structure (Gernsbacher, 1985; Kurby & Zacks, 2012). For example, Gernsbacher's (1985) original study of visual narrative asked participants to draw lines between static sequential images that marked the end of one episode and the beginning of another. In contrast, "online" segmentation tasks ask participants to actively segment visual narratives and events that unfurl temporally. This requires the segmentation task to occur concurrently to participants' comprehension of the narrative. For example, online segmentation tasks have been used to explicitly examine segmental structure and coherence relations in filmed narratives (Huff et al., 2014; Magliano et al., 2001; Magliano & Zacks, 2011; Zacks et al., 2009; Zacks et al., 2010), and comparable tasks have also been successful in showing hierarchic relationships between coarse- and fine-grained segmentations of event structure (Zacks, Braver, et al., 2001; Zacks & Tversky, 2001; Zacks, Tversky, & Iyer, 2001).

In our study, we used an offline segmentation task that expanded on the methodology in Gernsbacher (1985) to investigate the relative influences of narrative categories and coherence relationships on the segmentation of visual narrative constituent structure. Participants were given whole visual narrative sequences and asked to draw a line between panels that would divide a sequence into two parts that could make sense on their own. They then continued segmenting the sequence until all "panel bigrams" had been divided. Participants numerically labeled each of their segmentations in the order that they were made. Following the logic of classic psychological experiments on story structure (e.g., Gee & Grosjean, 1984; Gee & Kegl, 1983; Mandler, 1987), we assumed that the initial segmentation of a narrative sequence reflected the maximal constituent structure (i.e., topmost node in a tree structure), with each subsequent division reflecting an additional substructure. In addition, we expected that panels with close relations (e.g., within a constituent and/or with little continuity changes) would be segmented later than those with looser relationships (e.g., at the boundary between constituents and/or with

coherence shifts), which should be preferred as initial segmentations. Thus, participants' preferred order of segmentations should reveal intuitions for the internal structure of sequences.

We then compared participants' segmentations using regressions that analyzed the properties of each panel bigram in a sequence (five bigrams for six panels in each sequence), which included predictors of the expected boundary, narrative categories on both sides of a panel bigram, and coherence relations between panel bigrams. Similar methods have been used to examine the predictors influencing the segmentation of films (Zacks et al., 2009) and video games (Magliano, Radvansky, Forsythe, & Copeland, 2014) across various types of semantic relationships, yet no studies have previously included narrative category information as in VNG. If participants segment panels on the basis of linear changes in coherence, such as changes in characters or location, it would support prior work showing that semantic shifts signal breaks between structures (Magliano et al., 2001; Magliano & Zacks, 2011; Zacks et al., 2009). However, aspects of the narrative grammar should also provide cues for segmentation. For example, panel bigrams that use "illegal" strings of narrative categories (ex. *Peak-Establisher*)—within an otherwise well-formed sequence—should be cues for constituent breaks on the basis of narrative structure, whether or not they feature a shift in coherence (Cohn et al., 2014). Thus, we hypothesized that, as expected by VNG, both narrative categories and coherence relations would strongly predict participants' assessment of constituent boundaries, but that the narrative grammar would be more predictive of participants' segmentations.

2. Materials and Methods

2.1. Stimuli

Coherent graphic sequences were constructed using black and white panels from the *Complete Peanuts* volumes 1 through 9 (1950-1968) by Charles Schulz. In order to eliminate any effects of written language, we only used panels without text, or deleted the text from panels. All created sequences were six panels in length. Standard daily *Peanuts* strips are four panels long, whereas Sunday strips range in length between five and twelve panels long. We therefore deliberately created 332 novel, narratively coherent sequences by combining existing panels from different daily strips, by combining novel panels created by editing existing panels, or by deleting panels from existing Sunday strips. Some sequences were designed with no particular constituent structures in mind (i.e., not aiming to have particular grammatical patterns), yet others were created to test specific grammatical patterns. Subsets of these sequences have appeared in several other studies of visual narrative comprehension where they were all rated as narratively and semantically comprehensible (Cohn & Paczynski, 2013; Cohn et al., 2012), including in studies examining constituent structure (Cohn et al., 2014).

2.1.1. Coding of narrative structure

Two researchers experienced in the constructs of VNG coded the narrative and semantic characteristics of our stimuli. Coding was done collaboratively in a direct dialogue, with disagreements discussed until they were resolved. We coded the predicted narrative constituent structures of all sequences using theoretical diagnostic tests (deletion, movement, sliding window) outlined by VNG (see Cohn, 2013b, 2014a), and now described in a "tutorial" via Cohn (2015a). For example, a "sliding window test" assessed the well-formedness of only a 3-panel "window" of a sequence, while omitting the other panels. Because constituents should form a whole grouping, windowed sequences should be more comprehensible when comprising whole

constituents or parts of constituents than if they cross constituent boundaries (i.e., contain portions of one constituent and portions of another). Thus, a 6-panel strip would first analyze panels *1-2-3*, then *2-3-4*, then *3-4-5*, and *4-5-6*. If both *1-2-3* and *2-3-4* were deemed well-formed, but *3-4-5* and *4-5-6* were not, we might conclude that the break between constituents was located between panels 4 and 5, since this panel bigram existed within both less-felicitous strings.

These theoretical diagnostics were combined with empirical findings from an earlier study where participants made a single segmentation (participants only drew a line between maximal boundaries) of two-constituent sequences (Cohn et al., 2014). We thus identified an "expected boundary" as our predicted break between constituents, given these diagnostic tests and prior empirical data. Panel bigrams with a maximal "expected boundary" between constituents were coded with a "1", whereas subsequent divisions between nodes were coded with "2" or "3." In contrast with sequences containing only a single constituent break (e.g., Figure 2a and 2b), sequences with multiple constituents used several constituent breaks (e.g., Figure 2c and 2d). For these stimuli, we therefore assigned the "expected boundary" by ordinal position in the sequence (e.g., the first boundary was coded as "1", the second was "2", etc.).

Across all stimuli, many different patterns of narrative constituent structure were used. We focus here on constituents built of the core narrative schema, excluding modifiers and constructional patterns which carry additional predictions for the relations between linear coherence shifts and hierarchic structure (e.g., Cohn, 2015b). We chose several consistent patterns of constituent structure. Three major patterns all used two constituents, and varied depending on the location of the constituent boundary in the sequence. For example, "2-4" strips featured a constituent boundary between the second and third panels, thereby grouping the first two panels ("2") into a constituent and the last four panels ("4") into a constituent (as in Figure 1). Two constituent patterns included 3-3 strips (55), 4-2 strips (54), and 2-4 strips (51). Sequences with three constituents often used a center-embedded clause, where one fully-formed "embedded clause" was placed within another "matrix" sequence. This structure can be tested by separating the sequences to see whether the embedded and matrix clauses could stand alone. These sequences included 2-3-1 strips (34), 2-2-2 strips (16), 3-1-2 strips (15), 2-1-3 strips (14), and 3-2-1 strips (10). Other two and three constituent patterns had less than 10 strips per pattern. Less frequent sequences included left-branching structures and other complex patterns with multiple embedded constituents, however, for simplicity, our analysis focused on the aforementioned sequences with two and three constituents where primary constituent boundaries were expected to be most apparent (250 strips total; 75% of all sequences).

Figures 1 and 2 depict example sequence patterns. Figure 1 depicts a 2-4 strip, as discussed above. In Figure 2a, a 3-3 strip, Linus tears up paper in a first constituent (Initial), which is then played with, and lands on the head of, Snoopy (Peak) in the second constituent. Figure 2b, a 4-2 strip, shows Schroeder playing in sand until it gets too hot (Initial constituent), so he builds a sand pile so he can hide in the shade (Peak constituent). Figure 2c uses an embedded clause in a 2-3-1 pattern where Linus runs to catch a baseball hit in the air (matrix clause), but only before making a pit-stop to build a sandcastle (embedded clause). Finally, Figure 2d uses a 3-2-1 pattern where Charlie throws a newspaper (Initial constituent), which is retrieved by Snoopy and strewn all over the road by his sneeze (Peak constituent), only to have Charlie continue on his paper route oblivious to the mess caused by Snoopy (Release).

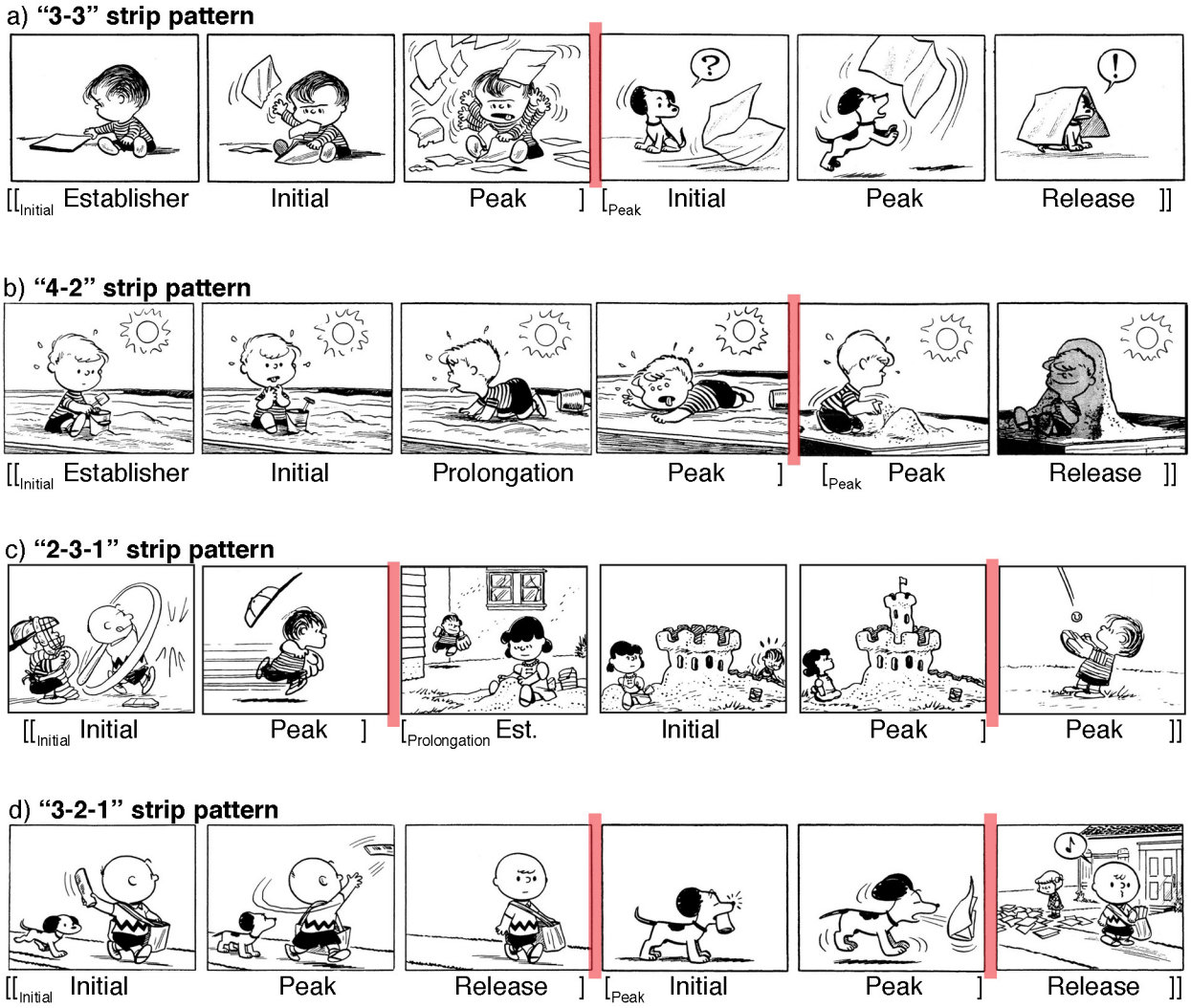

Figure 2. Various constituent structure patterns in visual narratives, with lines highlighting the breaks between constituents. Sequences (a) and (b) contain two constituents, whereas (c) and (d) both have three constituents, with a center embedded clause. Note that only some of the constituent boundaries align with changes in spatial location (a, c, d) and/or characters (a, c, d), but (b) has no major coherence changes between constituents. *Peanuts* is © Peanuts Worldwide LLC.

In addition to constituent structures of the sequences, we coded panels' narrative categories based on both their semantic content and their context in the sequence, again using theoretical diagnostic tests (Cohn, 2013b, 2014b). For example, deletion of some narrative categories (Peaks, Initials) renders a sequence less understandable, but omission of others (Establishers, Prolongations, Releases) is more acceptable (Cohn, 2014b). In addition, Peaks, but not any other category, can felicitously be substituted for an "action star" panel, which depicts a star-shaped "flash" commonly associated with impacts (Cohn, 2013a; Cohn & Wittenberg, 2015). Meanwhile, Releases, but not any other category, can have the phrase "Jeez, what a jerk!" added

as a speech balloon and retain the coherence of the sequence (Cohn, 2013a, 2013b; Sinclair, 2011). Additional diagnostics can be found in Cohn (2015a).

We recorded the narrative categories for each side of a panel bigram (i.e., first panel in a bigram or second panel in a bigram). Table 1 outlines the proportion of panel bigrams where each category appeared in first or second position for our 249 analyzed sequences. Note that the highest frequencies conform to the bigrams in the canonical sequence order of E-I-P-R, i.e., the bigrams of E1/I2 (i.e., E1 = Establisher as the first panel of a bigram followed by I2 = Initial as the second panel of the bigram), I1/P2, and P1/R2 (italicized and greyed). This canonical structure is maintained also in that Establishers and Initials appear more often as the first panel of a bigram than the second panel, whereas the reverse is true of Peaks and Releases. Other bigrams may reflect the divisions between constituents, such as P1/E2 reflecting a Peak ending one constituent and an Establisher starting another constituent (as in Figure 1). Altogether these categories constituted roughly 94% of all panel bigrams, with the remaining 6% comprised of various narrative modifiers, here excluded for simplicity.

Table 1. Proportion of all panel bigrams using particular narrative categories in first or second positions (i.e., bigrams with I1 and P2 means the first panel was an Initial and the second panel was a Peak: I-P). Total bigrams in our analysis=1246.

		Second panel of a bigram				
		Establisher (E2)	**Initial (I2)**	**Peak (P2)**	**Release (R2)**	**Total**
First panel of a bigram	**Establisher (E1)**	0.019	*0.157*	0.013	0.0	**0.19**
	Initial (I1)	0.003	0.026	*0.313*	0.008	**0.35**
	Peak (P1)	0.052	0.060	0.039	*0.180*	**0.33**
	Release (R1)	0.012	0.035	0.016	0.012	**0.075**
	Total	**0.087**	**0.277**	**0.38**	**0.20**	

It should also be noted that our segmentation task dividing panel bigrams yields an inherently binary branching structure. VNG does not exclusively predict a binary branching structure, but rather uses a flat structure more consistent with syntactic models from construction grammar (Culicover & Jackendoff, 2005; Goldberg, 1995; Jackendoff, 2002). Nevertheless, we believed that a binary division could inform us both about the broader intuitions of constituent structures in a whole sequence and about relationships between categories within a constituent.

2.1.2. Coding of coherence relationships

Coherence changes were coded along three salient dimensions discussed in the event-indexing model (Zwaan & Radvansky, 1998): characters, causality, and spatial location. We considered coherence shifts to be non-mutually exclusive (i.e., panel relationships could have multiple coherence changes) and non-exhaustive (i.e., coherence changes could be both full and partial). This granularity was important because, in VNG, degrees of changes may predict different types of processing. For example, partial changes in characters (characters are added or omitted between panels) would be expected to incur costs of updating a mental model (consistent with various discourse theories), and possibly constituent breaks. However, they would not be

expected to signal modifying "grammatical constructions" that may arise from full changes between characters (e.g., Cohn, 2015b).

Changes in characters were coded as a "1" for a complete change in characters between panels, with ".5" for a partial change (i.e., characters held constant but others added or omitted), and no change in characters was coded as "0." Changes in spatial location were coded as "1" for complete changes in location, ".5" for partial changes (such as changes within a common space, such as moving from one room to another in the same building), and "0" for no changes in location. Shifts in causation were coded as "1" where the events depicted in one panel were directly caused by the events in a prior panel (i.e., depicted the direct effect of the prior panel's events; ex. Charlie Brown falling because Lucy pulls a football away from being kicked), "0" for no causal relations, and ".5" for causal relations that were not related to full actions (an action in one panel did not cause a full action in another, but led to a change in a character's emotional state; ex. Lucy scowling after Snoopy rolls by on roller-skates). We considered this difference between causal changes as reflecting modulated degrees of intensity rather than as "partial" or "full" in the sense of shifts between characters or locations.

The proportion of coherence relations across all sequences by ordinal panel position is provided in Table 2, including both full and partial coherence shifts. Across all panel relationships, more bigrams showed changes in characters (34%) than causal shifts (25%) or changes in spatial location (23%). These shifts were most pronounced at bigram 2-3, which is consistent with the idea that coherence shifts signal constituent boundaries: Nearly half (115 of 249; 46%) of all analyzed sequences had a constituent boundary at bigram 2-3 (2-4 strips, 2-3-1 strips, 2-2-2 strips). The high proportion of causal changes at bigram 5-6 is also consistent with VNG: A final sequence panel is often a Release, which will prototypically depict the aftermath of the events in the prior panels (i.e., a causal change). In addition, coherence changes did align with the topmost expected boundary between constituents. Expected boundaries typically used both character changes (50%), and spatial location changes (35%), but far fewer causal shifts (16%).

Table 2. Proportion of all panel bigrams across ordinal sequence featuring shifts in characters, spatial locations, and causal relations. Total bigrams = 1246.

Coherence Changes	**Bigram 1-2**	**Bigram 2-3**	**Bigram 3-4**	**Bigram 4-5**	**Bigram 5-6**	***Total***
Characters	0.053	0.085	0.066	0.063	0.075	**0.342**
Spatial Location	0.028	0.058	0.046	0.039	0.055	**0.225**
Causation	0.034	0.050	0.040	0.051	0.079	**0.254**

Finally, we compared all the main predictors of our analysis using correlations. Table 3 depicts the r-values between our predictors. Note that these values do not necessarily reflect frequencies of panels within bigrams, but rather correlations between panels found in bigrams throughout sequences. For example, the bigram E1/I1 never occurs because no bigram can have two panels in its first position. However, a panel bigram of E1/I2 may precede a bigram starting with that same Initial (I1), resulting in a correlated relationship between E1 and I1. Yet, not all I1's will first be an I2, as in sequence-starting Initials. It should be immediately apparent that nearly all predictors correlated significantly with each other. Of particular interest, the "expected boundary"—our predicted break between major constituents—correlated significantly with all

predictors except causal changes, which trended towards significance (p=.096). Peaks and Releases as the first panel of a bigram, Establishers and Initials as the second panel of a bigram, and character and spatial location changes all positively correlated with the expected boundary. Establishers and Initials as the first panel of a bigram, Peaks and Releases as the second panel of a bigram, and causal changes were all negatively correlated with the expected boundary. Also worth noting is that the bigrams reflecting a canonical narrative schema (*E-I-P-R*, highlighted in grey) are the most highly correlated of all bigrams.

Table 3. Correlation coefficients between all predictor variables used in the regression analysis. Again, bigram pairs reflecting a canonical narrative schema are highlighted in grey. "Expected boundary" was the predicted major boundary between constituents. Total bigrams = 1246, Bold = $p < .05$, Italics = $p < .1$

	Expected Boundary	E1	I1	P1	R1	E2	I2	P2	R2	Character change	Spatial change
E1	**-0.18**										
I1	**-0.33**	**-0.37**									
P1	**0.30**	**-0.34**	**-0.54**								
R1	**0.34**	**-0.14**	**-0.22**	**-0.20**							
E2	**0.47**	0.02	**-0.21**	**0.17**	**0.07**						
I2	**0.17**	**0.59**	**-0.36**	**-0.15**	**0.11**	**-0.19**					
P2	**-0.28**	**-0.33**	**0.70**	**-0.41**	**-0.11**	**-0.25**	**-0.51**				
R2	**-0.20**	**-0.26**	**-0.36**	**0.55**	-0.04	**-0.16**	**-0.33**	**-0.34**			
Character changes	**0.32**	**-0.08**	**-0.19**	**0.19**	**0.15**	**0.25**	0.00	**-0.19**	**0.06**		
Spatial changes	**0.27**	**-0.11**	**-0.14**	**0.17**	**0.10**	**0.21**	0.00	**-0.15**	0.04	**0.39**	
Causal changes	*-0.05*	**-0.20**	-0.01	**0.22**	**-0.08**	**-0.09**	**-0.17**	0.04	**0.24**	-0.01	-0.05

2.2. Participants

We recruited 54 experienced comic readers (27 male, 27 female, mean age 22.9) from the Tufts University community, who were paid for their participation. All participants gave informed written consent according to Tufts University's Human Subjects Review Board guidelines. We assessed participants' experience reading comics by using the "Visual Language Fluency Index" (VLFI), which generates a "fluency score" based on participants' answers from a pretest questionnaire that asked them to rate their habits for reading and drawing various types of visual narratives (for details, see Cohn et al., 2012). VLFI scores correlate with both behavioral and neurophysiological effects in online comprehension of visual narratives (e.g., Cohn & Kutas, 2015; Cohn et al., 2012). In this metric, "average" fluency falls at 12, with low fluency below 8 and high fluency at or above 22. Participants had a wide range of VLFI scores (low=4.38, high=35.38), but had an "average" mean fluency of 14.35 (SD=6.24). Data from one participant was excluded from analyses due to their not properly carrying out the task.

2.3. Procedure

Participants were given a stack of paper, where each sheet depicted an experimental sequence. Because of time restraints, participants only viewed half of the 332 overall stimuli

sequences, roughly 165 sequences each. The order and choice of sequences shown to each participant were randomized.

We first asked participants to draw a line between panels where they thought the strip could best be divided into two sections that still made sense on their own. Next, we asked them to continue dividing the remaining segments into smaller pieces that "made sense" until all panel breaks had been segmented. To assess the order that participants drew each line, we asked them to label each division with a number, such that the first division was marked as "1" and subsequent divisions were labeled up to "5." Participants were told that there were no right or wrong answers, and to go with their first instinct. After finishing the segmentation task, participants answered a short questionnaire where they rated how difficult they found the task overall, and at each individual division (divisions 1 through 5) on a 1 to 5 scale (1=easy, 5=difficult). We also asked them to describe any conscious strategies they used in choosing their divisions. On average, participants took roughly 45 minutes to an hour to complete the task, depending on the number of stimuli that they viewed.

2.4. Data analysis

Because time restrictions allowed participants to view only half of the 332 overall stimuli sequences, each item was viewed by between 25 and 29 participants (mean: 27.28). For each sequence, we recorded the order of divisions made by each participant. Our analysis focused on participants' first and second segmentations of sequences that had only two or three constituents and had more than 10 strips per sequence pattern (see above).

Our primary analysis followed those in other studies of segmentation (Magliano et al., 2014; Zacks et al., 2009), which used separate logistical regressions on each participant's data, as developed by Lorch and Myers (1990). This methodology allowed us to address the question of "What properties do participants use as cues to segment visual narrative sequences?", as opposed to a question of "What properties do constituent breaks have?" that would be addressed by a single regression collapsing across participants. The dependent variable was the participant's segmentation for a particular panel bigram (a binary 0/1 assessment, "0" if they did not segment a bigram, "1" if they did bigram a segment), meaning that each strip contributed five datapoints for each bigram (panel break) in a 6-panel sequence. Predictor variables included the expected boundary, narrative categories (categories appearing either first or second in a bigram), and coherence relations (shifts in characters, space, or causation), with each predictor coded as "1" if that variable was used by a given panel bigram, or "0" if it was not (or ".5" where appropriate for coherence relations). This analysis yielded b-weights for each predictor for each participant. We extracted these b-weights from the regression analyses and compared them against 0 using a t-test to determine whether each predictor was significant. Following this, we used a one-way, repeated measures ANOVA to assess the relative influence of each predictor against each other, along with follow up t-tests between the b-weights of each predictor.

In addition, responses to post-experiment questionnaires were analyzed with a subject's analysis averaging participants' ratings for each segmentation's difficulty (1=easy, 5=difficult). Participants' descriptions for conscious strategies of segmentation were coded for terms describing changes in linear coherence ("I looked for scene changes and new characters"), event knowledge ("one event ended and another began", "cause and effect") or narrative structure ("punch-line panels"). Data from three participants were excluded from this analysis due to not completing the questionnaire. The included participants' data were analyzed using repeated-measures ANOVAs, followed by t-tests to compare pairwise relations between strategies.

Finally, to assess any possible influence of participants' comic reading frequency on these results, both b-weights of predictors from the regression analyses and participants' difficulty ratings were correlated with VLFI scores using a Pearson's correlation set to .05.

3. Results

3.1. Segmentation

We first report whether participants' segmentations corresponded to the location of boundaries predicted by VNG, and the consistency of those segmentations between participants. Overall, a modest proportion of participants chose our expected boundary as their first segmentation (44%), though this well exceeded the threshold of chance (20% = 1 out of 5 panel bigram possibilities), $t(52)=15.3$, $p<.001$. This was comparable to the proportion of participants (47%) who shared the most common first segmentation for a sequence (i.e., the mode for first segmentation), regardless of our expected boundary, which also exceeded the 20% threshold of chance, $t(52)=19.6$, $p<.001$.

We next report our primary analysis, which used regressions on each participant's choice of first segmentation to examine the predictors of the expected boundary, narrative category bigrams, and coherence shifts. B-weights were produced by each regression and then averaged across participants. Mean b-weights are depicted in Figure 3. A t-test showed that, at the expected first segmentation boundary, Establishers and Initials as the second panels of a bigram, character changes, and spatial changes were all significant as positive predictors of a segmentation (all $ts > 6.4$, all $ps < .001$). Establishers, Initials, and Releases as the first panels of a bigram, Releases as the second panel of a bigram, and causal changes were all significant negative predictors of segmentation (all $ts < -2.8$, all $ps < .01$).

A repeated-measures ANOVA confirmed that these b-weights for predictors were all significantly different from each other, $F(11,572)=53.9$, $p<.001$. Establishers as the second panel of a bigram were significantly more influential than all other positive predictors (all $ts > 2.3$, all $ps < .05$), followed by the expected boundary (all $ts > 2.5$, all $ps < .05$). Initials as the second panel of a bigram were more influential than Peaks of either bigram position (all $ts > 3.5$, all $ps < .005$), but did not differ from character or spatial changes (all $ps > .128$). Character and spatial changes were also larger than Peaks of either bigram position (all $ts > 2.5$, all $ps < .05$), but Peaks did not differ from each other ($p=.828$). Of the negative predictors, Establishers and Initials as the first panel of a bigram and Releases as the second panel of a bigram were significantly more negative than Releases as the first panel of a bigram and causal changes (all $ts > 2.7$, $p < .01$), but did not differ from each other ($p>.097$). Releases as the first panel of a bigram were also more influential than causal changes, $t(52)=2.0$, $p<.05$.

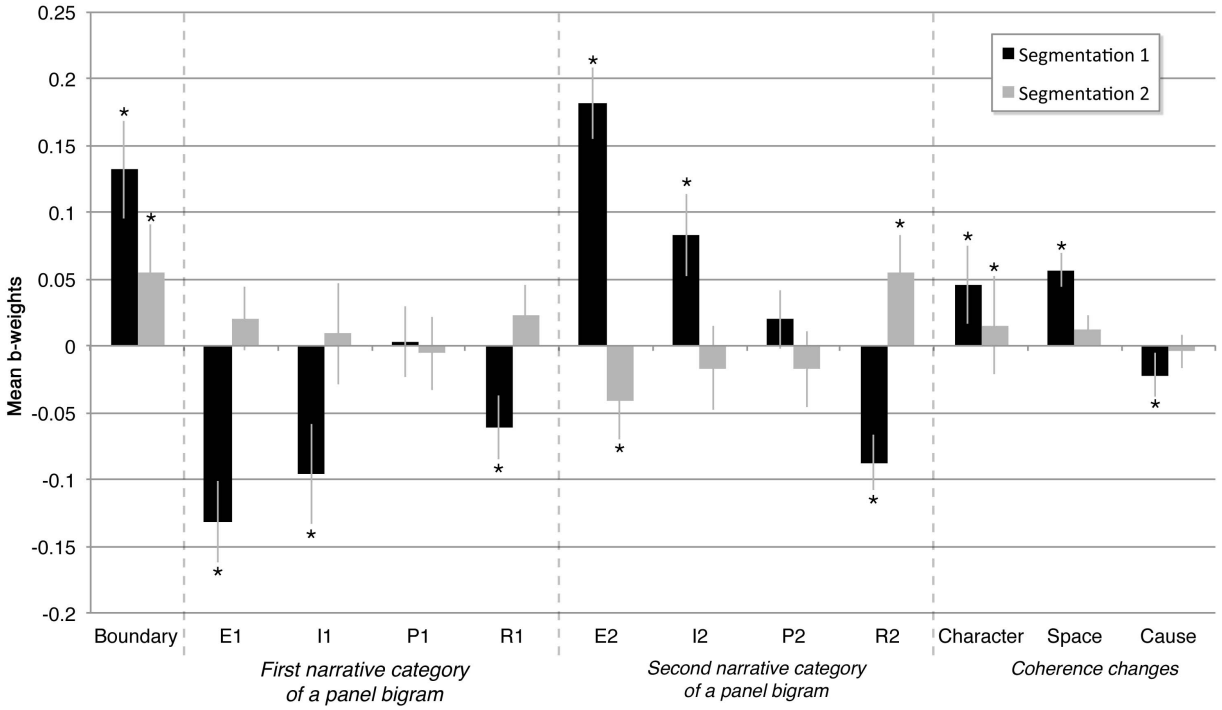

Figure 3. Mean b-weights from regressions averaged across participants depicted for predictors related to participants' mean agreement for first and second segmentations of visual narrative sequences. Error bars show standard deviation. N=53; * = $p < .05$.

Figure 3 also depicts the b-weights for each predictor across participants' second segmentation. Here, the expected second segmentation boundary, Releases as the second panel of a bigram, and character changes all were significant positive predictors (all $ts > 2.5$, all $ps < .05$). Only Establishers as the second panel of a bigram were a significant negative predictor, $t(52)=-2.5$, $p<.05$. When compared against each other, all predictors significantly differed, $F(11,572)=4.9$, $p<.001$. The expected boundary and Releases as the second panel of a bigram were more influential than all other predictors (all $ts > 1.9$, all $ps <.06$), but did not differ from each other ($p=.877$). No other negative predictors differed from each other (all $ps > .47$).

Because several of the factors in our regression were highly correlated (for example, because of the ordering of the canonical narrative schema, certain categories will naturally precede or follow each other), we sought to investigate any issues with multicollinearity in our data. Following the method of our regression analysis, we calculated the variance inflation factors (VIF) for each participant's regression, and again averaged them across participants. Though VIFs were expectedly high for categorical information (Table 4), they did not exceed the recommended level of 10.

Table 4. Variance inflation factors for all independent variables in the regression analysis.

	Segmentation 1	Segmentation 2
Expected Boundary	2.65	1.10
E1	7.59	6.83
I1	9.41	9.06
P1	8.07	8.08
R1	3.37	3.35
E2	3.25	3.06
I2	6.16	6.02
P2	5.53	5.55
R2	5.19	4.74
Character change	1.30	1.30
Spatial change	1.26	1.26
Causal change	1.12	1.12

Finally, we considered the impact of "visual language fluency" on participants' segmentations. For the first segmentation, VLFI scores approached significance for positively correlating with Peaks in the first panel of a bigram, $r(51)=.246$, $p=.076$, and significantly correlated with character changes, $r(51)=.308$, $p<.05$. Both correlations suggested that these predictors were more influential for participants with higher fluency scores. For the second segmentation, negative correlations appeared between VLFI scores and Establishers, Initials and Peaks as the first panel of a bigram (all $rs < -.270$, all $ps < .051$), suggesting that these were all less influential for higher fluency participants.

3.2. Participant assessments

Next, we report participants' assessments for the difficulty of the segmentation task. Using a 1 (=easy) to 5 (=difficult) scale, participants reported that their choices for segmentations were not overly difficult, with an overall mean of 3.12 (1.02). Participants found the first segmentation of bigrams to be the most easy, 1.63 (.69), with each subsequent segmentation becoming progressively more difficult until the final division: S2: 2.53 (.95), S3: 3.51 (.78), S4: 3.79 (1.02), S5: 3.36 (1.5). These ratings differed across divisions, $F(4,196) = 39.14$, $p < .001$, with all segmentations significantly different from each other (all $ts > 3.06$, all $ps < .005$) except for a trending significance between segmentations 3 and 4, $t(49) = 1.86$, $p = .07$, and no significance between segmentation 3 and 5, $t(49) = .493$, $p = .624$.

Further examination revealed that first-segmentation difficulty ratings negatively correlated with participants' VLFI scores, $r(48) = -.329$, $p < .05$, suggesting that participants with higher fluency considered their first segmentation to be easier than did those with less fluency. Although correlations were not significant after the first segmentation, we observed an interesting trend across correlation coefficients: The r-values for the correlation between segmentation and VLFI score increased with each segmentation, following the pattern of the difficulty ratings (1st line: -.349, 2nd line: -.147, 3rd line: .086, 4th line: .25, 5th line: .173). We interpreted this trend as suggesting that fluency became less advantageous for sequence-segmentation the further participants progressed in the task (where they had fewer choices for where to draw segmentation lines).

Finally, we also examined how participants explained their choices in the segmentation task. More participants consciously explained their decisions for segmentations by describing aspects of the start and end of events (51%) and coherence relations (49%) than purely narrative aspects of the structure (20%). Narrative explanations were used significantly less than strategies relying on coherence relations and events (all ts > 3.1, all ps < .005), which did not differ from each other, $t(49) = .198$, $p = .844$.

4. Discussion

This study investigated the factors that influenced participants' intuitions about the segmental structure of visual narratives. First, despite the modest agreement for first segmentations as a whole, the regression analysis suggested that expected boundaries were strong predictors of both first and second segmentations. Next, both narrative category information and coherence relations predicted segmentation, though categorical information was a consistently stronger predictor than linear coherence changes for both segmentations. Despite the greater influence of narrative categories, participants' conscious explanations for segmentations focused more on the linear semantic changes, consistent with previous findings of filmed visual narratives (Magliano et al., 2001; Magliano & Zacks, 2011; Zacks et al., 2009). These results suggest that segmentation of visual narrative sequences relies more on *narrative* structure, despite it being less consciously accessible than semantic features. However, overall, these results support the claim that two processing streams of narrative and semantics contribute to the whole understanding of constituent structures in sequential images. Below, we discuss these findings in more detail.

Overall, we found a modest agreement (44%) across participants for segmenting our expected boundary. This proportion exceeded the threshold of chance (20% = 1 out of 5 panel bigram possibilities), suggesting that participants shared intuitions for the division of sequences. However, this proportion was noticeably lower than found in our prior work (71%) for 135 of these 250 analyzed stimuli (Cohn et al., 2014). However, this prior study asked only for a single segmentation of two-constituent sequences, whereas this project asked for repeated segmentations of variable sequence patterns. It is possible that participants in the present study were more flexible about their first segmentation, knowing that they could also choose other bigrams on subsequent segmentations. Finally, we here coded only a single expected boundary (the first in the ordinal sequence). Yet, for the 90 three-constituent sequences there were two feasible initial boundaries (i.e., the two boundaries dividing the three segments), meaning that selection of the alternate boundary in these cases may have lowered the overall agreement.

Our regression analysis more clearly illustrated the factors influencing segmentation. Narrative category information most predicted the segmentation of the visual sequences. Establishers and Initials after the segmentation (second panel in a divided bigram) were more influential than any other predictor besides the expected boundary (which had mean b-weights falling between these predictors). Because Establishers and Initials typically begin a narrative schema, their presence as the second panel in a divided bigram is indicative of their starting a new constituent. The opposite finding occurred in the reversed stepwise pattern for Establishers and Initials as negative predictors when occurring as the first panel in a bigram. Because participants would not choose these panels to end a constituent (as the first panel of a divided bigram), they most negatively predict segmentation choices. In addition to Establishers and Initials, participants dispreferred Releases both before and after the boundary (second panel of a bigram). Altogether, the results for all categories reflect the canonical order of narrative

sequences (E-I-P-R), maintaining categories towards the front of the arc as the start of new constituents (E, I), whereas the categories towards the end of the arc finish constituents (P, R). Thus, these segmentation results provide further evidence for the presence of narrative categories stored in a canonical order.

In the second segmentation, none of the mean b-weights were as strong as in the first segmentation, but the strongest predictor was Releases as the second panel of a bigram. This division of a Release may align with the fact that our coding revealed that bigrams in position 5/6 had fairly substantial numbers of coherence shifts. Since Releases often end narrative schemas, this coherence shift may thus have been a cue for this second segmentation. However, although Releases as the second panel of a bigram correlated with causal changes ($r = .24$, $p < .001$) and somewhat with character changes ($r = .06$, $p = .04$), they were not significantly correlated with spatial location changes ($r = .04$, $p = .2$). Furthermore, the b-weights for the Release as the second panel of a bigram were still larger than for all coherence relations for the second segmentation. This suggests that coherence relations alone were not the primary motivator of this segmentation, though they may have factored into that decision.

Other possible reasons for this segmentation of second-panel Releases may rely on narrative structure and/or the prototypical semantic content of Release panels. First, some sequence patterns might separate a Release into a second constituent (as in many 2-3-*1* sequences), meaning that this panel was part of an actual boundary, not a segmentation made within a constituent. Second, if they were within a constituent, it may reflect a distancing of Releases from the remaining categories within a narrative schema. This is consistent with the idea of Releases being one of the "peripheral" categories of a narrative schema (along with Establishers and Prolongations) compared to the "core" categories of Initials and Peaks (Cohn, 2014b). Such separation may also align with observations that the endpoints of paths are more salient than starting points, whether in language, perception, and attention (Lakusta & Landau, 2005; Regier, 1996, 1997). Because endpoints of actions—which are prototypical of Releases—should be emphasized in a situation, participants choose to segment a narrative schema that individuates these panels.

Semantic coherence relations between panels also influenced participants' segmentations. Participants significantly relied on changes in characters and spatial location—but not causal shifts—to influence their first segmentations. Second segmentations also used changes in characters, but less so than the first segmentation. These first-constituent segmentations are consistent with the idea that changes in referential coherence (characters, location) may align with breaks in narrative constituents (Gernsbacher, 1990; Zacks & Magliano, 2011; Zacks et al., 2009), whereas causal actions likely correspond to the internal structure of constituents, where characters progress through actions. Thus, participants likely do use major coherence shifts as breaks between constituent structures, while building up "structure" (i.e., semantic coherence, motivated by causal relations) within constituents. Because of this, causal changes did not arise as a significant predictor at constituent breaks, but may be more likely within constituents (though this was not explored by our analysis).

Nevertheless, for both segmentations, these coherence relationships were less predictive than narrative category information. Such results appeared even though most narrative categories in bigrams were proportionally smaller than coherence shifts: Establishers as the second panel of a bigram comprised only 9% of all total bigrams, which was very small compared to the panel bigrams with changes in spatial location (22%) or characters (34%). Yet, Establishers as the second panel of a bigram had average b-weights almost four times larger than spatial location

and character changes. These results confirm that coherence shifts do co-occur with constituent boundaries, and such shifts may factor into participants' segmentations, but that these changes in semantic features are not the primary motivator of structure in visual narratives. Such findings are consistent with previous work where neural responses to disruptions of constituent structure occurred prior to comprehenders reaching shifts in coherence (Cohn et al., 2014), meaning that these brain modulations were due to predictive processing motivated by the content of preceding panels, not crossing a break in semantics. Thus, although semantics do indeed influence and co-occur with breaks in narrative structure, coherence shifts alone do not seem to determine recognition of structural boundaries.

Participants' explanations of their segmentation choices also emphasized semantic aspects of coherence relationships and events. This is particularly important because prior research emphasizing linear coherence relationships have often relied upon participants' conscious recognition of these factors (Magliano, Dijkstra, & Zwaan, 1996; Magliano et al., 2001). The results of the present study do support the theory that coherence relations factor into participants' choices for segmenting visual narrative sequences, but these semantic factors were ultimately less predictive than aspects of narrative structure. However, these narrative structures appeared to be less consciously accessible to participants. Similar findings have been found in previous work where participants reported observations about manipulations to the semantics of visual sequences, but made almost no mention of noticing manipulations to narrative structure (Cohn et al., 2012). That structure seems more "invisible" than semantics may also relate to the longstanding observations that recall for semantic information persists, whereas structure of a narrative rapidly disappears from memory (e.g., Gernsbacher, 1985; van Dijk & Kintsch, 1983).

Accordingly, researchers must thus be sensitive to the abilities of tasks and measurements to capture observations of desired structures. Certain methods may be more effective at assessing the semantics than the narrative structure, and vice versa. For example, given that semantic information is retained in memory, whereas "structural" information is not (e.g., Gernsbacher, 1985; van Dijk & Kintsch, 1983), memory paradigms may therefore not be appropriate for investigating the properties of a narrative grammar. This was the case for most studies of "story grammars" and "scripts" (e.g. Black & Bower, 1979; Mandler & Johnson, 1977; Stein & Glenn, 1979; Thorndyke, 1977), which were criticized for positing "grammatical" constructs that were actually closer to semantics (Black & Wilensky, 1979; de Beaugrande, 1982). Nevertheless, such methods may be useful for detailing aspects of the semantic structure (though not mechanisms of online processing). Similar limitations may also hold for studies relying on conscious assessments of stimuli, including segmentation tasks (as opposed to unconscious processing of manipulated structures). Although they are likely useful for investigating the semantics and event structure of visual narratives, tasks emphasizing conscious awareness of unmanipulated sequences may be unable to address the complexity found in a narrative structure.

On these points, it is worth highlighting some differences in methodology between this study and prior works examining visual narrative segmentation (though their results are not necessarily in opposition). First, unlike the many "online" segmentation tasks that have used filmed narratives or events (Magliano et al., 2001; Magliano & Zacks, 2011; Zacks et al., 2009; Zacks et al., 2010), this study used an "offline" task of drawn visual narratives laid out spatially on a page (as in Gernsbacher, 1985). Participants could thus see the entire sequence all at once, rather than engage it temporally as it unfurled. This meant that participants did not have to negotiate basic processing of the sequence and the segmentation task simultaneously, and instead could assess the whole sequence before making their segmental judgments. In addition, they

could assign preferential importance to some segmentations over others (i.e., "first segmentation" vs. "second segmentation," etc.), which would be much harder with temporally progressing stimuli. It may be possible that offline segmentation increases the salience of the narrative categories, where the global structure can be assessed at once. Meanwhile the online procedure may raise the salience of linear coherence relations, where such broader structure is less accessible. This would again be consistent with our findings here and elsewhere (Cohn et al., 2012) that narrative structure is less consciously accessible than semantic structure. However, since recent work found little difference between online and offline tasks with regard to event segmentation (Mura et al., 2013), it is an open question whether this procedural difference may impact the recognition of narrative and semantic aspects of visual narratives.

Related to this, a second difference is in the phrasing of the task. Previous works have phrased the segmentation task in terms of identifying changes in "events," "activities," or "situations" (Magliano et al., 2001; Magliano & Zacks, 2011; Zacks et al., 2009)—which may be based on semantic criteria (a "situation" being determined by the meaningful parts that occur within it). In contrast, the task here focused on the division of sequences of visible length into constituent parts—a task with no implicit semantics in the instructions. Whether these differences in procedure (online vs. online) or task (implicit semantics vs. sequencing alone) would push participants towards different segmentations would be interesting to investigate in future studies.

4.1. Conclusion

This study investigated the influences on the segmentation of visual narrative sequences, and found that participants relied on cues from both narrative structure and coherence relationships. However, even though participants were more consciously aware of coherence shifts in their segmentations, these semantic relations were less influential than aspects of narrative structure. Thus, coherence shifts may be a "low-level" aspect of semantic comprehension that is tracked across sequences (Magliano & Zacks, 2011), which may align with the "higher level" structural aspects narrative grammar. Coherence shifts are thus not exclusively used as breaks between narrative constituent structures, but rather likely index prototypical correspondences between these processing streams. Altogether, these results further support claims that a narrative grammar and semantics mutually interface to provide the whole of visual narrative comprehension at multiple levels of structure.

Acknowledgements

Gina Kuperberg is thanked for funding this research through grants provided by NIMH (R01 MH071635), NICHD (HD25889) and NARSAD (with the Sidney Baer Trust). Analysis of data and early drafts benefited from insights offered by Ariel Goldberg, Stephanie Gottwald, Ray Jackendoff, Ross Metusalem, Anastasia Smirnova, and Eva Wittenberg. Fantagraphics Books is thanked for their generous donation of volumes of *The Complete Peanuts*.

References

Asher, N., & Lascarides, A. (2003). *Logics of Conversation*. Cambridge: Cambridge University Press.

Bateman, J. A., & Wildfeuer, J. (2014). A multimodal discourse theory of visual narrative. *Journal of Pragmatics, 74*, 180-208. doi:10.1016/j.pragma.2014.10.001

Black, J. B., & Bower, G. H. (1979). Episodes as chunks in narrative memory. *Journal of Verbal Learning and Verbal Behavior, 18*, 187-198.

Black, J. B., & Wilensky, R. (1979). An evaluation of story grammars. *Cognitive Science, 3*, 213-230.

Carroll, J. M. (1980). *Toward a Structural Psychology of Cinema*. The Hague: Mouton

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

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. (2014a). The architecture of visual narrative comprehension: The interaction of narrative structure and page layout in understanding comics. *Frontiers in Psychology, 5*, 1-9. doi:10.3389/fpsyg.2014.00680

Cohn, N. (2014b). 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. (2015a). How to analyze visual narratives: A tutorial in Visual Narrative Grammar. Retrieved from http://www.visuallanguagelab.com/P/VNG_Tutorial.pdf

Cohn, N. (2015b). Narrative conjunction's junction function: The interface of narrative grammar and semantics in sequential images. *Journal of Pragmatics, 88*, 105-132. doi:10.1016/j.pragma.2015.09.001

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

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:10.1016/j.cogpsych.2013.07.002

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

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

Culicover, P. W., & Jackendoff, R. (2005). *Simpler Syntax*. Oxford: Oxford University Press.

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

Fodor, J., & Bever, T. G. (1965). The psychological reality of linguistic segments. *Journal of Verbal Learning and Verbal Behavior, 4*(5), 414-420.

Friederici, A. D. (2002). Towards a neural basis of auditory sentence processing. *Trends in Cognitive Sciences, 6*(2), 78-84.

Garrett, M. F., & Bever, T. G. (1974). The Perceptual Segmentation of Sentences. In T. G. Bever & W. Weksel (Eds.), *The Structure and Psychology of Language*. The Hague: Mouton and Co.

Gee, J. P., & Grosjean, F. (1984). Empirical evidence for narrative structure. *Cognitive Science, 8*, 59-85.

Gee, J. P., & Kegl, J. A. (1983). Narrative/Story Structure, Pausing, and American Sign Language. *Discourse Processes, 6*, 243-258.

Gernsbacher, M. A. (1985). Surface information loss in comprehension. *Cognitive Psychology, 17*, 324-363.

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

Goldberg, A. (1995). *Constructions: A Construction Grammar Approach to Argument Structure*. Chicago, IL: University of Chicago Press.

Hagoort, P. (2003). How the brain solves the binding problem for language: a neurocomputational model of syntactic processing. *NeuroImage, 20*, S18-S29. doi:http://dx.doi.org/10.1016/j.neuroimage.2003.09.013

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. doi:10.1037/0033-2909.133.2.273

Jackendoff, R. (1990). *Semantic Structures*. Cambridge, MA: MIT Press.

Jackendoff, R. (2002). *Foundations of Language: Brain, Meaning, Grammar, Evolution*. Oxford: Oxford University Press.

Jackendoff, R. (2007). *Language, Consciousness, Culture: Essays on Mental Structure (Jean Nicod Lectures)*. Cambridge, MA: MIT Press.

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.

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

Kurby, C. A., & Zacks, J. M. (2012). Starting from scratch and building brick by brick in comprehension. *Memory & Cognition, 40*(5), 812-826. doi:10.3758/s13421-011-0179-8

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.

Lakusta, L., & Landau, B. (2005). Starting at the end: the importance of goals in spatial language. *Cognition, 96*, 1-33.

Lorch, R. F., & Myers, J. L. (1990). Regression analyses of repeated measures data in cognitive research. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 16*(1), 149.

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., Radvansky, G. A., Forsythe, J. C., & Copeland, D. E. (2014). Event segmentation during first-person continuous events. *Journal of Cognitive Psychology, 26*(6), 649-661. doi:10.1080/20445911.2014.930042

Magliano, J. P., & Zacks, J. M. (2011). The Impact of Continuity Editing in Narrative Film on Event Segmentation. *Cognitive Science, 35*(8), 1489-1517. doi:10.1111/j.1551-6709.2011.01202.x

Mandler, J. M. (1987). On the psychological reality of story structure. *Discourse Processes, 10*, 1-29.

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.

Mura, K., Petersen, N., Huff, M., & Ghose, T. (2013). IBES: A Tool for Creating Instructions Based on Event Segmentation. *Frontiers in Psychology, 4*. doi:10.3389/fpsyg.2013.00994

Neville, H. J., Nicol, J. L., Barss, A., Forster, K. I., & Garrett, M. F. (1991). Syntactically based sentence processing classes: Evidence from event-related brain potentials. *Journal of Cognitive Neuroscience, 3*(2), 151-165.

Newtson, D. (1973). Attribution and the unit of perception of ongoing behavior. *Journal of Personality and Social Psychology, 28*(1), 28.

Newtson, D., & Engquist, G. (1976). The perceptual organization of ongoing behavior. *Journal of Experimental Social Psychology, 12*, 436-450.

Patel, A. D. (2003). Language, music, syntax and the brain. *Nature Neuroscience, 6*(7), 674-681. doi:10.1038/nn1082

Patel, A. D., Gibson, E., Ratner, J., Besson, M., & Holcomb, P. J. (1998). Processing syntactic relations in language and music: An event-related potential study. *Journal of Cognitive Neuroscience, 10*(6), 717-733.

Radvansky, G. A., & Zacks, J. (2014). *Event Cognition*. Oxford, UK: Oxford University Press.

Regier, T. (1996). *The human semantic potential: Spatial language and constrained connectionism*. Cambridge, MA: MIT Press.

Regier, T. (1997). Constraints on the learning of spatial terms: A computational investigation. In R. Goldstone, P. Schyns, & D. Medin (Eds.), *Psychology of learning and motivation: Mechanisms of perceptual learning* (Vol. 36, pp. 171-217). San Diego, CA: Academic Press.

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.

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

Sinclair, R. (2011). Christ, It Works for Everything. Retrieved from http://www.robertsinclair.net/comic/asshole.html

Speer, N. K., & Zacks, J. M. (2005). Temporal changes as event boundaries: Processing and memory consequences of narrative time shifts. *Journal of Memory and Language, 53*, 125-140.

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.

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

Zacks, J. M., Braver, T. S., Sheridan, M. A., Donaldson, D. I., Snyder, A. Z., Ollinger, J. M., . . . Raichle, M. E. (2001). Human brain activity time-locked to perceptual event boundaries. *Nature Neuroscience, 4*(6), 651-655.

Zacks, J. M., & Magliano, J. P. (2011). Film, narrative, and cognitive neuroscience. In D. P. Melcher & F. Bacci (Eds.), *Art and the Senses* (pp. 435-454). New York: Oxford University Press.

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.

Zacks, J. M., Speer, N. K., Swallow, K. M., Braver, T. S., & Reynolds, J. R. (2007). Event perception: A mind-brain perspective. *Psychological Bulletin, 133*(2), 273-293.

Zacks, J. M., Speer, N. K., Swallow, K. M., & Maley, C. J. (2010). The brain's cutting-room floor: Segmentation of narrative cinema. *Frontiers in Human Neuroscience, 4*, 1-15.

Zacks, J. M., & Tversky, B. (2001). Event structure in perception and conception. *Psychological Bulletin, 127*(1), 3-21.

Zacks, J. M., Tversky, B., & Iyer, G. (2001). Perceiving, remembering, and communicating structure in events. *Journal of Experimental Psychology, 130*(1), 29-58.

Zwaan, R. A., Langston, M. C., & Graesser, A. C. (1995). The construction of situation models in narrative comprehension: An event-indexing model. *Psychological Science, 6*, 292-297.

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