



Topics in Cognitive Science (2019) 1–35

© 2019 The Authors Topics in Cognitive Science published by Wiley Periodicals, Inc. on behalf of Cognitive Science Society

ISSN: 1756-8765 online

DOI: 10.1111/tops.12421

This article is part of the topic “Visual Narrative Research: An Emerging Field in Cognitive Science,” Neil Cohn and Joseph Magliano (Topic Editors). For a full listing of topic papers, see [http://onlinelibrary.wiley.com/journal/10.1111/\(ISSN\)1756-8765/earlyview](http://onlinelibrary.wiley.com/journal/10.1111/(ISSN)1756-8765/earlyview)

# Your Brain on Comics: A Cognitive Model of Visual Narrative Comprehension

Neil Cohn

*Department of Communication and Cognition, Tilburg University*

Received 3 August 2018; received in revised form 21 January 2019; accepted 18 March 2019

---

## Abstract

The past decade has seen a rapid growth of cognitive and brain research focused on visual narratives like comics and picture stories. This paper will summarize and integrate this emerging literature into the *Parallel Interfacing Narrative-Semantics Model* (PINS Model)—a theory of sequential image processing characterized by an interaction between two representational levels: semantics and narrative structure. Ongoing semantic processes build meaning into an evolving mental model of a visual discourse. Updating of spatial, referential, and event information then incurs costs when they are discontinuous with the growing context. In parallel, a narrative structure organizes semantic information into coherent sequences by assigning images to categorical roles, which are then embedded within a hierarchic constituent structure. Narrative constructional schemas allow for specific predictions of structural sequencing, independent of semantics. Together, these interacting levels of representation engage in an iterative process of retrieval of semantic and narrative information, prediction of upcoming information based on those assessments, and subsequent updating based on discontinuity. These core mechanisms are argued to be domain-general—spanning across expressive systems—as suggested by similar electrophysiological brain responses (N400, P600, anterior negativities) generated in response to manipulation of sequential images, music, and language. Such similarities between visual narratives and other domains thus pose fundamental questions for the linguistic and cognitive sciences.

---

Correspondence should be sent to Neil Cohn, Department of Communication and Cognition, Tilburg University, P.O. Box 90153, 5000 LE Tilburg, The Netherlands. E-mail: [neilcohn@visuallanguagelab.com](mailto:neilcohn@visuallanguagelab.com)

This is an open access article under the terms of the Creative Commons Attribution-NonCommercial-NoDerivs License, which permits use and distribution in any medium, provided the original work is properly cited, the use is non-commercial and no modifications or adaptations are made.

*Keywords:* Visual language; Visual narrative; Comics; Narrative; Discourse; Linguistic processing

---

## 1. Introduction

Sequential images span across human history and cultures from cave paintings, tapestries, and scrolls to visual narratives like contemporary comics and picture stories (McCloud, 1993; Petersen, 2011). The perceived transparency of understanding visual narratives has made them a popular experimental stimulus for cognitive scientists investigating many domains. Yet various healthy individuals who have little experience with graphics cannot construe meaning across drawn sequential images (Byram & Garforth, 1980; Fussell & Haaland, 1978; Liddell, 1997; Núñez & Cooperrider, 2013), implying that this ability does not rely on basic, universal perceptual processes alone (Berliner & Cohen, 2011; Magliano & Zacks, 2011; McCloud, 1993). *How then do we comprehend a sequence of images?*

This question has only recently begun to be examined in the cognitive sciences, where an emerging literature has contributed to better understanding visual narrative comprehension. This paper integrates this literature into a processing theory—the *Parallel Interfacing Narrative-Semantics Model (PINS Model)*—for the comprehension of sequential narrative images. This account will thus provide a framework for how comprehenders process the content of image sequences unit by unit using the representations posited in theories of visual narrative (Cohn, 2013b, 2015). The focus here will be on wordless visual narratives, though the basic mechanisms in principle should extend to multimodal visual narratives, albeit with additional mechanisms for the complexity that arises from such interactions (Cohn, 2016b). Thus, this work may inform the processing of film (Amini, Riche, Lee, Hurter, & Irani, 2015; Barnes, 2017; Cohn, 2016a), textual discourse (Fallon & Baker, 2016; Versluis, 2017), and their multimodal interactions (Gernsbacher, 1990; Magliano, Loschky, Clinton, & Larson, 2013).

The essence of the PINS Model is that sequential image comprehension combines processing across two representational levels of *semantics* and a *narrative structure* in a parallel architecture (Cohn, 2016b). Both of these components involve forward-looking and backward-looking mechanisms (Friederici, 2011; Hagoort, 2005, 2014; Jackendoff, 2002; Kuperberg, 2013) here characterized as *access*, *prediction*, and *updating*. These broader mechanisms pervade both representational levels of semantic and narrative structures. A sketch of this model is provided in Fig. 1. Such core mechanisms are essential features stressed by most all psychological theories of processing at both the sentence (Friederici, 2011; Hagoort, 2005; Kutas & Federmeier, 2011) and discourse levels (Kintsch, 1988; McNamara & Magliano, 2009; Zwaan & Radvansky, 1998). Indeed, they likely reflect domain-general processes that extend across expressive modalities to language, music, and other domains. We will thus begin by discussing each representational level, before addressing the implications of this work for domain-general and aspects of fluency.

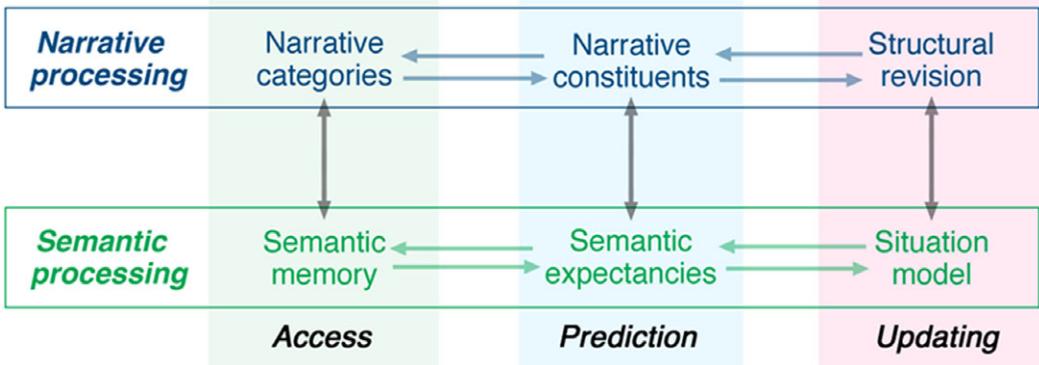


Fig. 1. Mechanisms operating over representational levels of both semantics and narrative in the processing of visual narrative sequences. Single-headed arrows represent feedforward and feed-backward connections within representational levels. Double-headed arrows represent the interfaces between semantic and narrative processing for different stages.

## 2. Semantic processing

The processing of meaning in visual narratives negotiates several levels of information, as outlined by discourse research (van Dijk & Kintsch, 1983; McNamara & Magliano, 2009). A narrative's *surface form* in verbal or written discourse would be the phonological and syntactic form. In the visual-graphic domain, the surface form would correspondingly be the graphic representation of images (including its layout) and the narrative structure that orders images into sequences (see below). This surface information links to encoded representations in semantic memory, which then become activated and incorporated into a *situation model*—a mental model constructed out of the elements and events of the progressing scene. This situation model then updates with subsequent information as the discourse unfolds, with increased costs occurring when an incoming stimulus has greater discontinuity with the preceding context (Zwaan & Radvansky, 1998). While surface features generally fade from memory, information in the situation model persists in memory into the future (van Dijk & Kintsch, 1983; Gernsbacher, 1985).

As suggested above, this overall orientation to processing has not been tied to any particular modality, and indeed has been applied to textual discourse (Radvansky & Zacks, 2014; Zwaan & Radvansky, 1998), film (Magliano & Zacks, 2011; Radvansky & Zacks, 2014), and wordless visual narratives (Gernsbacher, 1985; Magliano, Kopp, Higgs, & Rapp, 2016; Magliano, Larson, Higgs, & Loschky, 2015). Below, we further elaborate on the application of this overall dynamic to the processing of visual narratives, as depicted in Fig. 2. In particular, the PINS Model derives from observations from studies measuring event-related brain potentials (ERPs), an online measure of the electrical activity of the human brain as it unfolds in time. ERPs not only provide excellent temporal resolution of direct brain processing, but they often provide insight into functional mechanisms of processing beyond behavioral measures.

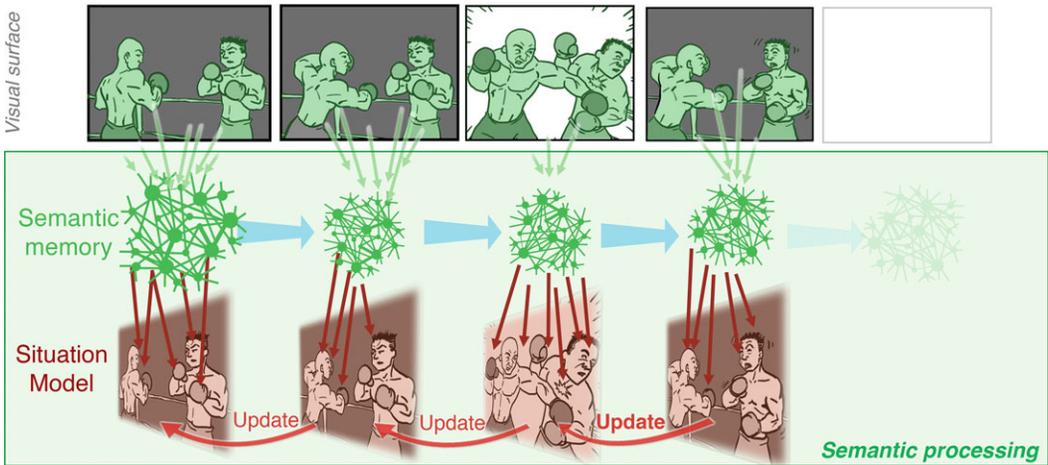


Fig. 2. Illustration of the semantic representational level for visual narratives, where a reader accesses the semantic information in images, which thereby is incorporated into a situation model of the elements and events of that scene.

### 2.1. Access: Semantic memory

The basic unit of a visual narrative is a “panel”—an encapsulated image unit usually depicting referential and event-based information, often (though not always) with a delineated border. When a reader engages a panel, she must extract its information via the surface structure of the depicted image. This process of information extraction is considered as a *front-end process*, contrasted from the *back-end processes* involved in the building of a situation model (Loschky, Hutson, Smith, Smith, & Magliano, 2018; Loschky, Magliano, Larson, & Smith, 2019; Magliano et al., 2013). This decoding involves attentional selection to guide object and scene perception to extract the relevant cues of an image (Loschky et al., 2018, 2019). In many cases, properties of the images themselves motivate such cues, as readers by and large focus on the same visual regions of interest whether images belong to a coherent or scrambled sequence (Foulsham, Wybrow, & Cohn, 2016).

Nevertheless, most images in visual narratives are created (i.e., *drawn*) intentionally to belong to a sequence, and readers in turn are tasked with finding the specific cues relevant for that context. In doing so, readers seem fairly directed in filtering the relevant content of a panel’s sequential understanding (Foulsham et al., 2016; Laubrock, Hohenstein, & Kümmerer, 2018), and comprehension of sequential images persists even at fairly rapid exposure (Hagmann & Cohn, 2016; Inui & Miyamoto, 1981). This means that the extraction of relevant cues must happen quickly and with insight from context (below). These attentional and perceptual processes may be the most modality-specific aspects of processing sequential images, as subsequent back-end processes appear to be more domain-general (Cohn, 2013a; Loschky et al., 2018; Magliano, Higgs, & Clinton, in press).

Attentional selection and extraction thus facilitate which content activates information in semantic long-term memory. This activated information may include knowledge about objects and entities (including roles like agents and patients), spatial locations, and events and actions. It may also include knowledge about specific visual narrative conventions, like that light bulbs above the head mean inspiration, or knowledge about specific visual narratives, like that Lucy typically pulls a ball away when Charlie Brown tries to kick it in *Peanuts*. In ERPs, *access* or retrieval of such semantic information is indexed by an *N400* (Fig. 3a), a negative polarity brain response peaking around 400 ms after the onset of a stimulus, like the appearance of a word or image (Kutas & Federmeier, 2011; Kutas & Hillyard, 1980). The *N400* is thought to reflect the brain’s default spread of activation through long-term semantic memory induced by an incoming stimulus—regardless of modality—relative to the prior activation established by a preceding context (Kuperberg,

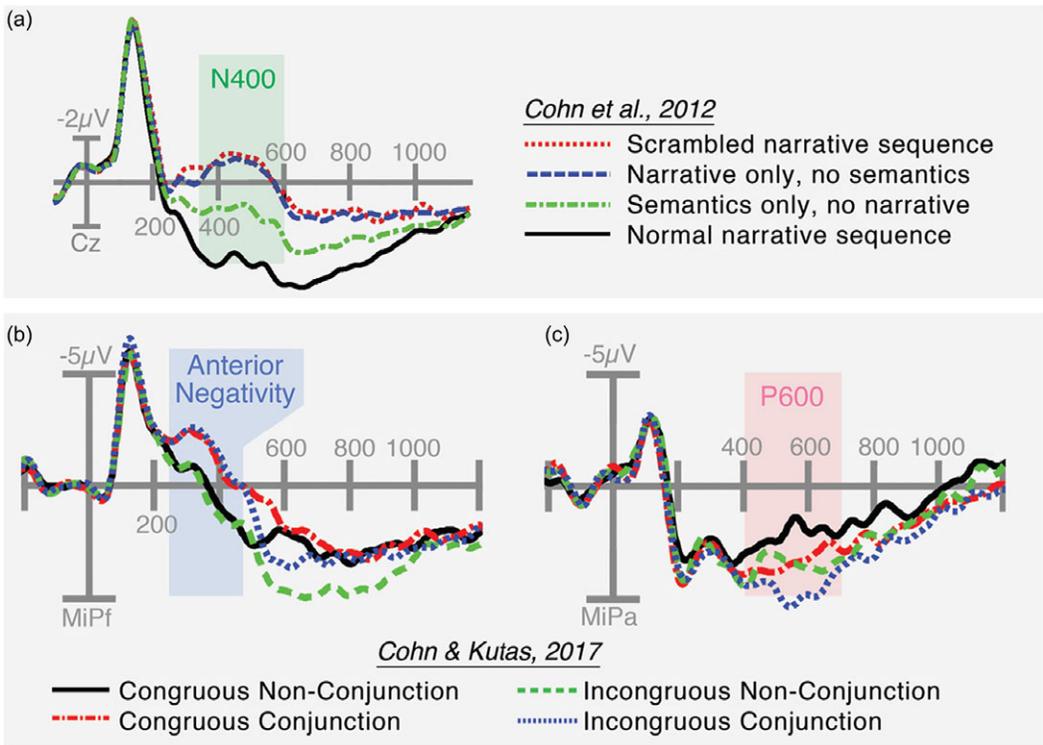


Fig. 3. Event-related potentials to manipulation of narrative and/or semantic structures in two experiments: (a) the *N400* elicited by semantic incongruity is insensitive to the presence of narrative structure (Cohn et al., 2012), while (b) an anterior negativity elicited by narrative patterning is insensitive to the presence of semantic incongruity, though the (c) *P600* is modulated by both narrative and semantics (Cohn & Kutas, 2017). Each graph depicts one electrode site, with (a) being at the midline central point of the scalp (Cz), (b) being the midline prefrontal (MiPf), and (c) being the midline parietal (MiPa). The x-axis depicts the time course of processing in milliseconds, while the y-axis depicts amplitude, with negative up. Separation of waves indicates a difference in processing, with relevant epochs highlighted.

2016; Kutas & Federmeier, 2011). Insofar as a prior context may preactivate upcoming semantic features, it may thus provide a predictive state of availability for incoming bottom-up information.

Within single images, the N400 has been observed to the recognition of visual objects (Viggiano & Kutas, 1998) or faces (Olivares, Iglesias, & Bobes, 1999). Incongruous aspects of scenes also modulate the N400, such as when unexpected objects appear within a scene, like soccer players kicking a roll of toilet paper instead of a ball (Ganis & Kutas, 2003; Sauv , Harmand, Vanni, & Brodeur, 2017; V  & Wolfe, 2013). This N400 to images may also follow an *N300*, an additional frontal negativity peaking around 300 ms which has been taken to reflect the semantic identification or categorization of *visual* objects (Draschkow, Heikel, V , Fiebach, & Sassenhagen, 2018; Hamm, Johnson, & Kirk, 2002; McPherson & Holcomb, 1999).

## 2.2. Prediction: Semantic expectancies

As context can influence subsequent semantic activation, the sequencing of visual narratives further modulates the N400. As in sentence and discourse processing, the N400 to images in a sequence is modulated by the degree to which the semantic features of an incoming stimulus overlap with those activated by its prior context (Kuperberg, 2016; Kutas & Federmeier, 2011). This means that larger N400s are observed to incongruous (Amoruso et al., 2013; Cohn, Paczynski, Jackendoff, Holcomb, & Kuperberg, 2012; West & Holcomb, 2002) or unexpected information in a visual sequence (Amoruso et al., 2013; Cohn & Kutas, 2015; Reid & Striano, 2008). For example, larger N400s are evoked by fully semantically incongruous panels compared to incongruities that maintain semantic associations with prior panels (Cohn, 2012; Cohn et al., 2012), a finding that also manifests in longer self-paced viewing times and lower comprehensibility ratings (Cohn, 2012). This effect holds even when crossing modalities: Larger N400s are observed to incongruous words compared to congruous words which replace the climax of a visual narrative sequence (Manfredi, Cohn & Kutas, 2017).

This modulation based on sequence occurs because comprehenders make various probabilistically based *expectations* about the way an incoming image will relate to a prior and subsequent sequential context. In a spatially juxtaposed narrative, comprehenders likely maintain a high probability expectancy that the objects/characters and events in one image will be retained in subsequent images (i.e., a *continuity constraint*). Also, the information extraction operating on such continuity is not perceptually trivial: Objects and events may change in their visual representations across images not only in different postures or states but also shown with different viewpoints, sizing, framing, or visual style. Indeed, recognition of such referential continuity appears impaired in populations with less exposure to visual narratives (Byram & Garforth, 1980; Fussell & Haaland, 1978; Liddell, 1997; N n ez & Cooperrider, 2013).

Within semantic memory, representations already activated should in turn be reactivated with less cost by an incoming image with high semantic feature overlap, which is why N400 amplitudes become attenuated for images that are congruous with their context

compared to those that are incongruous. As a reader progresses panel by panel in a coherent narrative sequence, semantic access of each image becomes facilitated as structural and semantic expectancies are cyclically confirmed relative to what came before (Cohn et al., 2012; Giglio, Minati, & Boggio, 2013). This buildup is evident in shorter viewing times and attenuated N400s across each subsequent image in a coherent sequence (Cohn & Paczynski, 2013; Cohn & Wittenberg, 2015; Cohn et al., 2012; Foulsham et al., 2016; Giglio et al., 2013). Thus, exposure to a prior congruous context makes subsequent information easier to process.

These back-end semantic processes may in turn affect front-end processing. Although readers fixate the same basic regions of images in both coherent and scrambled image sequences, eye movements appear less dispersed to the content of images in meaningful narrative sequences compared to the more widespread fixations in scrambled sequences (Foulsham et al., 2016). That is, a coherent sequential context constrains expectations about where to look within an image to find the relevant content. Such findings are consistent with work showing event knowledge guiding eye movements in other domains, such as discourse (Swets & Kurby, 2016) and event perception (Eisenberg, Zacks, & Flores, 2018). Thus, confirmation of expectancies about a visual narrative sequence feeds back to make information extraction and the access of semantic memory easier across a sequence. Such facilitation of semantic information across ordinal position in a sequence occurs across other sequentially meaningful domains, as it is also observed in discourse processing (Haberlandt, 1980, 1984) and sentence processing (Van Petten & Kutas, 1991).

As context aids facilitation, semantic access should thus be more demanding at the start of a sequence where initial information has yet to be established. The discourse literature posits a mechanism of *laying a foundation* (Gernsbacher, 1985, 1990) to describe how the basic semantics are established at the outset of a sequence. Evidence for such a process was initially suggested by slower self-paced reading times at the outset of a textual discourse (e.g., Glanzer, Fischer, & Dorfman, 1984), and indeed, longer self-paced viewing times have also been observed at the starting panel of visual narratives (Cohn, 2014; Cohn & Paczynski, 2013; Cohn & Wittenberg, 2015; Foulsham et al., 2016). Some work has speculated that these longer viewing times arise because laying a foundation demands increased fixations when starting a visual sequence (Loschky et al., 2018). With a view of semantic access, laying a foundation arises from the cost of retrieving semantic information with reduced or absent prior context. This may in part manifest in perceptual processes of attention and extraction (Foulsham et al., 2016), but it may be motivated by back-end cognitive processes. This is suggested because larger N400 amplitudes appear to images at the start of a sequence compared with attenuated amplitudes in subsequent panel positions (Cohn et al., 2012; Giglio et al., 2013), and in such ERP experiments, participants must make minimal eye movements (since it creates muscle artifact). In addition, larger amplitude N400s are similarly observed to the first *words* of a sentence, which also become attenuated across ordinal position (Van Petten & Kutas, 1991). This domain-general N400 attenuation suggests that costs of processing at the outset of a sequence are not a facet of narrative/discourse per se, but of accessing semantic information in a sequence more generally.

In addition, limiting exposure durations to first panels does not impede comprehension (Hagmann & Cohn, 2016) nor does omitting less semantically informative first panels that function to absorb such a scene-setting process (Cohn, 2014; Magliano et al., 2016). This suggests that dedicating time or units to laying a foundation is helpful to processing, but not essential, which would be consistent with the view that such costs are a consequence of the default access of semantic information in a sequential context. Also, unexpected panels at the start of a sequence demand longer viewing times beyond even those to coherent starting panels, such as in scrambled sequences (Cohn, 2014; Cohn & Wittenberg, 2015; Foulsham et al., 2016). Thus, some images are harder than others to process even at the start of a sequence, suggesting that longer viewing times reflect an increased cost of accessing the semantics due to minimal prior context.

This process of facilitation based on upholding continuity may or may not involve the preactivation of specific representations in memory (DeLong, Troyer, & Kutas, 2014; Kuperberg, 2016; Kutas, DeLong, & Smith, 2011). Rather, contiguous information upholds a degree of probabilistic *predictability*, which may involve reactivation, but does not necessarily involve overt predictions. Some content may generate anticipations for “what will happen next” though, in similar fashion to “predictive inferences” (McKoon & Ratcliff, 1986). Experimental work has suggested that various cues can motivate expectancies for foreshadowing in the comprehension of films (Magliano, Dijkstra, & Zwaan, 1996), but most work on prediction implicates more locally constrained processes. For example, characters in preparatory postures about to carry out a subsequent action (e.g., an agent reaching back an arm to punch) have been shown to elicit more agreement about their subsequent actions than the panels of characters who might receive those actions (i.e., patients) (Cohn & Paczynski, 2013). These expectancies appear to facilitate subsequent information, as panels following such preparatory postures are viewed faster than those following panels of patients (Cohn & Paczynski, 2013), and removing such preparatory cues lead to neural costs (Cohn, Paczynski, & Kutas, 2017). Such findings are consistent with other work showing anticipatory processing in event cognition (Eisenberg et al., 2018; Zacks, Kurby, Eisenberg, & Haroutunian, 2011). Such predictability is thus fairly local and constrained by specific event-based cues.

Although theories of the N400 have posited a role of feedforward stimulation (Kutas & Federmeier, 2011), extant work on visual narratives thus far has not shown evidence of semantic preactivation, as suggested by studies of the N400 in language (van Berkum, Brown, Zwitserlood, Kooijman, & Hagoort, 2005; DeLong, Urbach, & Kutas, 2005; Szewczyk & Schriefers, 2018). Indeed, in language, the N400 is inversely correlated with the expectancy of an upcoming word (Kutas & Federmeier, 2011), as measured through cloze probability (i.e., quantification of what happens next in a sequence, given a prior context). To the extent that the N400 indexes the same resources in visual narratives as in language, the PINS Model expects similar probabilistically modulated predictive processing in visual narratives. Investigation of such semantic expectancies could follow visual analogs of cloze probability (i.e., “what happens next?”), which would be promising for future research both for visual narratives and to test theories of the N400 in a nonverbal domain.

### 2.3. Updating: Situation model revision

As described above, the semantic information accessed in an image sequence must be integrated into the unfolding sequential representation. This knowledge thus becomes incorporated into a *situation model* of the aggregated meaning of a construed discourse (McNamara & Magliano, 2009; Zwaan & Radvansky, 1998). While being constructed during the reading of a visual narrative, a situation model remains held in working memory, but shifts to being stored in episodic long-term memory as its understanding is retained into the future (Magliano, Kopp, McNeerney, Radvansky, & Zacks, 2012). During online processing, the content of each image triggers an *update* of the situation model, involving integration, reanalysis, and/or reorganization of prior information established by the preceding context. As a result, greater updating occurs with greater discontinuity of the incoming information given the preceding context (Huff, Meitz, & Papenmeier, 2014; Magliano & Zacks, 2011). For example, updating of situational changes may occur across dimensions of characters, spatial locations, or event information, as posited by theories of visual and verbal narrative (Bateman & Wildfeuer, 2014; Hoeks & Brouwer, 2014; Huff & Schwan, 2012; Magliano, Miller, & Zwaan, 2001; Saraceni, 2016; Stainbrook, 2016; Zwaan & Radvansky, 1998). Because a situation model is always being built in reference to a progressing (visual) discourse, such updating processes occur iteratively at each unit of a (visual) discourse, not just to incongruities. These mappings may be incremental in nature (Huff et al., 2014; Kurby & Zacks, 2012), but when they become untenable, there may be a shift to a new situation model (Gernsbacher, 1990; Loschky et al., 2019; McNamara & Magliano, 2009; Zwaan & Radvansky, 1998).

In ERPs, updating processes are associated with the *P600* (Fig. 3c), a positivity typically peaking 600 ms after the onset of a word or image with a posterior distribution across the scalp. Although P600s were first associated with syntactic processing (Hagoort, Brown, & Groothusen, 1993; Osterhout & Holcomb, 1992) (elaborated below), evidence of their elicitation by both structural and semantic violations has led to a broad interpretation of P600s as indexing an updating, integration, and/or reanalysis processes triggered by an incoming discontinuity with a prior context (Brouwer, Crocker, Venhuizen, & Hoeks, 2016; Brouwer & Hoeks, 2013; Kuperberg, 2013, 2016; Van Petten & Luka, 2012). Such an interpretation is also consistent with arguments linking the P600 to other positivities associated with mental model updating (Donchin & Coles, 1988; King & Kutas, 1995).

In visual sequences, P600s have been observed in both congruous and incongruous circumstances, supporting the idea that updating persists continuously across each image in a sequence, with increased situational change demanding greater updating (Cohn & Kutas, 2015; Magliano & Zacks, 2011). Both congruous and incongruous character changes between panels elicit P600s (Cohn & Kutas, 2015, 2017). They have also been observed to alterations of the semantic cues signifying events in visual narratives, such as the explicitness of event structures (Cohn & Kutas, 2015) or omission or reversal of motion lines that depict the trajectory of path actions (Cohn & Maher, 2015). Such findings align with P600s observed to the updating of referential and inferential information in discourse (Ferretti,

Rohde, Kehler, & Crutchley, 2009; Nieuwland & Van Berkum, 2005) and with P600s observed to manipulations of real-world visual events, outside the context of narratives (Amoruso et al., 2013; Sitnikova, Holcomb, & Kuperberg, 2008). Indeed, other measures of event perception have similarly implicated processes of mental model updating (Papeneier, Boss, & Mahlke, 2018; Zacks, Speer, Swallow, Braver, & Reynolds, 2007). Thus, the P600 appears to index a backward-looking process of updating a mental model given the degree to which an incoming signal aligns with a prior context.

Additional revision of a situation model may occur when information is missing in the surface structure of visual cues, thus demanding an inference. For example, if a boxer is shown reaching back to punch, and then, his opponent is shown on the ground (i.e., if Fig. 4 omitted the third panel), an inference will be required to understand its cause as a knockout punch. Inferences have long been a focus of discourse research on situation models (McNamara & Magliano, 2009) and are a primary aspect of theories of visual narratives (McCloud, 1993; Saraceni, 2016). In the processing of visual narratives, P600s have been observed in inferential contexts involving backward-looking situational discontinuity, such as constructing a spatial inference out of disparate characters (Cohn & Kutas, 2017), or to the

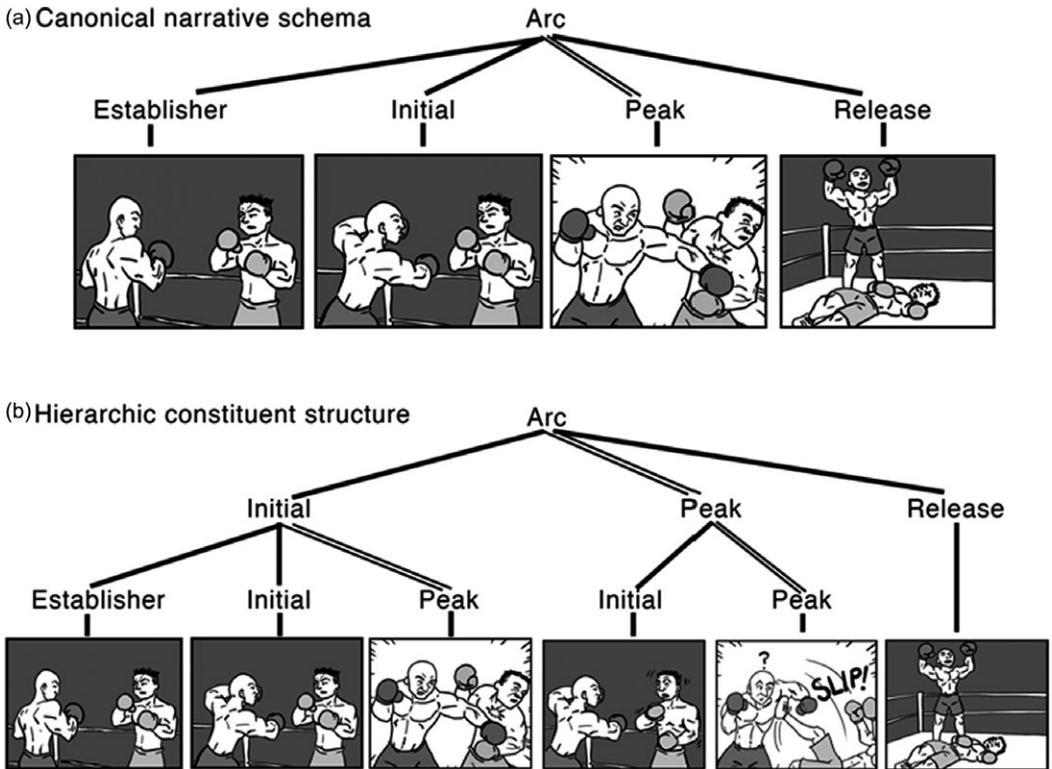


Fig. 4. Visual sequences showing (a) canonical narrative schema in Visual Narrative Grammar, and (b) schema combined in hierarchic constituent structures.

differential activity between a panel with inexplicit event information and a subsequent image which resolves that event-based inference (Cohn & Kutas, 2015).

Inferential demand may also trigger sustained late negativities, thought to index working memory processes, such as searching through a mental model to resolve inferential or referential ambiguities (van Berkum, 2009; Hoeks & Brouwer, 2014). Such negativities have been observed to event-based inferences in visual narratives, such as to panels following an event that is omitted from a scene (Cohn & Kutas, 2015). Behavioral research has similarly implicated inferential processing by longer viewing times to panels following the position of omitted event information (Cohn & Wittenberg, 2015; Hutson, Magliano, & Loschky, 2018; Magliano et al., 2015, 2016), and these longer viewing times appear to be modulated by working memory demands (Magliano et al., 2015). Such an effect is again consistent with findings from language of sustained frontal negativities working to build inferred event information (Baggio, van Lambalgen, & Hagoort, 2008; Bott, 2010; Paczynski, Jackendoff, & Kuperberg, 2014; Wittenberg, Paczynski, Wiese, Jackendoff, & Kuperberg, 2014). Thus, inference generation incurs costs for updating a situation model for information that is not explicitly depicted in a narrative.

### 3. Narrative processing

The story so far is that the processing of visual narratives involves assessing the basic semantics of images, which lead to predictions related to its contiguity with a sequential context, and this information is then incorporated into a growing situation model. Integration of this information depends on the congruity of an image and its context, and this knowledge can potentially feed back on basic semantic processing, as the process repeats. Thus, the access-updating cycle iteratively occurs at each unit of a sequence (Brouwer & Hoeks, 2013; Brouwer et al., 2016). Overall, this process should be consistent with linguistic models of discourse (Kintsch, 1988; McNamara & Magliano, 2009; Zwaan & Radvansky, 1998), and with basic mechanisms described at both sentence and discourse levels of processing (van Berkum, 2012; Friederici, 2011; Hagoort, 2005; Kutas & Federmeier, 2011; Kutas, Kluender, Barkley, & Amsel, 2017).

Nevertheless, these mechanisms alone are not sufficient to account for the processing of visual narrative sequences. First, in that semantic processes for (visual) narratives involve domain-general process of mental model construction—not tied to any particular expressive modality—they are mostly consistent with those described as operating on discourse (Gernsbacher, 1990; Kintsch, 1988; McNamara & Magliano, 2009) and event cognition (Loschky et al., 2018, 2019; Radvansky & Zacks, 2014). This means that nothing *per se* about these mechanisms limit them to the understanding of *narratives*. Nevertheless, we recognize that everyday events differ from those packaged in narrative contexts. Thus, some nontrivial cognitive structure must allow us to distinguish narratives from everyday experiences.

Second, on this point, parts of a visual narrative sequence play discernable roles from each other within a narrative context—panels that set up actions function differently in a

sequence than those that depict the climax (Cohn, 2014). Indeed, narrative roles have been characterized as far back as Aristotle (Butcher, 1902), and narrative theories generally show convergence on identifying such roles (Brewer, 1985; Cohn, 2013b; Cutting, 2016). Visual narratives also use a wide range of identifiable sequencing patterns (Batesman, 2007; Branigan, 1992; Cohn, 2015, in press), which appear to differ cross-culturally in their frequency (Cohn, in press). Neither relative narrative roles nor sequencing patterns occur in everyday events, again warranting a system that encodes such entrenched knowledge.

Third, semantic processing alone cannot account for various relations between panels beyond image-to-image juxtapositions (Magliano & Zacks, 2011; McCloud, 1993; Saraceni, 2016; Stainbrook, 2016). As in sentences, units in a visual narrative can involve long-distance connections between nonjuxtaposed images, including center-embedded “clauses” (Cohn, 2013b). Some sequences may be ambiguous, where a single structural sequence has multiple interpretations and/or parsings (Cohn, 2010b, 2015), or may depict complex semantic relations like metaphor, which may not be motivated by an event structure (Cohn, 2010a; Tasić & Stamenković, 2015). In addition, the same general meaning can be expressed in multiple different ways that vary what is shown when and how (Brewer & Lichtenstein, 1981; Cohn, 2013b, 2015; McCloud, 1993)—warranting a system separate from meaning to allow such differences in presentation. Such phenomena require more than just monitoring perceptuo-semantic changes.

Finally, semantic processes rely on general functions which do not account for patterned differences between narrative systems. Yet visual narratives do systematically differ across cultures and time periods (Cohn, 2013a, in press; Cohn, Pederson, & Taylor, 2017), and readers process these patterns differently based on their frequency of exposure (Cohn & Kutas, 2017). In order for constructs to deviate cross-culturally, and for readers of those constructs to process them differently, these patterns must be encoded in long-term memory beyond just online tracking of discontinuity, which posits no stored representations.

Altogether, these observations necessitate a visual narrative *structure* that goes beyond semantic processing alone. Where the PINS Model describes the processes at work in comprehending sequential images, the theory of *Visual Narrative Grammar* (VNG) describes the representations that undergo those processes. VNG argues that a combinatorial structure runs parallel to semantic processing, which functions to organize this meaningful information into comprehensible sequences. The situation model is the constructed understanding of a (visual) narrative’s meaning, but a *narrative grammar* guides how that meaning is conveyed sequentially. Thus, while the information in the situation model should persist in memory (van Dijk & Kintsch, 1983; Gernsbacher, 1985), the narrative grammar should not, nor may it be as consciously apparent as the semantics of a sequence (Cohn & Bender, 2017; Cohn et al., 2012).

This narrative grammar operates on sequential images using similar architectural principles as a syntactic structure in sentences. Like syntax, narrative grammar assigns image units to categorical roles, and then organizes them using a constituent structure guided by constructional schemas. Narrative grammar operates at a higher level of semantics than in

sentences though, closer to the level of a discourse structure, since most images contain more information than individual words (Cohn, 2013b, 2015). Nevertheless, the basic principles of structure maintain in both syntax and narrative: Categorical roles are organized in schematic structures that allow for distance dependencies, structural ambiguities, and other complex patterns (Cohn, 2013b, 2015).

In that VNG makes an analogy with grammatical structure, it has similarities to previous grammars posited for stories (Mandler & Johnson, 1977; Rumelhart, 1975) or film (Carroll, 1980; Colin, 1995). However, VNG attempts to account for several of the critiques leveled at these prior comparisons between narrative and syntactic structure (de Beaugrande, 1982; Black & Wilensky, 1979; Garnham, 1983). Previous story grammars used phrase structure rules, based on early models of generative grammar (Chomsky, 1965), but were critiqued for characterizing semantics, not grammar (de Beaugrande, 1982; Black & Wilensky, 1979). Such limitations may have been related to many story grammars being operationalized through memory tasks, where semantic information is retained but structure is not (de Beaugrande, 1982).

VNG addresses these critiques in several ways. First, VNG is based on sequencing schema stored in memory, modeled after *construction grammars* (Culicover & Jackendoff, 2005; Goldberg, 1995). Here, sequencing patterns are encoded in long-term memory along with interface rules mapping to an unambiguously parallel structure of semantics (Cohn, 2013b, 2015). Unlike with generative procedural rules, stored schema use “unification” as a combinatorial mechanism (Hagoort, 2005, 2016; Jackendoff, 2002), which is the process of constructing larger structures by assembling pieces of structure stored in memory, given context. Unlike story grammars, VNG also posits additional schema that elaborate or modify the canonical order (Cohn, 2013b, 2015), as discussed below, and potentially other idiosyncratic patterns analogous to syntactic constructions and idioms (Cohn, 2013a). These constructs have not been based on memory tasks, but rather online measures of (neuro)cognition, which in some cases align with mechanisms found in syntactic processing, as discussed below.

### 3.1. Access: Narrative categories

#### 3.1.1. Theory

Within VNG, panels are assigned categorical roles for how they function within a sequence. There are four core narrative categories. *Establishers* (E) set up the referential entities in an interaction, often as a passive state. For example, in Fig. 4a, the Establisher simply depicts two boxers, without any actions. *Initials* (I) then mark the beginning of narrative tension, prototypically an “about to” event like the preparatory action and/or a source of a path, like the reaching back to punch of the boxer in Fig. 4a. A *Peak* (P) depicts the climax of the sequence, such as a completed or interrupted action and/or goal of a path, like the boxer’s punch in Fig. 4a. Finally, a *Release* (R) dissipates the narrative tension, prototypically mapped to a semantic coda or aftermath of an action. In Fig. 4a, this comes with one boxer standing victorious over the other. Other categories expand on these core states (see Cohn, 2013b).

Narrative categories are in part assigned by panels' internal semantic cues (Fig. 1), and certain cues license prototypical mappings to particular narrative roles. For example, a preparatory action like reaching back to punch (as in panel 2 of Fig. 4) would map prototypically to an Initial. However, semantic cues alone do not determine narrative categories, which can also be influenced by a panel's distribution in a sequence. This relationship is similar to syntactic categories in sentences (like nouns, verbs), which have prototypical correspondences with semantics (like objects, events), but ultimately are defined by their context in a sentence: for example, the word *dance*, which is semantically an event, can be either a noun (*the dance*) or verb (*they dance*) depending on context. Similarly, narrative categories balance this relationship between semantic content and sequence context (Cohn, 2013b, 2014), as elaborated below.

Narrative categories may also facilitate aspects of semantic processing. Establishers at the outset of a sequence may aid a comprehender with the greater demand of semantic access at the outset of a sequence (*laying a foundation*). This narrative category thus "absorbs the cost" of increased semantic processing with a unit that prototypically contains minimal event information, and indeed is fairly expendable (Cohn, 2014; Haggmann & Cohn, 2016). Releases may have a similar function for "wrap up" processing at the end of a narrative sequence (Cohn, 2014; Cohn & Wittenberg, 2015; Foulsham et al., 2016). Releases are also fairly expendable, but sequences missing them are deemed less comprehensible (Cohn, 2014).

### 3.1.2. Evidence for narrative categories

If they did not play narrative roles, we would expect panels to have uniform tendencies across a sequence, modulated only by the degree of (dis)continuity between them. In such a case, meaningful relations alone should distinguish panels, such as those that start a sequence incurring greater cost, where laying a foundation motivates greater access. In contrast, if panels function as categories, consistent and different behaviors should distinguish them from each other. A variety of tasks have implicated such varied tendencies.

First, some panels are recognized as more or less essential to a sequence. When participants are asked to omit panels from a sequence, they consistently choose to delete "peripheral" categories (Establishers, Releases) more often than "core" categories (Peaks, Initials) (Cohn, 2014). A complementary task asked participants to recognize where a panel had been deleted from a sequence, and here, the "core" categories were more accurately recognized as missing (Cohn, 2014; Magliano et al., 2016). Thus, panels vary in their importance in a sequence.

Panels also vary in the flexibility of their positioning in a sequence. In tasks asking participants to arrange unordered panels, some panel content can play multiple roles in a sequence, while others are less able to be rearranged. For example, panels acting as Establishers and Releases can be displaced more than other categories, and these panels do not vary in viewing times when their positions become reversed (Cohn, 2014). In contrast, Initials and Peaks are less flexible in their positioning in a sequence and incur costs when displaced in a sequence. Peaks moved to the start of a sequence evoke viewing times longer than other categories. This implies that laying a foundation at the start of a

sequence is not a uniform process, but modulated by expectations of what kinds of information start sequences. Finally, brain responses differ to panels that violate narrative category expectations compared to those that are congruous (Cohn, 2012; Cohn & Kutas, 2015). Overall, these findings support that panels do not behave in uniform ways in a sequence, and that narrative categories are differentiated by distributional behaviors.

### 3.2. Prediction: Narrative constituents

#### 3.2.1. Theory

Narrative categories do not just characterize isolated panel types, but rather are embedded in a *canonical narrative schema*, as in Table 1a. This canonical narrative schema places narrative categories in a preferred order, encoded as a pattern in memory, in line with construction grammar (Culicover & Jackendoff, 2005; Goldberg, 1995).

Several studies provide evidence that readers distinguish between canonical and noncanonical narrative sequences. Canonical sequences are easier to reconstruct and are rated as more comprehensible than noncanonical sequences, and panels in canonical order are viewed at shorter viewing times than those out of order (Cohn, 2014). In addition, the rates at which participants recognize and omit categories from sequences follow the general “shape” of the narrative schema: Peaks are most necessary, followed by Initials, and then Releases and Establishers (Cohn, 2014). Similarly, participants’ conscious segmentation of a visual sequence follows the canonical schema’s preferences, with placement of a segmental break being most likely prior to an Establisher, which typically starts a segment, then descending in likelihood along the canonical schema ( $E > I > P > R$ ), and the reverse preferences prior to a segmentation (Cohn & Bender, 2017).

Since categories inherently belong to a narrative schema, such roles may thus be influenced by their sequence context, beyond just semantic content. This means that narrative categories can depict semantic content that may not conform to the prototypical mappings. For example, a panel may still be categorized as an Initial if it depicts an ongoing process (like running) and not a preparatory event if it is placed between an Establisher and a Peak, as supported by the order of the schema. Or a panel depicting a passive event may play a role either as an Establisher or Release, depending on placement at the start or end of a sequence (Cohn, 2014).

Nevertheless, prototypical interfaces between semantic content and a narrative schema mean that a well-formed narrative can be satisfied without maintaining semantic coherence between panels. As in Fig. 5a, a couple sitting on a couch passively could be an Establisher. A preparatory action (reaching back to punch) could map to an Initial, while a subsequent totally unrelated climactic action (a building blowing up) would fulfill a

Table 1  
Basic constructional patterns in Visual Narrative Grammar

(a) Canonical narrative schema	[Phase X (Establisher) – (Initial) – Peak – (Release)]
(b) Conjunction schema	[Phase X X <sub>1</sub> - X <sub>2</sub> - . . . X <sub>n</sub> ]
(c) Head-modifier schema	[Phase X (Modifier) – X – (Modifier)]

Peak, followed by another unrelated Release of the response to an action (a dog hiding). As in Fig. 5b, the semantic cues would provide bottom-up mappings to narrative categories, and the top-down *Establisher-Initial-Peak-Release* sequence would be narratively well-formed, thus satisfying structural predictions and revision (discussed below). However, the accessed semantic information between panels would remain semantically incoherent. Such sequences are analogous to syntax-only sentences like *Colorless green ideas sleep furiously* (Chomsky, 1965), which has a well-formed syntax without semantic connections between words.

In research on these “narrative only” sequences, participants’ ratings have confirmed that both whole sequences and panel-to-panel bigrams are meaningfully incoherent. This incoherence makes semantic access just as difficult across panels in narrative only sequences as in a scrambled sequence of images, indexed by the N400 (Fig. 3a). However, the well-formedness of the narrative grammar confers an advantage beyond scrambled sequences: Response times to target panels in narrative only sequences are faster

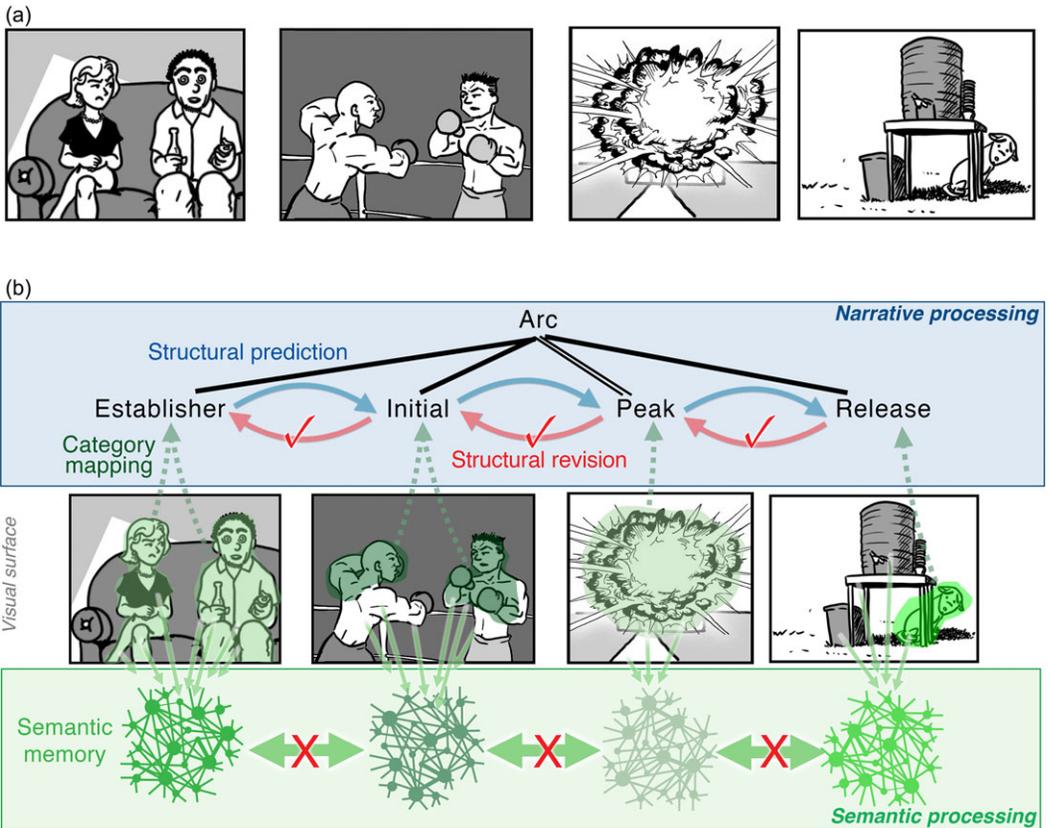


Fig. 5. (a) An example sequence with a well-formed narrative structure but no semantic relations between panels, and (b) a diagram of its processing across narrative and semantic representational levels. Semantic cues provide adequate mappings to categories, which are correctly ordered into the narrative schema. Yet the activated semantic information maintains no relations between panels.

than to those in scrambled sequences, despite no difference in the N400 amplitudes that they evoke (Cohn et al., 2012). Since the N400 is not sensitive to the narrative, despite that structure aiding in response times, it indicates that the narrative grammar operates on a separate representational level than semantics.

Within VNG, the canonical narrative schema can apply categorical roles both to panels and to other phases made up by groupings of panels. Consider Fig. 4b, which expands the narrative from Fig. 4a, by adding only two panels after the first Peak. Now, the surface structure is *E-I-P-I-P-R*, which on its own is not a “legal” sequence of categories. However, panels now form groupings, where the basic narrative schema appears at both the surface level of panels (within constituents) and at a higher level of structure (across constituents). In this case, the first constituent plays the role of an Initial, which sets up a subsequent Peak of the defeated boxer slipping—itsself built of an Initial (first boxer reaching back again) and a Peak (second boxer collapsing). The Peak of each constituent acts as its “head,” meaning that it motivates the internal structure of that constituent, and in turn motivates the constituent level categories (Cohn, 2013b, 2015). That means that the second Peak—which heads a Peak constituent—is the main event of the sequence. *Arcs* are a node that plays no other role in the sequence, often the topmost node. As narrative schemas are recursive, these structures can thus climb to high levels, including whole plotlines, which then consist of multiple subnarrative constituents.

VNG also differs from other formal models of narrative in that it posits additional schemas that elaborate or modify the canonical narrative schema. For example, a *conjunction schema* repeats a narrative category within a constituent of that same type (Table 1b), similar to the way that conjunction in syntax repeats words within a phrase of that type (e.g., conjoined nouns form a noun phrase: [<sub>NP</sub> [<sub>N</sub> salt] and [<sub>N</sub> pepper]]). Conjunction may manifest in different mappings to semantics. For example, *Environmental-Conjunction* uses panels showing various characters at the same narrative state, where the broader spatial location must then be inferred (Fig. 6a). Other types of conjunction depict parts of a character to imply the whole, parts or iterations of actions, or various elements connected by a broader semantic field (Cohn, 2015). Additional modification may use a panel that zooms in on the contents of another panel (a Refiner), which establishes a head-modifier relationship (Table 1c), as depicted in Fig. 6b.

Thus, VNG posits that three basic sequencing patterns operate in sophisticated visual narrative systems (canonical phase, conjunction, head-modifier). Together, these schema can combine to create substantial complexity to visual narrative sequencing (Cohn, 2015), and sometimes these combinations constitute stored patterns of their own (Cohn, In press). For example, successive conjunctions can interweave multiple narrative tracks (Cohn, 2015), as in “parallel-cutting” described in film (Bateman & Schmidt, 2012; Buckland, 2000; Carroll, 1980). In addition, because VNG is a construction grammar, it allows for other idiosyncratic constructions which may not be captured by these abstract schemas (Cohn, 2013a; Cohn & Kutas, 2015).

Several hypotheses arise from VNG’s encoding of schematic structures in memory. First, segmentation arises from the combination of these schemas, with boundaries at the breaks between narrative constituents. Second, schematic ordering should allow for

forward-looking predictions. Third, sequencing of categories that do not uphold schematic constraints should trigger a structural revision.

3.2.2. Structural predictions

As narrative categories are encoded in memory within a canonical narrative schema, incoming category information can instigate *structural predictions* of subsequent category information. This type of prediction is not about what type of semantic event might occur (as in predictive inference), but rather what narrative structure may probabilistically come next given the sequencing schemas. Such probabilistic predictions may thus be sensitive to the order of categories from the canonical schema and/or the patterns of various narrative constructions (Cohn, 2015, In press), and they may be modulated the familiarity with those patterns given the visual narratives a comprehender reads (Cohn & Kutas, 2017). Schematic ordering would thus predict that the boundaries between segments occur when

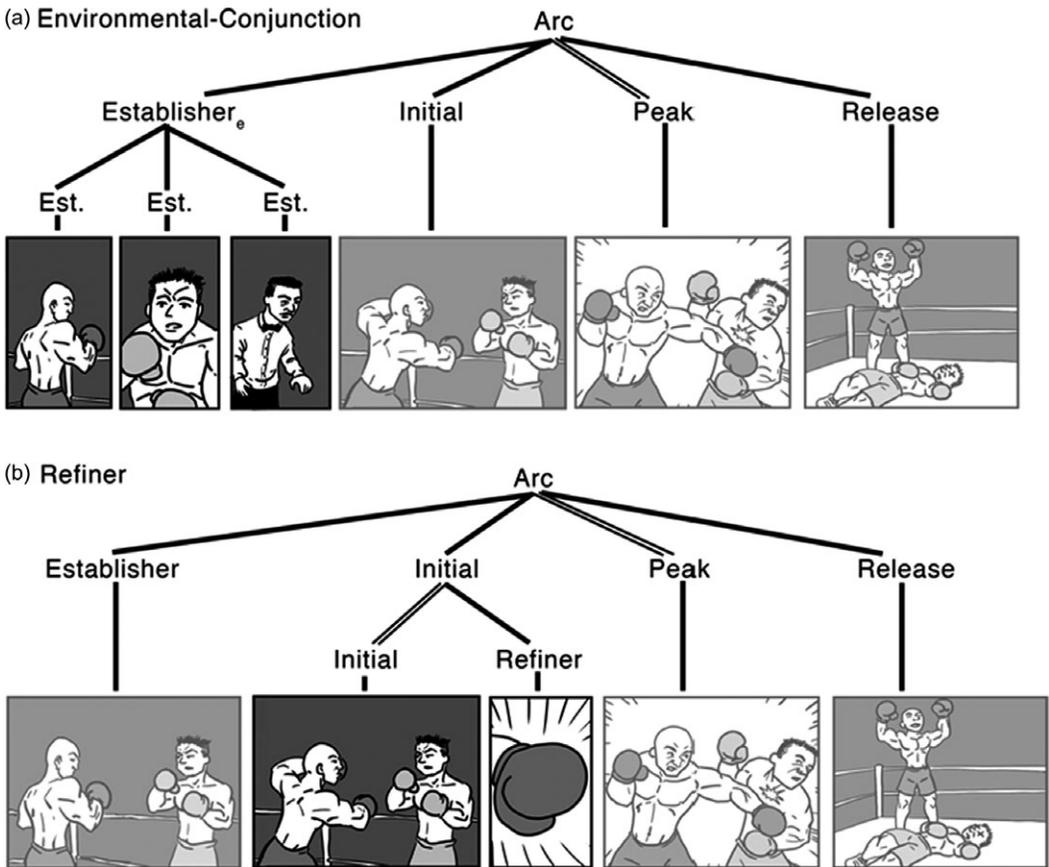


Fig. 6. Modifying schema in VNG applied to Fig. 4a using (a) a conjunction schema which repeats narrative categories within a node (here Establishers) and (b) a head-modifier schema which elaborates on a “head”—here an Initial modified by a Refiner.

panel relations do not conform to the canonical sequence. For example, a *Peak-Initial* order of a panel bigram confounds the sequencing in the canonical narrative schema (Table 1a), and thus, this bigram should cue a break in constituent structure, as in Fig. 7. Note that this contrasts with discourse models which argue that “segmentation boundaries” are triggered on the basis of semantic discontinuity, like changes in location or characters (Gernsbacher, 1990; Magliano & Zacks, 2011; Radvansky & Zacks, 2014). In the PINS Model, situational discontinuity may prototypically align with narrative breaks, hence the mapping between narrative constituents and semantic expectancies in Fig. 1 (Cohn & Bender, 2017; Cohn, Jackendoff, Holcomb, & Kuperberg, 2014; Hagmann & Cohn, 2016), but changes in meaning do not necessarily trigger the breaks themselves.

Work examining the breaks between segments in visual narratives has long used “segmentation tasks” which ask participants to mark where one segment ends and another begins. Participants have shown consistent intuitions for how to segment visual narratives into constituents (Cohn & Bender, 2017; Gernsbacher, 1985; Magliano et al., 2012), and because such segmentations often align with situational change, a causal relationship was assumed (Gernsbacher, 1990; Magliano & Zacks, 2011). However, when included into analyses, narrative category information is more predictive of segmentation than situational change (Cohn & Bender, 2017). Indeed, these preferences follow the order of a canonical narrative schema: Categories starting the schema (Establisher, Initial) are more predictive of beginning a new segment than categories ending the schema (Peak, Release). Such work implies that juxtaposition of noncanonical categories signal constituent boundaries, whether or not they align with situational changes.

If noncanonical orders of categories provide cues for constituent boundaries, the narrative schema should also provide predictions for the order of categories within a constituent. Thus, if a comprehender identifies a panel as an Initial, it carries a structural prediction that the subsequent image will be a Peak. Similarly, if they view a Peak or Release, a subsequent image

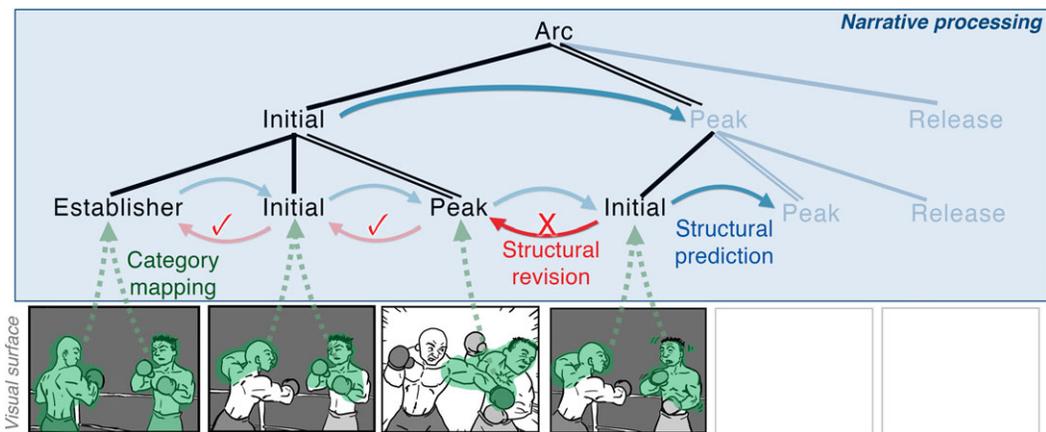


Fig. 7. Illustration of the narrative representational level of sequential images, where semantic cues in images map to categorical roles, which in turn sponsor predictions based on a schematic order (in blue), and are thereby revised in the face of incoming information.

should start a new constituent. Such forward-looking predictions have been implicated by results of an experiment where blank “disruption panels” were inserted within or between the narrative constituents of a visual sequence. Measuring ERPs, disruptions within constituents evoked larger *anterior negativities* than disruptions between constituents (Cohn et al., 2014). As a larger negativity appeared to disruptions within the first constituent compared to between...constituents, these disruptions occurred *prior* to the constituent break, thereby preceding any possible discontinuity caused by crossing a constituent boundary. Thus, segmentation *cannot* rely solely on backward-looking processes based on discontinuity of meaning, but must involve forward-looking structural expectations (Cohn et al., 2014).

In ERP research of visual narratives, anterior negativities (Fig. 3b), often with a leftward or bilateralized distribution, appear to index the cost of combinatorial processing of the narrative grammar (Cohn & Kutas, 2017; Cohn et al., 2012, 2014). Similar anterior negativities have been shown to violations of expectancies and combinatorial processing in language (Friederici, 2011; Neville, Nicol, Barss, Forster, & Garrett, 1991; Yano, 2018) and music (Koelsch, Gunter, Wittfoth, & Sammler, 2005; Patel, 2003). Anterior negativities in visual narratives appear to be insensitive to situational discontinuity, such as an incongruous change of characters, but they are sensitive to differences in narrative patterning (Cohn & Kutas, 2017). This is the reverse of the N400, which is not attenuated by the presence of narrative structure but is sensitive to semantics (Cohn et al., 2012). Indeed, panels in sequences with only a narrative grammar but not semantic associations (described above, also Fig. 5) do not differ in the N400s they evoke compared to those in scrambled image sequences (Fig. 3a). However, differences between these sequences do manifest as a left lateralized anterior negativity (Cohn et al., 2012). Such findings suggest processing that is sensitive to purely combinatorial aspects of a narrative, dissociated from semantics.

### 3.3. Updating: Structural revision

When an incoming panel does not conform to the patterns predicted by an activated narrative schema, an *updating* process revises this structure. *Structural revision* should be warranted when incoming narrative categories contrast the expectations of a schema, which in turn may signal a change in constituent structure, as forecasted above. For example, in Fig. 4b, the narrative categories of panels 3 and 4 constitute a bigram of *Peak-Initial*. As this order does not occur in the canonical narrative schema (Table 1a), it should trigger a structural revision whereby a reader ends one constituent and begins another, along with building the higher level constituent that connects these subordinate constituents, as in Fig. 7. This updating in narrative processing thus reflects consequences of building a structure from schematic parts (e.g., “unification”), rather than updating of a situation model in semantic processing.

In ERPs, structural updating is also indexed by the P600 (Fig. 3c), when associated with the revision or updating required to integrate an incoming stimulus with prior context (Cohn & Kutas, 2015, 2017; Cohn et al., 2014). Although discussed above with regard to the updating of semantic information into a situation model, similar P600s have

been observed for revisions of grammar or its integration with semantic structure (Brouwer & Hoeks, 2013; Kuperberg, 2013). Indeed, P600s were first observed to errors in syntactic processing that warranted structural revision (Hagoort et al., 1993; Osterhout & Holcomb, 1992). Thus, the P600 here is taken to reflect updating processes that could operate on both structure (narrative) and/or semantics, given the nature of the incoming information (Brouwer, Fitz, & Hoeks, 2012), again consistent with theories that the P600 is related to more general updating processes (Donchin & Coles, 1988; King & Kutas, 1995; Van Petten & Luka, 2012).

Structural revision should thus occur when incoming category information contrasts schematic expectations. For example, P600s appear to panels after incongruous violations of narrative categories, but not after violations of semantic associations, despite both being followed by longer self-paced viewing times (Cohn, 2012). In addition, as discussed above, situation model updating is required of inference generation, as suggested by P600s and/or longer viewing times required to understand images following the omission of event information (Cohn & Kutas, 2015; Cohn & Wittenberg, 2015; Magliano et al., 2015, 2016). In such cases, the narrative structure often uses an Initial panel followed by a Release, forcing the climactic content of the missing Peak to be inferred. However, when comparing sequences where a Peak is omitted versus one where it remains present—even when *inference generation is held constant*—larger P600s are evoked by panels following deleted Peaks (Cohn & Kutas, 2015). In such cases, the P600 results from the disconfirmation of expecting a Peak but getting a Release, and needing to revise the structure, not just the inference generation.

As discussed, confounded predictions of constituent boundaries should also trigger structural revision. Such processes are implicated by P600s evoked by ill-formed groupings of constituent structure (Cohn et al., 2014) or by unexpected narrative constructions (Cohn & Kutas, 2017). In addition, longer viewing times have been observed to panels following constituent breaks and to panels both immediately and several panels after disruptions of constituents (Cohn, 2012). Thus, the incoming segmental structure of a narrative sequence must be reanalyzed if it contrasts the established structural predictions, and such reanalysis may have further downstream effects on the processing of a sequence.

#### 4. Narrative–semantics interfaces

Altogether, the PINS Model posits that representational levels of semantics and narrative structure combine with each other in the comprehension of visual narrative sequences. Such parallel processing is implicated in that the electrophysiological responses indexing semantic processing (N400) are insensitive to narrative structure (Cohn et al., 2012), while those indexing combinatorial processing of narrative are insensitive to semantic discontinuity (Cohn & Kutas, 2017). Moreover, these neurocognitive responses appear to have a similar time course (~300–900 ms), thus implying parallel mechanisms. The PINS Model further suggests various points of interface between these levels of representation.

First, since narrative categories may be cued by bottom-up semantic information within individual images, an interface exists between the initial stages in each of the representational levels, as in the gray double-headed arrows in Fig. 1. Such connections predict prototypical correspondences between particular semantic content and narrative category assignment (Cohn, 2012), such as preparatory actions corresponding to narrative Initials. However, as discussed above, semantic content is not determinative of narrative categories, as similar content can play multiple roles in a sequence depending on position (Cohn, 2014).

A second point of interface is between semantic expectancies and the predictions generated by narrative constituent structures, which, when violated, leads to the interface in the updating of a situation model and revision of a narrative structure. As discussed, prior work has taken semantic discontinuity between discourse units as a signal for the break between narrative constituents (Gernsbacher, 1990; Magliano & Zacks, 2011; Radvansky & Zacks, 2014). In the PINS Model, semantic relations between panels interface both between and across narrative constituent structures, but they do not necessarily motivate those narrative relationships. Indeed, certain holistic constituents in VNG may be characterized by discontinuity, such as Environmental-Conjunction (Fig. 6a), a construction composed of panels that change between characters (Cohn, 2013b, 2015). In cases where discontinuity and constituent breaks do align, the updating of a situation model would co-occur with the revision of a narrative structure, thus marking the interface at the third “updating” tier of Fig. 1.

While semantic discontinuity is a predictor of narrative segmentation, it is less predictive than bigrams disallowed by the canonical narrative schema, such as the *Peak-Initial* bigram in Figs. 4a and 7, which cue the break between constituents (Cohn & Bender, 2017). Participants are also better at discriminating whether panels have been switched in position across constituents than between constituents—and even better for nonadjacent than adjacent switched panels—suggesting that violating the well-formedness of constituents matters more than just adjacent discontinuity (Hagmann & Cohn, 2016). In addition, as discussed above, brain responses differentiate between disruptions within and between narrative constituents, *prior* to viewing a second panel that would signal semantic discontinuity (Cohn et al., 2014). Altogether, this suggests that, while semantic expectancies and discontinuities may *interface* with narrative constituents, they are not dependent on them.

Finally, because narrative categories have a prototypical correspondence to semantic event structures (i.e., the interfacing arrow in the “access” tier of Fig. 1), and because the narrative schema allows forward-looking predictions of upcoming narrative categories, such predictions may facilitate semantic expectancies. For example, if a reader is at an Establisher, the narrative schema would predict a subsequent Initial (narrative), but this may thus carry an additional expectancy of a semantic preparatory action (semantics), given the prototypical correspondence between Initials and event preparations. Thus, the PINS Model predicts the possibility for event structures to be anticipated on the basis of their interface with narrative categories, in addition to the semantic expectancies themselves. If supported by future research, this would be consistent with findings in sentence processing of coherent but violated semantic expectancies when syntactic structure remains well-formed (e.g., Thornhill & Van Petten, 2012; Wlotko & Federmeier, 2012).

## 5. Further implications

In sum, the PINS Model hypothesizes that online processing of visual narrative sequences combines the representational levels of semantics and narrative structure. Semantic processing involves assessing cues in images and integrating that information into a situation model based on expectations established by the continuity of a sequence. Parallel to this, semantic cues are mapped to narrative categories embedded in a canonical schematic sequence stored in memory. This schema allows structural predictions for subsequent narrative categories, which, if violated, trigger revision processes. Thus, a narrative grammar interfaces with semantics throughout the online processing of a visual narrative sequence. In the end, this narrative structure may fade from memory, while a situation model transfers to episodic long-term memory where the visual narrative's meaning persists into the future.

### 5.1. Domain-generalality

Given that narratives appear across modalities, the PINS Model may be applicable beyond visual sequences specifically. The idea that (semantic) processing mechanisms extend across modalities has long been assumed by discourse researchers, be it written, graphic, or filmed narratives (Gernsbacher, 1990; Magliano & Zacks, 2011; Magliano et al., 2012, in press; Radvansky & Zacks, 2014). Such domain-generalality is posited for narrative grammar, which, although formulated for drawn visual narratives, has been proposed as a domain-general narrative structure, which adapts to the affordances of different modalities (Cohn, 2013b, 2016a). It has thus also been applied to film (Amini et al., 2015; Cohn, 2016a; Yarhouse, 2017), motion graphics (Barnes, 2017), discourse (Fallon & Baker, 2016; Versluis, 2017), health communication (Sontag & Barnes, 2017), and computational generation of narrative (Andrews & Baber, 2014; Kim & Monroy-Hernandez, 2015; Martens & Cardona-Rivera, 2016). How other modalities alter a narrative grammar or invoke similar processing costs remains an important test of future research.

Neurocognitive aspects of domain-generalality are also implicated by visual narrative research. Specifically, growing evidence points to overlap between the processing mechanisms underlying wordless sequential images and the processing of language. Behavioral work has implicated similar working memory resources operating between language and visual narratives in inference generation (Magliano et al., 2015) as well as connections between verbal and visual narrative segmentation (Magliano et al., 2012) and production (Johnels, Hagberg, Gillberg, & Miniscalco, 2013), including populations with specific language impairment (Bishop & Adams, 1992; Bishop & Donlan, 2005).

ERP research has also implicated similar neurocognitive responses to manipulations of both sentences and visual narratives (Fig. 3). It has long been established that the N400 indexes semantic processing across domains (Kutas & Federmeier, 2011). As in language, larger N400s are evoked by anomalous and/or unexpected information in visual events and visual narratives (Amoruso et al., 2013; Cohn et al., 2012; Sitnikova et al., 2008; West &

Holcomb, 2002). Furthermore, visual narrative sequences can modulate the N400 effect to words that replace Peak panels (Manfredi, Cohn, & Kutas, 2017), suggesting cross-modal semantic resources. In addition, attenuation of N400 effects has been shown to both visual and verbal narratives in individuals with autism compared to neurotypical controls, implicating similar processing mechanisms across modalities (Coderre et al., 2018).

Combinatorial processing has also been argued to originate in domain-general mechanisms (Corballis, 1991; Hagoort, 2014; Jackendoff, 2011). This was first suggested in ERPs when violations to musical syntax evoked similar neural responses as linguistic syntax: anterior negativities and P600s (Koelsch et al., 2005; Patel, 2003). As described above, ostensibly similar neural responses have been observed to violations of narrative grammar in sequential images (Cohn & Kutas, 2015, 2017; Cohn et al., 2012, 2014). In addition, both Broca's and Wernicke's areas—brain regions long associated with language processing (Hagoort, 2014)—have been implicated in online visual narrative processing (Cohn & Maher, 2015; Nagai, Endo, & Takatsune, 2007; Osaka, Yaoi, Minamoto, & Osaka, 2014; Saft et al., 2013) and in tasks using visual narratives with aphasics (Bihle, Brownell, Powelson, & Gardner, 1986; Huber & Gleber, 1982).

Altogether, these findings suggest growing evidence for the domain-general nature of both semantic and grammatical processing. As mentioned above, cross-domain processing is posited by discourse approaches (Gernsbacher, 1990; Magliano & Zacks, 2011; Magliano et al., 2012, in press; Radvansky & Zacks, 2014); however, these ERPs observed to combinatorial processing do not necessarily index “discourse” processing: Violations to sequential images appear at a *narrative* level, while those to syntax in language appear at the *sentence* level. Yet the electrophysiological responses for processing sentences, discourse, and visual narratives appear to be similar regardless of the “level” of processing. This implies that such similarities are not just constrained to parallel levels of information processing—that is, visual narrative corresponding to verbal discourse specifically (e.g., Bateman & Wildfeuer, 2014; Magliano et al., 2013)—but to mechanisms of sequence processing more generally (Christiansen, Conway, & Onnis, 2011). Future research thus should target the degree to which modalities' processing may overlap and/or deviate from each other.

These similarities between electrophysiological responses across modalities also raise questions related to processing involving *multimodal* interactions. Indeed, while we here focused on wordless visual narratives, image sequences typically appear in multimodal relationships, such as combined with written language. ERP research has shown that visual narratives integrated with text (Manfredi et al., 2017) and paired with auditory speech/sounds (Manfredi, Cohn, De Araújo Andreoli, & Boggio, 2018) elicit N400s and late effects with the same time course as in unimodal contexts, consistent with other studies of crossmodal and multimodal processing (Coco, Araujo, & Petersson, 2017; Liu, Wang, & Jin, 2009; Liu, Wang, Wu, & Meng, 2011; Weissman & Tanner, 2018; Wu & Coulson, 2005). Such findings support that the N400 reflects semantic activation that is only semisensitive to modality-specific inputs (as suggested by variance in the scalp distribution of the N400 to different modalities). Subsequent processing (like late positivities or late negativities) likewise follow a similar time course of processing across modalities. Such findings imply that multimodal semantic processing follows similar back-end

processes as unimodal processing, albeit receiving complex front-end stimulation from multiple sensory inputs (text, images, sounds, etc.). What is more in question is how these different semantic sources interact with “grammatical” processing, which can involve complex interactions depending on which modalities do or do not use combinatorial structures (Cohn, 2016b). Further investigation of such multimodal processing is a logical next step for the PINS Model and its relationship with neurocognitive models of language and other expressive systems.

## 5.2. Visual narrative fluency

Finally, though the cognitive mechanisms of access, prediction, and updating should be universally accessible to comprehenders, visual narratives appear to require more than just perceptual and event processing. Rather, visual narrative comprehension requires a “fluency” to be comprehended, which is developed as individuals age, given exposure to visual narratives. Various healthy populations who lack exposure to visual narratives have difficulty creating (Wilson, 2016) and comprehending visual narratives (Byram & Garforth, 1980; Fussell & Haaland, 1978; Liddell, 1997; Núñez & Cooperrider, 2013). In these cases, basic meaningful construals between images are not recognized; they do not connect that characters in one image repeat in subsequent images—a lack of sequential *continuity* of referential cohesion. Rather, each image is perceived as an independent scene. This basic ability to connect images appears to developmentally emerge between the ages of 4 and 6 (Bornens, 1990; Trabasso & Nickels, 1992), given exposure to visual narratives (Byram & Garforth, 1980; Fussell & Haaland, 1978; Liddell, 1997; Núñez & Cooperrider, 2013). Subsequent complexity, such as the accurate inference of omitted information, appears to develop at even later ages (Nakazawa, 2016).

Recent work has also assessed expertise in experiments on visual narrative research using the *Visual Language Fluency Index* (VLFI) score. This metric has suggested modulation of sequential image comprehension even between self-described “comic readers.” VLFI scores reveal that frequency of comic reading/drawing correlates with ERP amplitudes (Cohn & Kutas, 2015; Cohn & Maher, 2015; Cohn et al., 2012), reaction times (Cohn et al., 2012), self-paced viewing times (Cohn & Maher, 2015; Cohn & Wittenberg, 2015), accuracy detection (Hagmann & Cohn, 2016), and perceived ease of segmentation (Cohn & Bender, 2017).

Fluency modulation also extends beyond general familiarity with comics and visual narratives, but to readership of *specific types* of comics. A recent experiment examined the processing of Environmental-Conjunction (Fig. 6a), which appears more in Japanese manga than American comics (Cohn, 2013a, in press). Participants recruited cognitive resources differently depending on their readership of Japanese manga while growing up (Cohn & Kutas, 2017): More frequent manga readers showed a greater anterior negativity and a reduced P600, with the opposite pattern appearing to less frequent manga readers (Fig. 3b and c). This result points to a *fluency* not just for visual narratives generally but also for the particular patterns used in narrative grammars for specific visual languages, similar to the fluency for specific verbal or signed languages.

Overall, evidence of fluency for visual narratives contrasts popular assumptions that sequential image understanding is transparent or developmentally inevitable (McCloud, 1993), or relies solely on basic perceptual or event processing (Berliner & Cohen, 2011; Radvansky & Zacks, 2014) or general intelligence (Ramos & Die, 1986). Rather, comprehenders must acquire and encode specific knowledge from the visual narratives that they read, which in turn habituate them to understand sequential images on the basis of these patterns. In turn, familiarity with such patterns modulate the general cognitive processes that guide their comprehension (i.e., prediction, updating) (Cohn & Kutas, 2017). Thus, the instantiation of the processing mechanisms outlined in the PINS Model (Fig. 1) may vary depending on the frequency and type of patterns found in the visual narratives a person engages. A promising avenue for future research can investigate the degree to which processing a visual narrative is modulated by general fluency and/or specific narrative patterns.

## **6. Concluding remarks and future directions**

To conclude, the PINS Model posits that comprehension of visual narratives negotiates both semantic and narrative processing. Comprehenders access semantic information about objects and events in images. A narrative grammar assigns images categorical roles and groups them into hierarchic constituents so that semantic information can be organized sequentially and construct a coherent situation model. While the narrative grammar eventually fades from memory, a situation model shifts from working memory to episodic long-term memory as the meaning of a visual narrative is retained into the future. In online processing, both semantics and narrative use forward- and backward-looking mechanisms in an iterative cycle of prediction and updating in the ongoing processing of sequential images. These operations appear to recruit similar neural mechanisms as language processing and require fluency in both general and specific visual narrative systems. Such findings point toward domain-general mechanisms guiding the comprehension of both verbal and visual languages, modulated by experience.

Research on visual narrative processing has only been growing with seriousness over the past decade, and thus affords great potential for future research. The rich complexity of visual narrative systems transcends simple perceptual or spatial cognition, and studying such mechanisms can inform fundamental domains of cognition. Given that the ability to draw is unique to humans—and both single and sequential images are among our oldest records—perhaps, it is time to embrace the study of these systems within the cognitive sciences as rigorously as other aspects of human expression.

## **Acknowledgments**

Thanks are given to Katalin Bálint, Emily Coderre, Tom Foulsham, Ray Jackendoff, Gina Kuperberg, Fons Maes, Joe Magliano, and Joost Schilperoord for comments on earlier drafts

and/or discussions contributing to this manuscript. I dedicate this paper to my friend and collaborator, Dr. Martin Paczynski (1974–2018), whom I first brainstormed this model with, together on the whiteboard of a small conference room at Tufts University in 2011.

## References

- Amini, F., Riche, N. H., Lee, B., Hurter, C., & Irani, P. (2015). Understanding data videos: looking at narrative visualization through the cinematography lens. In *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems* (pp. 1459–1468). New York: ACM.
- Amoruso, L., Gelormini, C., Aboitiz, F., Alvarez González, M., Manes, F., Cardona, J., & Ibanez, A. (2013). N400 ERPs for actions: Building meaning in context. *Frontiers in Human Neuroscience*, 7, 1–16. <https://doi.org/10.3389/fnhum.2013.00057>
- Andrews, D., & Baber, C. (2014). Visualizing interactive narratives: Employing a branching comic to tell a story and show its readings. In *Proceedings of the 32nd annual ACM conference on Human factors in computing systems* (pp. 1895–1904). New York: ACM.
- Baggio, G., van Lambalgen, M., & Hagoort, P. (2008). Computing and recomputing discourse models: An ERP study. *Journal of Memory and Language*, 59(1), 36–53. <https://doi.org/10.1016/j.jml.2008.02.005>.
- Barnes, S. (2017). Studies in the efficacy of motion graphics: The impact of narrative structure on exposition. *Digital Journalism*, 5(10), 1–21. <https://doi.org/10.1080/21670811.2017.1279020>
- Bateman, J. A. (2007). Towards a grande paradigmatique of film: Christian Metz reloaded. *Semiotica*, 2007(167), 13–64.
- Bateman, J. A., & Schmidt, K.-H. (2012). *Multimodal film analysis: How films mean*. New York: Routledge.
- Bateman, J. A., & Wildfeuer, J. (2014). A multimodal discourse theory of visual narrative. *Journal of Pragmatics*, 74, 180–208. <https://doi.org/10.1016/j.pragma.2014.10.001>.
- de Beaugrande, R. (1982). The story of grammars and the grammar of stories. *Journal of Pragmatics*, 6, 383–422.
- van Berkum J. J. A. (2009). The neuropragmatics of “simple” utterance comprehension: An ERP review. In U. Sauerland & K. Yatsushiro (Eds.), *Semantics and pragmatics: From experiment to theory* (pp. 276–316). Basingstoke, UK: Palgrave.
- van Berkum, J. J. A. (2012). The electrophysiology of discourse and conversation. In M. J. Spivey, M. Joannisse, & K. McRae (Eds.), *The Cambridge handbook of psycholinguistics*. Cambridge, UK: Cambridge University Press.
- van Berkum, J. J. A., Brown, C., Zwitserlood, P., Kooijman, V., & Hagoort, P. (2005). Anticipating upcoming words in discourse: Evidence from ERPs and reading times. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 31(3), 443–467.
- Berliner, T., & Cohen, D. J. (2011). The illusion of continuity: Active perception and the classical editing system. *Journal of Film and Video*, 63(1), 44–63.
- Bihrlé, A. M., Brownell, H. H., Powelson, J. A., & Gardner, H. (1986). Comprehension of humorous and nonhumorous materials by left and right brain-damaged patients. *Brain and Cognition*, 5, 399–411.
- Bishop, D. V. M., & Adams, C. (1992). Comprehension problems in children with specific language impairment: Literal and inferential meaning. *Journal of Speech, Language, and Hearing Research*, 35(1), 119–129. <https://doi.org/10.1044/jshr.3501.119>.
- Bishop, D. V. M., & Donlan, C. (2005). The role of syntax in encoding and recall of pictorial narratives: Evidence from specific language impairment. *British Journal of Developmental Psychology*, 23(1), 25–46. <https://doi.org/10.1348/026151004X20685>.
- Black, J. B., & Wilensky, R. (1979). An evaluation of story grammars. *Cognitive Science*, 3, 213–230.
- Bornens, M.-T. (1990). Problems brought about by “reading” a sequence of pictures. *Journal of Experimental Child Psychology*, 49(2), 189–226. [https://doi.org/10.1016/0022-0965\(90\)90055-D](https://doi.org/10.1016/0022-0965(90)90055-D).

- Bott, O. (2010). *The processing of events*. Vol. 162. Amsterdam: John Benjamins.
- Branigan, E. (1992). *Narrative comprehension and film*. London: Routledge.
- Brewer, W. F. (1985). The story schema: Universal and culture-specific properties. In D. R. Olson, N. Torrance, & A. Hildyard (Eds.), *Literacy, language, and learning* (pp. 167–194). Cambridge, UK: Cambridge University Press.
- Brewer, W. F., & Lichtenstein, E. H. (1981). Event schemas, story schemas, and story grammars. In J. Long & A. D. Baddeley (Eds.), *Attention and performance IX* (pp. 363–379). Hillsdale, NJ: Erlbaum.
- Brouwer, H., Crocker, M. W., Venhuizen, N. J., & Hoeks, J. C. J. (2016). A neurocomputational model of the N400 and the P600 in language processing. *Cognitive Science*, 41(S6), 1318–1352. n/a-n/a. <https://doi.org/10.1111/cogs.12461>
- Brouwer, H., Fitz, H., & Hoeks, J. (2012). Getting real about semantic illusions: Rethinking the functional role of the P600 in language comprehension. *Brain Research*, 1446, 127–143. <https://doi.org/10.1016/j.brainres.2012.01.055>.
- Brouwer, H., & Hoeks, J. C. J. (2013). A time and place for language comprehension: Mapping the N400 and the P600 to a minimal cortical network. *Frontiers in Human Neuroscience*, 7, 758. <https://doi.org/10.3389/fnhum.2013.00758>.
- Buckland, W. (2000). *The cognitive semiotics of film*. Cambridge, UK: Cambridge University Press.
- Butcher, S. H. (1902). *The poetics of Aristotle* (3rd ed.). London: Macmillan.
- Byram, M. L., & Garforth, C. (1980). Research and testing non-formal education materials: A multi-media extension project in Botswana. *Educational Broadcasting International*, 13(4), 190–194.
- Carroll, J. M. (1980). *Toward a structural psychology of cinema*. The Hague: Mouton.
- Chomsky, N. (1965). *Aspects of the theory of syntax*. Cambridge, MA: MIT Press.
- Christiansen, M. H., Conway, C. M., & Onnis, L. (2011). Similar neural correlates for language and sequential learning: Evidence from event-related brain potentials. *Language and Cognitive Processes*, 27(2), 231–256. <https://doi.org/10.1080/01690965.2011.606666>.
- Coco, M. I., Araujo, S., & Petersson, K. M. (2017). Disentangling stimulus plausibility and contextual congruency: Electro-physiological evidence for differential cognitive dynamics. *Neuropsychologia*, 96, 150–163. <https://doi.org/10.1016/j.neuropsychologia.2016.12.008>.
- Coderre, E. L., Cohn, N., Slipher, S. K., Chernenok, M., Ledoux, K., & Gordon, B. (2018). Visual and linguistic narrative comprehension in autism spectrum disorders: Neural evidence for modality-independent impairments. *Brain and Language*, 186, 44–59.
- Cohn, N. (2010a). Extra! Extra! Semantics in comics!: The conceptual structure of *Chicago Tribune* advertisements. *Journal of Pragmatics*, 42(11), 3138–3146. <https://doi.org/10.1016/j.pragma.2010.04.016>.
- Cohn, N. (2010b). The limits of time and transitions: Challenges to theories of sequential image comprehension. *Studies in Comics*, 1(1), 127–147.
- 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. <https://doi.org/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. <https://doi.org/10.1111/cogs.12116>.
- Cohn, N. (2015). Narrative conjunction's junction function: The interface of narrative grammar and semantics in sequential images. *Journal of Pragmatics*, 88, 105–132. <https://doi.org/10.1016/j.pragma.2015.09.001>.
- Cohn, N. (2016a). From visual narrative grammar to filmic narrative grammar: The narrative structure of static and moving images. In J. Wildfeuer & J. A. Bateman (Eds.), *Film text analysis: New perspectives on the analysis of filmic meaning* (pp. 94–117). London: Routledge.
- Cohn, N. (2016b). A multimodal parallel architecture: A cognitive framework for multimodal interactions. *Cognition*, 146, 304–323. <https://doi.org/10.1016/j.cognition.2015.10.007>.

- Cohn, N. (In press). Structural complexity in visual narratives: Theory, brains, and cross-cultural diversity. In M. Grishakova & M. Poulaki (Eds.), *Narrative complexity and media: Experiential and cognitive interfaces* (pp. 174–199). Lincoln: University of Nebraska Press.
- Cohn, N., & Bender, P. (2017). Drawing the line between constituent structure and coherence relations in visual narratives. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, 43(2), 289–301. <https://doi.org/10.1037/xlm0000290>.
- Cohn, N., Jackendoff, R., Holcomb, P. J., & Kuperberg, G. R. (2014). The grammar of visual narrative: Neural evidence for constituent structure in sequential image comprehension. *Neuropsychologia*, 64, 63–70. <https://doi.org/10.1016/j.neuropsychologia.2014.09.018>.
- Cohn, N., & Kutas, M. (2015). Getting a cue before getting a clue: Event-related potentials to inference in visual narrative comprehension. *Neuropsychologia*, 77, 267–278. <https://doi.org/10.1016/j.neuropsychologia.2015.08.026>.
- Cohn, N., & Kutas, M. (2017). What's your neural function, visual narrative conjunction? Grammar, meaning, and fluency in sequential image processing. *Cognitive Research: Principles and Implications*, 2(27), 1–13. <https://doi.org/10.1186/s41235-017-0064-5>.
- Cohn, N., & Maher, S. (2015). The notion of the motion: The neurocognition of motion lines in visual narratives. *Brain Research*, 1601, 73–84. <https://doi.org/10.1016/j.brainres.2015.01.018>.
- Cohn, N., & Paczynski, M. (2013). Prediction, events, and the advantage of Agents: The processing of semantic roles in visual narrative. *Cognitive Psychology*, 67(3), 73–97. <https://doi.org/10.1016/j.cogpsych.2013.07.002>.
- Cohn, N., Paczynski, M., Jackendoff, R., Holcomb, P. J., & Kuperberg, G. R. (2012). (Pea)nuts and bolts of visual narrative: Structure and meaning in sequential image comprehension. *Cognitive Psychology*, 65(1), 1–38. <https://doi.org/10.1016/j.cogpsych.2012.01.003>.
- Cohn, N., Paczynski, M., & Kutas, M. (2017). Not so secret agents: Event-related potentials to semantic roles in visual event comprehension. *Brain and Cognition*, 119, 1–9. <https://doi.org/10.1016/j.bandc.2017.09.001>.
- Cohn, N., Pederson, K., & Taylor, R. (2017). A picture is worth more words over time: Multimodality and narrative structure across eight decades of American superhero comics. *Multimodal Communication*, 6(1), 19–37. <https://doi.org/10.1515/mc-2017-0003>.
- Cohn, N., & Wittenberg, E. (2015). Action starring narratives and events: Structure and inference in visual narrative comprehension. *Journal of Cognitive Psychology*, 27(7), 812–828. <https://doi.org/10.1080/20445911.2015.1051535>.
- Colin, M. (1995). The Grande Syntagmatique revisited. In W. Buckland (Ed.), *The film spectator: From sign to mind* (pp. 45–86). Amsterdam: Amsterdam University Press.
- Corballis, M. (1991). *The lopsided ape: Evolution of the generative mind*. Oxford, England: Oxford University Press.
- Culicover, P. W., & Jackendoff, R. (2005). *Simpler syntax*. Oxford, England: Oxford University Press.
- Cutting, J. E. (2016). Narrative theory and the dynamics of popular movies. *Psychonomic Bulletin & Review*, 23(6), 1713–1743. <https://doi.org/10.3758/s13423-016-1051-4>
- DeLong, K. A., Troyer, M., & Kutas, M. (2014). Pre-processing in sentence comprehension: Sensitivity to likely upcoming meaning and structure. *Language and Linguistics Compass*, 8(12), 631–645. <https://doi.org/10.1111/lnc3.12093>.
- DeLong, K. A., Urbach, T. P., & Kutas, M. (2005). Probabilistic word pre-activation during language comprehension inferred from electrical brain activity. *Nature Neuroscience*, 8(8), 1117–1121.
- van Dijk, T., & Kintsch, W. (1983). *Strategies of discourse comprehension*. New York: Academic Press.
- Donchin, E., & Coles, M. G. H. (1988). Is the P300 component a manifestation of context updating? *Behavioral and Brain Sciences*, 11(03), 357–374. <https://doi.org/doi:10.1017/S0140525X00058027>.
- Draschkow, D., Heikel, E., Vö, M. L. H., Fiebach, C. J., & Sassenhagen, J. (2018). No evidence from MVPA for different processes underlying the N300 and N400 incongruity effects in object-scene processing. *Neuropsychologia*, 120, 9–17. <https://doi.org/10.1016/j.neuropsychologia.2018.09.016>.

- Eisenberg, M. L., Zacks, J. M., & Flores, S. (2018). Dynamic prediction during perception of everyday events. *Cognitive Research: Principles and Implications*, 3(1), 53. <https://doi.org/10.1186/s41235-018-0146-z>.
- Fallon, T. J., & Baker, M. (2016). Seeking the effects of visual narrative grammar on the written dialogue production of ESL students at Japanese universities: A proposed experiment. *Journal of Nagoya Gakuin University: Language and Culture*, 28(1), 63–68. <https://doi.org/10.15012/00000777>.
- Ferretti, T. R., Rohde, H., Kehler, A., & Crutchley, M. (2009). Verb aspect, event structure, and coreferential processing. *Journal of Memory and Language*, 61(2), 191–205. <https://doi.org/10.1016/j.jml.2009.04.001>.
- Foulsham, T., Wybrow, D., & Cohn, N. (2016). Reading without words: Eye movements in the comprehension of comic strips. *Applied Cognitive Psychology*, 30, 566–579. <https://doi.org/10.1002/acp.3229>.
- Friederici, A. D. (2011). The brain basis of language processing: From structure to function. *Physiological Review*, 91(4), 1357–1392. <https://doi.org/10.1152/physrev.00006.2011>.
- Fussell, D., & Haaland, A. (1978). Communicating with pictures in Nepal: Results of practical study used in visual education. *Educational Broadcasting International*, 11(1), 25–31.
- Ganis, G., & Kutas, M. (2003). An electrophysiological study of scene effects on object identification. *Cognitive Brain Research*, 16(2), 123–144.
- Garnham, A. (1983). What's wrong with story grammars? *Cognition*, 15, 145–154.
- 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 Erlbaum.
- Giglio, A. C. A., Minati, L., & Boggio, P. S. (2013). Throwing the banana away and keeping the peel: Neuroelectric responses to unexpected but physically feasible action endings. *Brain Research*, 1532, 56–62. <https://doi.org/10.1016/j.brainres.2013.08.017>.
- Glanzer, M., Fischer, B., & Dorfman, D. (1984). Short-term storage in reading. *Journal of Verbal Learning and Verbal Behavior*, 17, 467–486.
- Goldberg, A. (1995). *Constructions: A construction grammar approach to argument structure*. Chicago, IL: University of Chicago Press.
- Haberlandt, K. (1980). Story grammar and reading time of story constituents. *Poetics*, 9(1–3), 99–118.
- 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.
- Hagmann, C. E., & Cohn, N. (2016). The pieces fit: Constituent structure and global coherence of visual narrative in RSVP. *Acta Psychologica*, 164, 157–164. <https://doi.org/10.1016/j.actpsy.2016.01.011>.
- Hagoort, P. (2005). On Broca, brain, and binding: A new framework. *Trends in Cognitive Sciences*, 9(9), 416–423.
- Hagoort, P. (2014). Nodes and networks in the neural architecture for language: Broca's region and beyond. *Current Opinion in Neurobiology*, 28, 136–141. <https://doi.org/10.1016/j.conb.2014.07.013>.
- Hagoort, P. (2016). MUC (Memory, Unification, Control): A model on the neurobiology of language beyond single word processing. In G. Hickok & S. L. Small (Eds.), *Neurobiology of language* (pp. 339–347). San Diego: Academic Press.
- Hagoort, Peter, Brown, Colin, & Groothusen, Jolanda (1993). The syntactic positive shift (sps) as an erp measure of syntactic processing. *Language and Cognitive Processes*, 8(4), 439–483. doi:10.1080/01690969308407585.
- Hamm, J. P., Johnson, B. W., & Kirk, I. J. (2002). Comparison of the N300 and N400 ERPs to picture stimuli in congruent and incongruent contexts. *Clinical Neurophysiology*, 113(8), 1339–1350. [https://doi.org/10.1016/S1388-2457\(02\)00161-X](https://doi.org/10.1016/S1388-2457(02)00161-X).
- Hoeks, J. C. J., & Brouwer, H. (2014). Electrophysiological research on conversation and discourse. In T. M. Holtgraves (Ed.), *The Oxford handbook of language and social psychology* (pp. 365–386). Oxford, UK: Oxford University Press.
- Huber, W., & Gleber, J. (1982). Linguistic and nonlinguistic processing of narratives in aphasia. *Brain and Language*, 16, 1–18.

- Huff, M., Meitz, T. G. K., & Papenmeier, F. (2014). Changes in situation models modulate processes of event perception in audiovisual narratives. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, *40*(5), 1377–1388. <https://doi.org/10.1037/0033-2909.133.2.273>.
- Huff, M., & Schwan, S. (2012). Do not cross the line: Heuristic spatial updating in dynamic scenes. *Psychonomic Bulletin & Review*, *19*(6), 1065–1072. <https://doi.org/10.3758/s13423-012-0293-z>.
- Hutson, J. P., Magliano, J., & Loschky, L. C. (2018). Understanding moment-to-moment processing of visual narratives. *Cognitive Science*, *42*(8), 2999–3033. doi: 10.1111/cogs.12699.
- Inui, T., & Miyamoto, K. (1981). The time needed to judge the order of a meaningful string of pictures. *Journal of Experimental Psychology: Human Learning and Memory*, *7*(5), 393–396.
- Jackendoff, R. (2002). *Foundations of language: Brain, meaning, grammar, evolution*. Oxford, UK: Oxford University Press.
- Jackendoff, R. (2011). What is the human language faculty?: Two views. *Language*, *87*(3), 586–624.
- Johnels, J. Å., Hagberg, B., Gillberg, C., & Miniscalco, C. (2013). Narrative retelling in children with neurodevelopmental disorders: Is there a role for nonverbal temporal-sequencing skills? *Scandinavian Journal of Psychology*, *54*(5), 376–385. <https://doi.org/10.1111/sjop.12067>.
- Kim, J., & Monroy-Hernandez, A. (2016). Storia: Summarizing Social Media Content based on Narrative Theory using Crowdsourcing. In *Proceedings of the 19th ACM Conference on Computer-Supported Cooperative Work & Social Computing* (pp. 1018–1027). New York, NY: ACM.
- King, J. W., & Kutas, M. (1995). Who did what and when? using word- and clause-level ERPs to monitor working memory usage in reading. *Journal of Cognitive Neuroscience*, *7*(3), 376–395.
- Kintsch, W. (1988). The role of knowledge in discourse comprehension: A construction-integration model. *Psychological Review*, *95*(2), 163–182.
- Koelsch, S., Gunter, T. C., Wittfoth, M., & Sammler, D. (2005). Interaction between syntax processing in language and in music: An ERP study. *Journal of Cognitive Neuroscience*, *17*(10), 1565–1577.
- Kuperberg, G. R. (2013). The pro-active comprehender: What event-related potentials tell us about the dynamics of reading comprehension. In B. Miller, L. Cutting, & P. McCardle (Eds.), *Unraveling the behavioral, neurobiological, and genetic components of reading comprehension* (pp. 176–192). Baltimore: Paul Brookes Publishing.
- Kuperberg, G. R. (2016). Separate streams or probabilistic inference? What the N400 can tell us about the comprehension of events. *Language, Cognition and Neuroscience*, *31*(5), 602–616. <https://doi.org/10.1080/23273798.2015.1130233>.
- Kurby, C. A., & Zacks, J. M. (2012). Starting from scratch and building brick by brick in comprehension. *Memory & Cognition*, *40*(5), 812–826. <https://doi.org/10.3758/s13421-011-0179-8>.
- Kutas, M., Delong, K. A., & Smith, N. J. (2011). A look around at what lies ahead: Prediction and predictability in language processing. In M. Bar (Ed.), *Predictions in the brain: Using our past to generate a future* (pp. 190–207). Oxford, UK: Oxford University 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.
- Kutas, M., Kluender, R., Barkley, C., & Amsel, B. D. (2017). Language. In J. Cacioppo, L. Tassinary, & G. Berntson (Eds.), *Handbook of psychophysiology* (4th ed) (pp. 511–525). Cambridge, UK: Cambridge University Press.
- Laubrock, J., Hohenstein, S., & Kümmerer, M. (2018). Attention to comics: Cognitive processing during the reading of graphic literature. In A. Dunst, J. Laubrock, & J. Wildfeuer (Eds.), *Empirical comics research: Digital, multimodal, and cognitive methods* (pp. 239–263). New York: Routledge.
- Liddell, C. (1997). Every picture tells a story—or does it?: Young South African children interpreting pictures. *Journal of Cross-Cultural Psychology*, *28*(3), 266–283.

- Liu, B., Wang, Z., & Jin, Z. (2009). The integration processing of the visual and auditory information in videos of real-world events: An ERP study. *Neuroscience Letters*, *461*(1), 7–11. <https://doi.org/10.1016/j.neulet.2009.05.082>.
- Liu, B., Wang, Z., Wu, G., & Meng, X. (2011). Cognitive integration of asynchronous natural or non-natural auditory and visual information in videos of real-world events: An event-related potential study. *Neuroscience*, *180*, 181–190.
- Loschky, L. C., Hutson, J. P., Smith, M. E., Smith, T. J., & Magliano, J. (2018). Viewing static visual narratives through the lens of the scene perception and event comprehension theory (SPECT). In A. Dunst, J. Laubrock, & J. Wildfeuer (Eds.), *Empirical comics research: Digital, multimodal, and cognitive methods* (pp. 217–238). London: Routledge.
- Loschky, L. C., Magliano, J., Larson, A. M., & Smith, T. J. (2019). The Scene Perception & Event Comprehension Theory (SPECT) applied to visual narratives. *Topics in Cognitive Science*.
- Magliano, J. P., Dijkstra, K., & Zwaan, R. A. (1996). Generating predictive inferences while viewing a movie. *Discourse Processes*, *22*, 199–224.
- Magliano, J. P., Higgs, K., & Clinton, J. A. (In press). Sources of complexity in comprehension across modalities of narrative experience. In M. Grishakova & M. Poulaki (Eds.), *Narrative complexity and media: Experiential and cognitive interfaces*. Lincoln: University of Nebraska Press.
- Magliano, J. P., Kopp, K., Higgs, K., & Rapp, D. N. (2016). Filling in the gaps: Memory implications for inferring missing content in graphic narratives. *Discourse Processes*, *54*(8), 569–582. <https://doi.org/10.1080/0163853x.2015.1136870>
- Magliano, J. P., Kopp, K., Mc Nerney, M. W., Radvansky, G. A., & Zacks, J. M. (2012). Aging and perceived event structure as a function of modality. *Aging, Neuropsychology, and Cognition*, *19*(1–2), 264–282. <https://doi.org/10.1080/13825585.2011.633159>.
- Magliano, J. P., Larson, A. M., Higgs, K., & Loschky, L. C. (2015). The relative roles of visuospatial and linguistic working memory systems in generating inferences during visual narrative comprehension. *Memory & Cognition*, *44*(2), 207–219. <https://doi.org/10.3758/s13421-015-0558-7>.
- Magliano, J. P., Loschky, L. C., Clinton, J. A., & Larson, A. M. (2013). Is reading the same as viewing? an exploration of the similarities and differences between processing text- and visually based narratives. In B. Miller, L. Cutting, & P. McCardle (Eds.), *Unraveling the behavioral, neurobiological, and genetic components of reading comprehension* (pp. 78–90). Baltimore, MD: Brookes.
- 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. <https://doi.org/10.1111/j.1551-6709.2011.01202.x>.
- Mandler, J. M., & Johnson, N. S. (1977). Remembrance of things parsed: Story structure and recall. *Cognitive Psychology*, *9*, 111–151.
- Manfredi, M., Cohn, N., De Araújo Andreoli, M., & Boggio, P. S. (2018). Listening beyond seeing: Event-related potentials to audiovisual processing in visual narrative. *Brain and Language*, *185*, 1–8. <https://doi.org/10.1016/j.bandl.2018.06.008>.
- Manfredi, M., Cohn, N., & Kutas, M. (2017). When a hit sounds like a kiss: An electrophysiological exploration of semantic processing in visual narrative. *Brain and Language*, *169*, 28–38. <https://doi.org/10.1016/j.bandl.2017.02.001>.
- Martens, C., & Cardona-Rivera, R. E. (2016). *Generating Abstract Comics*. Paper presented at the Interactive Storytelling: 9th International Conference on Interactive Digital Storytelling, ICIDS 2016, Los Angeles, CA, November 15–18, 2016, Proceedings, Los Angeles, CA.
- McCloud, S. (1993). *Understanding comics: The invisible art*. New York: HarperCollins.
- McKoon, G., & Ratcliff, R. (1986). Inferences about predictable events. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *12*(1), 82.
- McNamara, D. S., & Magliano, J. (2009). Toward a comprehensive model of comprehension. *Psychology of learning and motivation*, *51*, 297–384.

- McPherson, W. B., & Holcomb, P. J. (1999). An electrophysiological investigation of semantic priming with pictures of real objects. *Psychophysiology*, *36*(1), 53–65.
- Nagai, M., Endo, N., & Takatsune, K. (2007). Measuring brain activities related to understanding using Near-Infrared Spectroscopy (NIRS). In M. J. Smith & S. G. (Eds.), *Human interface and the management of information: Methods, techniques and tools in information design* (Vol. 4557, pp. 884–893). Heidelberg: Springer Berlin.
- Nakazawa, J. (2016). Manga literacy and manga comprehension in Japanese children. In N. Cohn (Ed.), *The visual narrative reader* (pp. 157–184). London: Bloomsbury.
- 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.
- Nieuwland, M. S., & Van Berkum, J. J. A. (2005). Testing the limits of the semantic illusion phenomenon: ERPs reveal temporary semantic change deafness in discourse comprehension. *Cognitive Brain Research*, *24*(3), 691–701. <https://doi.org/10.1016/j.cogbrainres.2005.04.003>.
- Núñez, R., & Cooperrider, K. (2013). The tangle of space and time in human cognition. *Trends in Cognitive Sciences*, *17*(5), 220–229. <https://doi.org/10.1016/j.tics.2013.03.008>.
- Olivares, E. I., Iglesias, J., & Bobes, M. A. A. (1999). Searching for face-specific long latency ERPs: A topographic study of effects associated with mismatching features. *Cognitive Brain Research*, *7*(3), 343–356. [https://doi.org/10.1016/s0926-6410\(98\)00038-x](https://doi.org/10.1016/s0926-6410(98)00038-x)
- Osaka, M., Yaoi, K., Minamoto, T., & Osaka, N. (2014). Serial changes of humor comprehension for four-frame comic Manga: An fMRI study. *Scientific Reports*, *4*(5828), 1–9. <https://doi.org/10.1038/srep05828>
- Osterhout, L., & Holcomb, P. (1992). Event-related potentials elicited by syntactic anomaly. *Journal of Memory and Language*, *31*, 758–806.
- Paczynski, M., Jackendoff, R., & Kuperberg, G. (2014). When events change their nature: The neurocognitive mechanisms underlying aspectual coercion. *Journal of Cognitive Neuroscience*, *26*(9), 1905–1917. [https://doi.org/10.1162/jocn\\_a\\_00638](https://doi.org/10.1162/jocn_a_00638).
- Papenmeier, F., Boss, A., & Mahlke, A.-K. (2018). Action goal changes caused by agents and patients both induce global updating of event models. *Journal of Experimental Psychology: Learning, Memory, & Cognition*. <https://doi.org/10.1037/xlm0000651>
- Patel, A. D. (2003). Language, music, syntax and the brain. *Nature Neuroscience*, *6*(7), 674–681. <https://doi.org/10.1038/nn1082>.
- Petersen, R. S. (2011). *Comics, manga, and graphic novels: A history of graphic narratives*. Santa Barbara, CA: ABC-CLIO.
- Radvansky, G. A., & Zacks, J. (2014). *Event cognition*. Oxford, UK: Oxford University Press.
- Ramos, M. C., & Die, A. H. (1986). The Wais-R Picture Arrangement Subtest: What do scores indicate? *The Journal of General Psychology*, *113*(3), 251–261. <https://doi.org/10.1080/00221309.1986.9711036>.
- Reid, V. M., & Striano, T. (2008). N400 involvement in the processing of action sequences. *Neuroscience Letters*, *433*(2), 93–97. <https://doi.org/10.1016/j.neulet.2007.12.066>.
- Rumelhart, D. E. (1975). Notes on a schema for stories. In D. Bobrow & A. Collins (Eds.), *Representation and understanding* (pp. 211–236). New York: Academic Press.
- Saft, C., Lissek, S., Hoffmann, R., Nicolas, V., Tegenthoff, M., Juckel, G., & Brüne, M. (2013). Mentalizing in preclinical Huntington’s disease: An fMRI study using cartoon picture stories. *Brain Imaging and Behavior*, *7*(2), 154–162. <https://doi.org/10.1007/s11682-012-9209-9>.
- Saraceni, M. (2016). Relatedness: Aspects of textual connectivity in comics. In N. Cohn (Ed.), *The visual narrative reader* (pp. 115–129). London: Bloomsbury.
- Sauvé, G., Harmand, M., Vanni, L., & Brodeur, M. B. (2017). The probability of object–scene co-occurrence influences object identification processes. *Experimental Brain Research*, *235*(7), 2167–2179. <https://doi.org/10.1007/s00221-017-4955-y>.
- 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.

- Sontag, J. M., & Barnes, S. R. (2017). The visual framing of graphics when used in preventative health digital news packages: Exploring the use of a narrative structure as the message infrastructure. *Journal of Visual Communication in Medicine*, 40(3), 109–119. <https://doi.org/10.1080/17453054.2017.1366824>.
- Stainbrook, E. J. (2016). A little cohesion between friends; Or, we're just exploring our textuality: Reconciling cohesion in written language and visual language. In N. Cohn (Ed.), *The visual narrative reader* (pp. 129–154). London: Bloomsbury.
- Swets, B., & Kurby, C. A. (2016). Eye movements reveal the influence of event structure on reading behavior. *Cognitive Science*, 40(2), 466–480. <https://doi.org/doi:10.1111/cogs.12240>.
- Szewczyk, J. M., & Schriefers, H. (2018). The N400 as an index of lexical preactivation and its implications for prediction in language comprehension. *Language, Cognition and Neuroscience*, 33(6), 665–686. <https://doi.org/10.1080/23273798.2017.1401101>.
- Tasić, M., & Stamenković, D. (2015). The interplay of words and images in expressing multimodal metaphors in comics. *Procedia - Social and Behavioral Sciences*, 212, 117–122. <https://doi.org/10.1016/j.sbspro.2015.11.308>.
- Thornhill, D. E., & Van Petten, C. (2012). Lexical versus conceptual anticipation during sentence processing: Frontal positivity and N400 ERP components. *International Journal of Psychophysiology*, 83(3), 382–392. <https://doi.org/10.1016/j.ijpsycho.2011.12.007>.
- Trabasso, T., & Nickels, M. (1992). The development of goal plans of action in the narration of a picture story. *Discourse Processes*, 15, 249–275.
- Van Petten, C., & Kutas, M. (1991). Influences of semantic and syntactic context on open- and closed-class words. *Memory and Cognition*, 19, 95–112.
- Van Petten, C., & Luka, B. J. (2012). Prediction during language comprehension: Benefits, costs, and ERP components. *International Journal of Psychophysiology*, 83(2), 176–190. <https://doi.org/10.1016/j.ijpsycho.2011.09.015>.
- Versluys, C. (2017). *Why this now? A genre analytic approach to aphasic/non-aphasic interactive events*. Utrecht, the Netherlands: LOT.
- Viggiano, M. P., & Kutas, M. (1998). The covert interplay between perception and memory: Event-related potential evidence. *Electroencephalography and Clinical Neurophysiology/Evoked Potentials Section*, 108(5), 435–439. [https://doi.org/10.1016/S0168-5597\(98\)00004-5](https://doi.org/10.1016/S0168-5597(98)00004-5).
- Võ, M. L.-H., & Wolfe, J. M. (2013). Differential electrophysiological signatures of semantic and syntactic scene processing. *Psychological Science*, 24(9), 1816–1823. <https://doi.org/10.1177/0956797613476955>.
- Weissman, B., & Tanner, D. (2018). A strong wink between verbal and emoji-based irony: How the brain processes ironic emojis during language comprehension. *PLoS ONE*, 13(8), e0201727.
- West, W. C., & Holcomb, P. (2002). Event-related potentials during discourse-level semantic integration of complex pictures. *Cognitive Brain Research*, 13, 363–375.
- Wilson, B. (2016). What happened and what happened next: Kids' visual narratives across cultures. In N. Cohn (Ed.), *The visual narrative reader* (pp. 185–227). London: Bloomsbury.
- Wittenberg, E., Paczynski, M., Wiese, H., Jackendoff, R., & Kuperberg, G. (2014). The difference between “giving a rose” and “giving a kiss”: Sustained neural activity to the light verb construction. *Journal of Memory and Language*, 73, 31–42. <https://doi.org/10.1016/j.jml.2014.02.002>.
- Wlotko, E. W., & Federmeier, K. D. (2012). So that's what you meant! Event-related potentials reveal multiple aspects of context use during construction of message-level meaning. *NeuroImage*, 62(1), 356–366. <https://doi.org/10.1016/j.neuroimage.2012.04.054>.
- Wu, Y. C., & Coulson, S. (2005). Meaningful gestures: Electrophysiological indices of iconic gesture comprehension. *Psychophysiology*, 42(6), 654–667. <https://doi.org/10.1111/j.1469-8986.2005.00356.x>.
- Yano, M. (2018). Predictive processing of syntactic information: Evidence from event-related brain potentials. *Language, Cognition and Neuroscience*, 33(8), 1017–1031. <https://doi.org/10.1080/23273798.2018.1444185>
- Yarhouse, R. B. (2017). Examining film engagement through the visual language of comics. *International Journal of Comic Art*, 19(1), 384–404.

- Zacks, J. M., Kurby, C. A., Eisenberg, M. L., & Haroutunian, N. (2011). Prediction error associated with the perceptual segmentation of naturalistic events. *Journal of Cognitive Neuroscience*, 23(12), 4057–4066. [https://doi.org/10.1162/jocn\\_a\\_00078](https://doi.org/10.1162/jocn_a_00078).
- 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.
- Zwaan, R. A., & Radvansky, G. A. (1998). Situation models in language comprehension and memory. *Psychological Bulletin*, 123(2), 162–185.