

Visual narratives and the mind: Comprehension, cognition, and learning

Neil Cohn

Tilburg University
Department of Communication and Cognition
P.O. Box 90153, 5000 LE Tilburg
The Netherlands
Email: neilcohn@visuallanguagelab.com

This is a pre-proof version. Please consult the final version for referencing:

Cohn, Neil. 2019. Visual narratives and the mind: Comprehension, cognition, and learning. In Federmeier, Kara D. and Diane M. Beck (Eds). *Psychology of Learning and Motivation: Knowledge and Vision*. Vol. 70. (pp. 97-128). London: Academic Press

Abstract

The way we understand a narrative sequence of images may seem effortless, given the prevalence of comics and picture stories across contemporary society. Yet, visual narrative comprehension involves greater complexity than is often acknowledged, as suggested by an emerging field of psychological research. This work has contributed to a growing understanding of how visual narratives are processed, how such mechanisms overlap with those of other expressive modalities like language, and how such comprehension involves a developmental trajectory that requires exposure to visual narrative systems. Altogether, such work reinforces visual narratives as a basic human expressive capacity carrying great potential for exploring fundamental questions about the mind.

Keywords: visual language; visual narrative; narrative; scene perception; event cognition; ERPs

1 Introduction

Visual narratives are so prevalent in modern society that they may seem effortless to understand. Narrative sequential images in picture books are often the first exposure that children get to textual storytelling, and comics provide a source of entertainment, and potentially education, throughout the lifespan. Visual narratives also appear in many non-entertainment contexts, be they storyboards, instruction manuals, or stimuli in a range of psychological experiments. This widespread use underscores a general belief that visual narratives are transparent to understand (McCloud, 1993; Szawerna, 2017), requiring little learning beyond basic cognition like perceptual and event processing (Loschky, Magliano, Larson, & Smith, Under review), sequential reasoning (Zampini et al., 2017), and theory of mind (Baron-Cohen, Leslie, & Frith, 1986).

Nevertheless, emerging research on visual narrative comprehension suggests that this perception of effortless understanding masks more complex structure and processes. Consider Figure 1, a sequence from Ben Costa's *Pang: The Wandering Shaolin Monk, Volume 2*. In the first panel, a monk held in bondage, Pang, encounters a tiger. Pang then offers to sacrifice himself instead of a character not depicted here (panel 2), while the tiger looks at him in preparation to attack (panel 3). Pang then closes his eyes (panel 4), preparing to be eaten (panel 5), only to look up (panel 6) and see the tiger walking away (panel 7) to which he gasps in relief (panel 8).

This sequence has numerous levels of complexity. First, we must know that this depiction is not just a whole single image, but constitutes numerous sub-images (panels) which must be read in a particular order. Within these panels, the images convey basic meaningful information, like a person, a tiger, and a forest. We must also make the connection that the characters are the same across panels, not totally different characters and scenes in each frame. In addition, not all characters are shown at all times, meaning that we have to infer what is going on "off-panel" given what we see. We also make expectations about the events and actions that are taking place, such as a possibility that the tiger might eat Pang at the start of the sequence, and thus leading to a relief when it goes away at the end. In this, we understand that the characters have desires and goals, such as Pang to not be eaten, and the tiger to have (or not have) a meal. The sequence is also conveyed to us in a particular way—certain information is depicted or withheld to evoke a particular response—thereby suggesting a particular narrative structure. Comprehending this sequencing also extends beyond the linear relations between images: panels 3 and 7 of the tiger are separated by three panels of only Pang—which all could be deleted—meaning they require a distance connection. All of this complexity (and more) underlies what may seem like a fairly simple comprehension process.

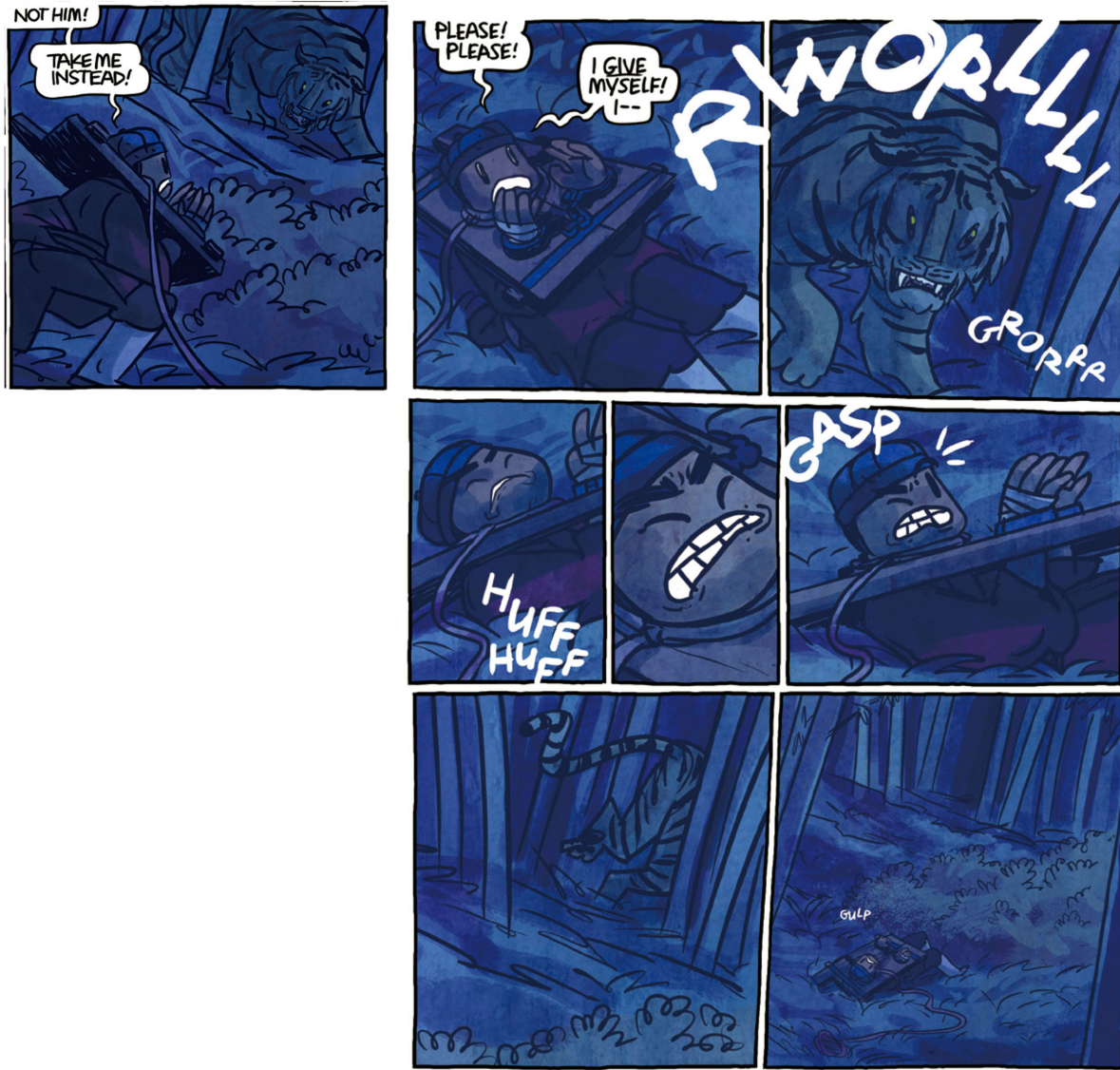


Figure 1. Example sequence from *Pang: The Wandering Shaolin Monk, Volume 2*, by Ben Costa, where a captured monk, Pang, is confronted by a tiger. The first panel comes from a separate page, for which the layout has been altered for clarity. Sequence © Ben Costa, used with permission.

This complexity has been recognized with increasing interest by researchers in psychology. Growing experimental work has led to models of sequential image comprehension which suggest sophisticated interactions between domain-general and domain-specific processing (Cohn, Under review; Loschky et al., Under review). Meanwhile, a scattered literature across several decades has also provided a wealth of data about the development of sequential image understanding, but this work has largely not been integrated into a coherent framework, nor has it informed research on the cognition of visual narratives. This paper attempts to integrate these threads of research with an eye towards future research by addressing three primary questions: First, how do we comprehend drawn visual narratives? Second, are the cognitive mechanisms used to process visual narratives specialized or do they overlap with other expressive modalities? Finally, how might people learn to understand visual narratives?

2 Comprehending visual narratives

Visual narratives of sequential images have often been characterized as simple to understand, relying on fairly uniform processes of comprehension (Bateman & Wildfeuer, 2014; McCloud, 1993). Yet, growing empirical research has implicated a complex interaction between cognitive mechanisms at various levels of structure. Here, I combine insights from two recent models of visual narrative processing: The *Scene Perception and Event Comprehension Theory (SPECT)* (Loschky, Hutson, Smith, Smith, & Magliano, 2018; Loschky et al., Under review) emphasizes how perceptual processing combines with mental model construction throughout visual narrative comprehension. The *Parallel Interfacing Narrative-Semantics (PINS) Model* (Cohn, Under review) meanwhile emphasizes neurocognition, with visual narratives involving the interface of two levels of representation: semantic information that provides the meaning, and narrative structure which organizes it into a sequence.

2.1 Semantic processing

With regard to the processing of meaning across sequential images, SPECT differentiates two general domains of processing (Loschky et al., 2018; Loschky et al., Under review). **Front-end processes** are the perceptual and attentional processes used to extract information from visual images, primarily within a single eye fixation. Such processes are largely perceptually driven. Following this, **back-end processes** are the stages involved in constructing a **situation model** of the scene, i.e., the mental model incorporating the understanding of the entities and events of an unfolding (visual) discourse (Zwaan & Radvansky, 1998). In general, front-end processes precede back-end processes, though there may be feedback in the other direction, such as when demands of a mental model affect eye fixations.

2.1.1 Front-end processing: Information extraction

Front-end processes include the **attentional selection** involved in determining which information to process and **information extraction** to pull out the relevant meaning. An eye tracking study looking at over a million fixations to naturalistic, unmanipulated comics suggests that readers fixate characters and their parts faster and longer than backgrounds (Laubrock, Hohenstein, & Kümmerer, 2018). Overall though, fixation durations were faster to panels than what is typically observed in scene perception. This may occur because of the reduced complexity of panels, which are drawn with communicative intent, compared to photographs used in typical scene perception experiments. The conventionalization of graphic schemas in panels may also play a role. Indeed, machine learning techniques trained on photographs do significantly worse when assessing the content of comic panels than on natural percepts (Khetarpal & Jain, 2016; Takayama, Johan, & Nishita, 2012). Thus, while mechanisms from scene perception operate on visual narratives, they differ from those used with natural percepts.

The faster viewing of panels may also be caused by the fact that the narrative sequencing informs comprehenders about where to direct their attention. In contrast to the idea that visual images are fairly unconstrained in how content directs attention, images in a sequence receive focused attention to particular areas of interest. These areas contain cues that are relevant for what the sequence is about (Hutson, Magliano, & Loschky, 2018), and attentional selection is more focal to such areas when images are in a coherent order than in a scrambled sequence (Foulsham, Wybrow, & Cohn, 2016).

Figure 2 depicts a schematized unfolding of semantic processing in Figure 1 (here turned greyscale), where the front-end processing of information extraction is represented by the second

tier. As suggested by eye-tracking studies, extracted information from the full images may be fairly constrained. In the second panel, the primary information might be the text (here excluded), Pang’s worried face and his bound hands, not the whole image and its background. Recent work has suggested that such focal cues provide enough information for comprehending a visual narrative sequence (Foulsham & Cohn, In prep). We created panels based on the areas of the top 10% of fixations from a prior eye-tracking experiment (Foulsham et al., 2016), and then compared the processing of sequences with these “fixation-zoom panels” and full-scene panels. We found no differences between the self-paced viewing times of full-scene and zoomed panels, suggesting that such focal information was sufficient for the sequential meaning.

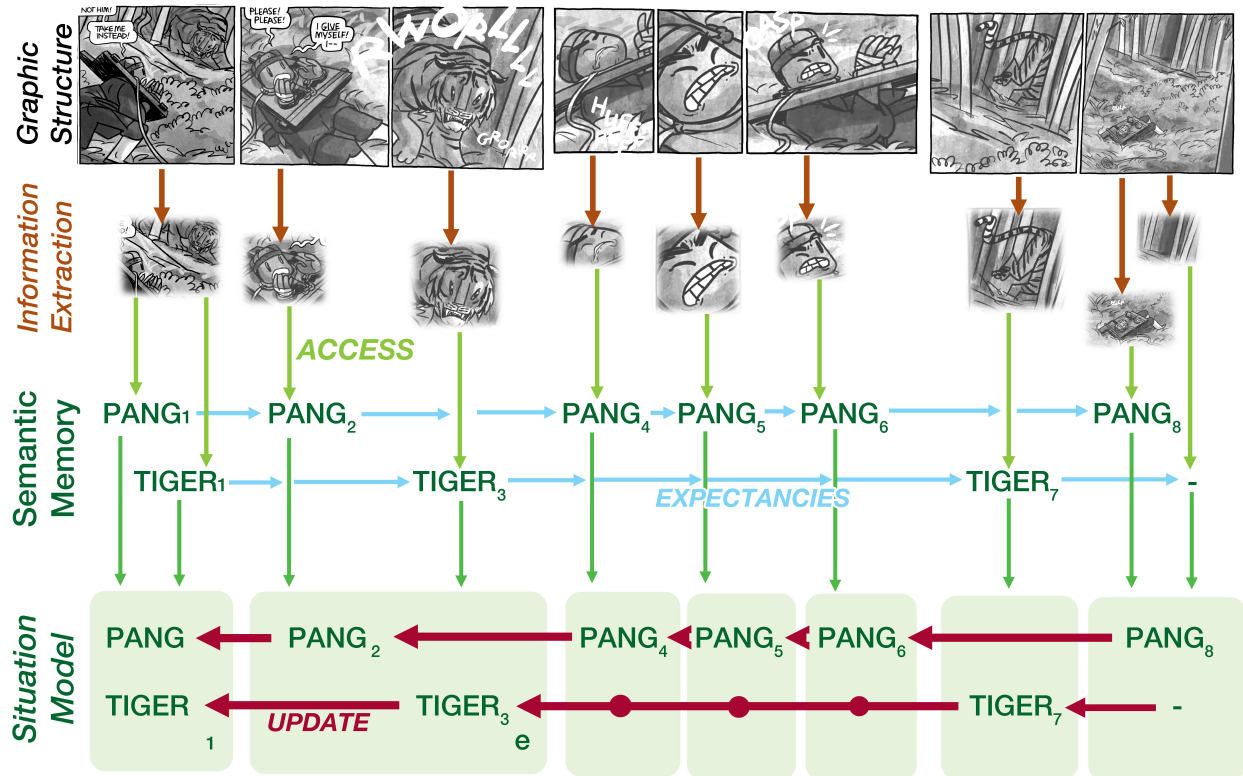


Figure 2. Aspects of semantic processing characterized for a visual narrative sequence. Forward-looking expectancies are noted in blue within semantic memory, while backward-looking updating is notated in red within the situation model.

2.1.2 Back-end processing: Semantic access

In SPECT, the front-end processes of visual search and attentional selection guide information extraction that feeds into “back end” processes that create an understanding of the sequence. The PINS model emphasizes that such processes begin with mapping this information into semantic memory. This process of *semantic access* has been indexed in neurocognition by the “N400”—a neural response recorded using event-related potentials (ERPs), a measurement of the electrical activity of the brain, in this case while participants read visual narratives. Though the N400 was first discovered in language processing (Kutas & Hillyard, 1980), it has been observed across many domains (Kutas & Federmeier, 2011), including to images in visual narrative processing (Cohn, Paczynski, Jackendoff, Holcomb, & Kuperberg, 2012; West & Holcomb, 2002).

The N400 is a default neural response that occurs to all meaningful information, modulated by the expectancy of an incoming stimulus (like an image or word) given its prior context (Kutas & Federmeier, 2011). Unexpected or incongruous information thus elicits larger N400s than more expected information. When expectancies are continually affirmed, attenuation of the N400 occurs. For example, N400 amplitudes decrease across the ordinal position of words in a sentence (Van Petten & Kutas, 1991) or images in a coherent visual narrative sequence (Cohn et al., 2012).

Though N400 effects appear across modalities (Kutas & Federmeier, 2011), in images they are often preceded by another negative deflection peaking near 300ms, the “N300” (McPherson & Holcomb, 1999). This negativity has often been taken to index the semantic identification or categorization of visual objects (Hamm, Johnson, & Kirk, 2002), potentially in line with a stage of visual information extraction prior to semantic access. Scene perception research has largely taken the N300 to be inseparable from the N400, based on comparing typical percepts with incongruous or unexpected elements in a visual scene (Draschkow, Heikel, Vö, Fiebach, & Sassenhagen, 2018; Lauer, Cornelissen, Draschkow, Willenbockel, & Vö, 2018). However, we recently used ERPs to compare our fixation-based panels to full-scene panels, as described above (Cohn & Foulsham, In prep). Fixation-zoomed panels elicited a significantly smaller N300 than full-scene panels, but differed minimally in N400 amplitudes. This implied that zoomed-in content based on eye-fixations required less information identification than full scenes, since the relevant content was already extracted and framed directly (N300), but they provided similar access to semantic memory for the sequence (N400).

Within Figure 2, green lines depict how the extracted information feeds into semantic memory. Figure 2 focuses only on referential information for simplicity, but full comprehension would incorporate information about events, goals, spatial location, and other situational dimensions (Zwaan & Radvansky, 1998). Within semantic memory, this information activates features for the main characters directly (Pang, Tiger) and their associations (Chinese monks, big cats, etc.). These activations within semantic memory then set a precedence for (re)activation by subsequent frames. For example, referential information (Pang, the Tiger) or spatial information (a bamboo forest) present in one panel should be easier to reactivate if they appear in the next panel, thus leading to an attenuated N400 across the coherent narrative sequence (Cohn et al., 2012).

2.1.3 Back-end processing: Situation model construction

Information from semantic memory is then incorporated into a growing situation model of the unfolding meaning of the visual discourse (Cohn, Under review; Loschky et al., 2018; Loschky et al., Under review). A situation model is first held in working memory while being constructed throughout engagement with a (visual) narrative, before becoming stored in episodic long term memory as the story is retained into the future (Magliano, Kopp, McNerney, Radvansky, & Zacks, 2012). As a reader progresses through a visual narrative sequence, they generate various expectancies for the semantic information that may occur in a subsequent panel, given the current image. At the most basic level, a *continuity constraint* maintains that the characters in one panel may be the same as those in prior frames (Bornens, 1990). Thus, visual features that carry across panels may reference the same entity in a situation model, even given changes in a character’s (or object’s) posture, or the framing or viewpoint taken on them. Without such a continuity constraint, readers would view each frame as if it depicted different entities. With recognition of these cues, an *activity constraint* then allows an inference that a change of state has occurred between panels, such as an unfolding event, or a change in time or viewpoint. Without recognition of the activity

constraint, each panel would be considered as its own isolated scene, unconnected to the other images in a sequence.

As a reader progresses through the sequence, they monitor for changes in characters, postures, events, spatial locations and other situational dimensions across panels. Discontinuities of semantic information extracted from panels triggers an *updating* process whereby the situation model must be revised given the new information. SPECT here draws a distinction between *mapping* and *shifting* (Loschky et al., 2018; Loschky et al., Under review). Backward-looking “mapping” occurs from detecting incoherence between incoming information and the current situation model, a disconfirmation of previously made expectancies, and/or exceeding a threshold of change between dimensions. Such principles echo theories of coherence across image sequences (Bateman & Wildfeuer, 2014; Saraceni, 2016; Stainbrook, 2016), such as more inference being required when panels retain fewer entities or less semantic associations across panels (Saraceni, 2016). Such updating is posited as ongoing across each unit of a discourse, but when such change becomes untenable for mapping, the continuous activity is segmented and “shifts” to a new situation model (Loschky et al., 2018; Loschky et al., Under review).

In Figure 2, the situation model is represented with green boxes, with the updating process represented with red lines marking the change in the situation model relative to the preceding image. For example, panel 2 shows Pang, but panel 3 shows the tiger. This change in character incurs a cost (Cohn & Kutas, 2017). Further though, because each panel highlights only a portion of the broader scene of that moment, the broader spatial environment would need to be inferred, as notated by “e” (for “environment”). Updating may also be required for maintaining the tiger in working memory across panels 4-6, which focus on Pang, while the tiger remains off-panel until panel 7. Panel 7 may thus incur an updating cost both for reactivation of the tiger, and because the tiger walks away, in contrast to the foreshadowed expectation that it may be attacking Pang off-panel in panels 4-6.

One consequence of such updating processes may be the need to derive an inference when the objects and/or events in the incoming panel differ from the established situation model. Imagine if Figure 1 omitted the penultimate panel. We would thus need to reconcile the incoming final panel (Pang alone in the woods) with the prior information (a tiger was going to attack him) to derive the intended meaning with the absence of that event (the tiger left). Such backward-looking inferencing would incur working memory demands in order to update the situation model with the required information (Cohn & Kutas, 2015; Cohn & Wittenberg, 2015; Hutson et al., 2018; Magliano, Larson, Higgs, & Loschky, 2015).

Backward-looking updating processes have been posited in ERPs to be indexed by a P600—a positive deflection that appears over the posterior part of the scalp, peaking around 600ms after the appearance of an image. In language research, P600s have been associated with the update, revision, and/or integration of incoming information given the expectancies of a prior context (Brouwer, Crocker, Venhuizen, & Hoeks, 2016; Kuperberg, 2013), and may be associated with a broader family of brain responses associated with mental model updating (Donchin & Coles, 1988; Van Petten & Luka, 2012). In visual narratives, P600s have been observed for changes in the characters or events of an ongoing scene, for both congruous (Cohn & Kutas, 2015) and incongruous contexts (Cohn & Kutas, 2017; Cohn & Maher, 2015), and for mapping of inferential processing (Cohn & Kutas, 2015). Updating for inference may also be seen behaviorally in the form of longer viewing times to panels following the omission of key event information (Cohn & Wittenberg, 2015; Magliano, Kopp, Higgs, & Rapp, 2016; Magliano et al., 2015), or via visual search processes seeking to reconcile the missing information (Hutson et al., 2018).

Inference has also been associated with sustained ERP negativities, often with a central or frontal distribution across the scalp. In language research, such negativities have been associated with working memory demands for searching through or holding onto information in a situation model to resolve inferences related to referential (Hoeks & Brouwer, 2014; van Berkum, 2009) or event-based ambiguities (Baggio, van Lambalgen, & Hagoort, 2008; Bott, 2010; Paczynski, Jackendoff, & Kuperberg, 2014; Wittenberg, Paczynski, Wiese, Jackendoff, & Kuperberg, 2014). Similar sustained negativities have also been observed in visual narratives requiring the inference of a missing event (Cohn & Kutas, 2015). Behavioral research has also shown slower self-paced viewing times to panels following the position of an omitted event, which can be modulated by intervening working memory demands (Magliano et al., 2015).

Thus, processing the meaning of visual narratives breaks down into front-end processes for negotiating information in the visual modality, and back-end processes for constructing a situation model. Front-end processes use attentional selection and information extraction to feed into back-end processes activating semantic memory to then construct a progressively updating situation model. Such processes involve both forward-looking expectancies on the basis of activated information, and backward-looking updating to reconcile those expectations with incoming information. Taken together, these processes are consistent with established theories for the processing of discourse in the verbal domain (Graesser, Millis, & Zwaan, 1997; McNamara & Magliano, 2009; van Dijk & Kintsch, 1983), yet adapted to the unique affordances of the visual-graphic modality (discussed below).

2.2 Narrative processing

While comprehension of a visual narrative sequence aims at constructing the type of understanding characterized above, such semantic processing alone remains insufficient. As described, authors make choices for which information is shown, when it is shown, and how that information coheres in sequencing. For example, in Figure 1, panels 4-6 only show Pang, while the tiger's actions remain out of view. Because these panels separate the two panels of the tiger (panels 3 and 7), we must connect the tiger panels across a distance. Also, this Pang-only sequence comes after the clear set-up of the situation (with only the first panel showing both Pang and the tiger), and is followed by a "climax" where the tiger goes away. All of these features reflect choices made by the author, and constitute an explicit *narrative structure* of the sequence.

Thus, within the PINS Model (Cohn, Under review), a narrative level of representation runs parallel to semantic processing. This narrative processing is depicted in Figure 3, where the horizontal labels and colors (marking processes of access, prediction, and updating) interface with the vertical levels of Figure 2. The characteristics of this level of representation have been outlined by the theory of Visual Narrative Grammar (VNG), which argues that a *narrative grammar* organizes the meaningful information of a visual narrative sequence analogous to how syntactic structure organizes the semantic information of a sentence (Cohn, 2013b).

VNG is similar to previous approaches comparing narrative to syntax, such as story grammars of discourse (Mandler & Johnson, 1977; Rumelhart, 1975). However, these older models used procedural phrase structure rules based on Chomskyan generative grammars (Chomsky, 1965), and had ambiguous distinctions between narrative and semantics. In contrast, VNG is modeled after construction grammars (Culicover & Jackendoff, 2005; Goldberg, 1995), which posit that sequencing schemas are entrenched in long-term memory with mappings to an explicitly separate semantics (see Cohn, 2013b, 2015).

Table 1. Basic sequencing patterns used in Visual Narrative Grammar.

a) **Canonical narrative schema:** [Phase X (Establisher) – (Initial) – (Prolongation) – Peak – (Release)]

b) **Conjunction schema** [Phase X X₁ - X₂ - ... X_n]

The most basic sequencing pattern in VNG is the canonical narrative schema (see Table 1a). This pattern specifies that a narrative progresses through various states, which differentiate parts in the sequence. A sequence may begin with a panel that functions as an Establisher, which sets up the actors and events of the situation, such as Pang meeting the tiger in panel 1 of Figure 1. The anticipation or start of events occurs in an Initial, such as Pang and the tiger in panels 2 and 3. This anticipation may be delayed across a Prolongation, which in Figure 1 occurs in panels 4-6 as it is unknown whether Pang will be eaten. The climax of a sequence occurs in a Peak, here with the tiger leaving in panel 7. The dissolution of that narrative tension occurs in the Release, reflected in Pang’s relief in the final panel. The canonical schema specifies these narrative categories in this particular order, though not all categories are mandatory.

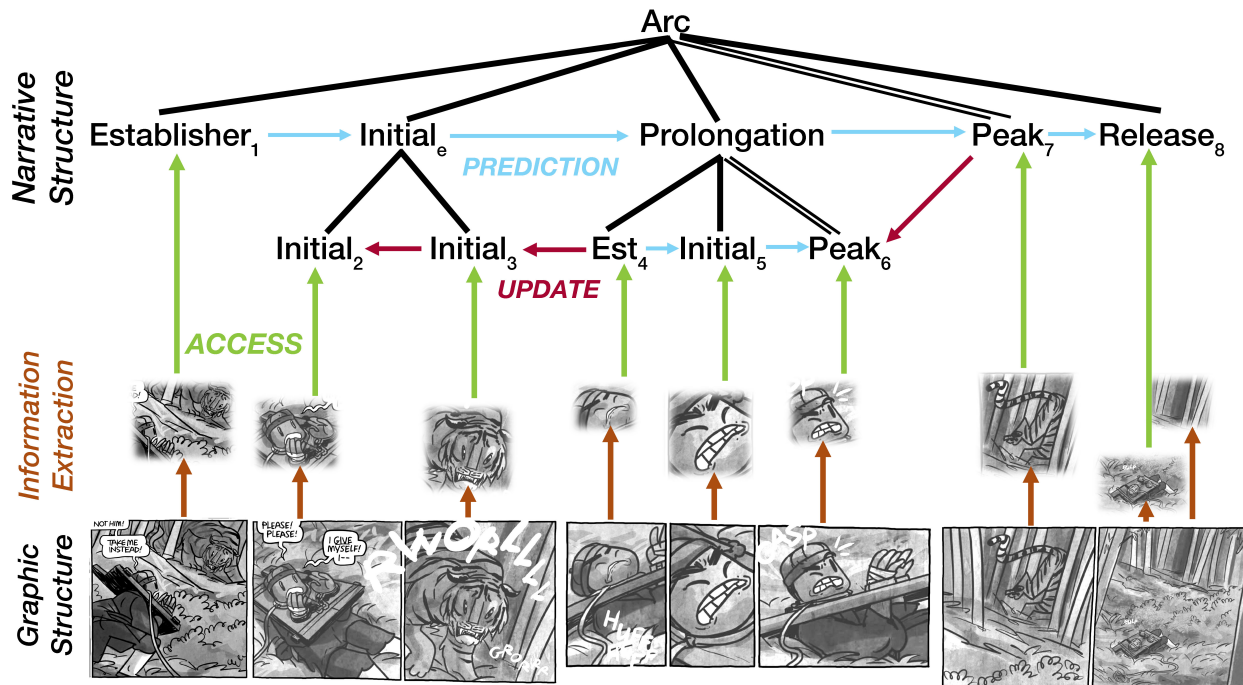


Figure 3. The narrative structure of a visual narrative sequence illustrated by a hierarchic tree structure. Green lines represent the access of narrative categories via extracted visual cues. Blue lines represent forward-looking structural predictions, while red lines represent backward-looking structural revision would be necessary.

Complexity can be added to a sequence because narrative categories recursively can apply both to individual panels and to groupings of panels. Figure 3 depicts the narrative structure of Figure 1. At the top level, panels follow the canonical narrative schema. However, panels 4-6 constitute their own narrative constituent, with an Establisher of Pang looking worried (about being eaten), an Initial of his anticipation of the attack, and then a mini-climax of opening his eyes to see what is happening. This sequence is essentially a *center-embedded clause*: it could stand alone as a sequence, and indeed could also be omitted without disrupting the congruence of the

broader sequence. However, in separating the Initials (below) of the potential for the tiger to attack (panel 3), and its subsequent departure (panel 7), this embedded constituent delays the climax of the sequence, and thereby, as a whole, functions as a Prolongation. This grouping is motivated internally by panel 6, the Peak, which acts as the “head” of the sequence. That is, this panel carries the primary message of this grouping relevant for the higher-level sequence. At the top, an “Arc” reflects a constituent that plays no other role in a sequence.

Modifiers further elaborate a visual sequence, such as constructions that switch between different characters within a narrative state, or zooms of information in another panel (Cohn, 2015). In Figure 3, panels 2 and 3 each show a portion of the overall spatial environment—Pang and the Tiger—but it does not show them together in the same frame. Yet, both panels serve as narrative Initials—joined together through a Conjunction schema (Table 1b) which unites two categories of the same type into a common constituent, just as conjunction operates in syntax of sentences (ex. *salt and pepper* unites two nouns in a noun phrase). In this case, the two Initials correspond to different characters, which demands the inference of a shared environment. This inference in the situation model then corresponds to the Initial constituent, as notated with the subscript “e” (which corresponds to the “e” in the situation model in Figure 2).

Constituent structures and conjunction as *narrative patterns* illustrate the necessity for a separate level of representation from a visual narrative’s *semantics*, because they characterize choices for which information is shown and when. Using conjunction, panels 2 and 3 separately bring focus to each character, although a single image containing both characters would also be possible. Similarly, panels 4 through 6 show Pang without framing the tiger, and indeed could also be collapsed into a single panel, but extending it on Pang alone raises the *narrative* tension. These structural choices relate to how meaning is presented (and subsequently processed) amidst multiple options, thus requiring a separate structure from the semantics itself.

2.2.1 *Narrative categories*

Because this narrative grammar packages how information is conveyed to a reader, it constitutes a level of representation parallel to the semantics, which a comprehensive model of comprehending visual narratives cannot ignore. First, narrative categories are *accessed* via semantic cues in panels, extracted as the most relevant parts of an image for a sequence. For example, the prototypical semantic cues of an Establisher are passive actions that introduce entities, as in panel 1 of Figure 3. Initials are typically cued by preparatory actions, such as the tiger approaching Pang in panels 2 and 3. Peaks are often completed actions, though Figure 3 departs from this prototypicality, depicting the tiger *leaving* instead of attacking, here climaxing by *denying* such event-based expectations. Finally, Releases often depict codas of actions, such as the aftermath of the tiger leaving. These semantic cues provide the bottom-up assignment of a narrative category to a panel.

Because of the canonical narrative schema, top-down information can also guide narrative category assignment. For example, when participants arrange unordered panels into a sequence, the same image content can function as Establishers and Releases, and self-paced viewing times do not differ for panels switched between these categories (Cohn, 2014). Because the same content can play differing roles in a sequence, it implies that top-down position can also guide narrative categories beyond bottom-up cues (and/or that different cues can function in different positions).

2.2.2 *Narrative constituents and prediction*

Such top-down assignment is facilitated because the canonical narrative schema allows for *structural predictions*. Thus, if a panel is assigned to the category of an Initial, there is a structural prediction that a Peak will follow, given their order in the canonical narrative schema. If the incoming panel satisfies the bottom-up constraints, or allows for such assignment, the top-down schema can thereby determine the narrative category. In ERPs, violations of those forward-looking predictions results in an anterior negativity, which indexes the disconfirmation of structural predictions (Kluender & Kutas, 1993; Yano, 2018), or increased cost of combinatorial structure building (Hagoort, 2017). In visual narratives, such anterior negativities have been observed to comparisons between semantically incongruous sequences that varied in the presence of narrative structure (Cohn et al., 2012), to violations of narrative constituent structure (Cohn, Jackendoff, Holcomb, & Kuperberg, 2014), and to narrative conjunction compared to non-conjunction sequences (Cohn & Kutas, 2017).

Theories of discourse processing have often emphasized that narrative segments are triggered in response to semantic discontinuity, due to a revision of the constructed situation model (Gernsbacher, 1990; Loschky et al., Under review; Magliano & Zacks, 2011). Under such a view, narrative constituents would emerge only out of a backward-looking updating process, from the recognition that a narrative break had already been passed. While semantic discontinuities may indeed correlate with narrative constituent boundaries (Cohn & Bender, 2017), narrative processing operates independently from semantic processing. First, non-canonical narrative bigrams (like *Peak-Peak* in Figure 3) are more predictive of narrative segmentation than semantic discontinuity (Cohn & Bender, 2017). Second, comparisons of sequences with the presence or absence of narrative structure and semantic associations indicated that N400s—which index semantic processing—are not sensitive to narrative structure (Cohn et al., 2012). Third, the inverse relationship is indicated by anterior negativities, which are sensitive to narrative patterning, but not situational discontinuity (Cohn & Kutas, 2017). Thus, selective ERP effects distinguish between narrative structure (anterior negativities) and semantic processing (N400).

In addition, left anterior negativities have been observed when backward-looking updating would have been impossible. In this study, participants viewed panels one at a time while, inserted into the sequence, blank white “disruption” panels were presented either at the natural break between narrative constituents, or disrupting the segmented constituents themselves (Cohn et al., 2014). A larger left anterior negativity was evoked by the disruption panels within narrative constituents compared to those at the natural break between constituents. Importantly, this negativity thus appeared to disruptions that occurred *prior* to a participant reaching a panel following the break between constituents. It thus could not have been triggered by reanalyzing the sequence on the basis of looking backward at a discontinuity, because no panel had yet been reached to signal such a contrast. Rather, it must have been the disconfirmation of a forward-looking prediction.

2.2.3 *Structural revision*

In other cases, *structural revision* may be warranted if the incoming structure does not conform to the predictions. For example, if the predictions of a particular narrative category differ from the cues of an incoming panel, a reanalysis may occur. Such revision is again indexed in ERPs by a P600, which was indeed originally observed in language in contexts of syntactic revision (Hagoort, Brown, & Groothusen, 1993; Osterhout & Holcomb, 1992) before its interpretation was extended to more general updating or integration. In visual narrative sequences, P600s appear to an unexpected change in narrative categories (Cohn, 2012), even when inferential

demands are held constant (Cohn & Kutas, 2015). That is, if the reader is at a Peak and a subsequent Release is thus predicted, incoming information prototypical of an Initial will trigger a reanalysis of the panel. Such a revision may also evoke a reanalysis of the sequencing, whereby a constituent boundary is recognized when given an illegal Peak-Initial bigram (Cohn & Bender, 2017). Such structural revision is suggested by P600s to reanalysis of the constituent structure of a sequence (Cohn et al., 2014), and to unexpected sequencing patterns (Cohn & Kutas, 2017).

Thus, in sum, visual narrative comprehension involves at least two levels of representation for semantic and narrative structures. Information extracted from surface-level visual cues triggers the accesses of semantic memory, which feeds into a situation model, which, in turn, is updated based on the congruity of that incoming information with the expectancies generated at the prior panel. Parallel to this, a narrative structure packages this meaning sequentially. Here, extracted information cues narrative categories, which belong to a schema that facilitates structural predictions. Disconfirmed structural expectations evoke a revision process to update the overall narrative structure. Together, these parallel levels of representation interface across the comprehension of a visual narrative sequence.

3 Domain-specificity and generality in visual narrative processing

Given these mechanisms implicated by SPECT and the PINS model in the comprehension of visual narratives, we next ask: to what degree are these processes specific to visual narrative comprehension, and to what degree do they overlap with other domains, like language? In these comparisons, I have hypothesized a guiding *Principle of Equivalence*, which states that *we should expect the mind/brain to treat expressive capacities in similar ways, given modality-specific constraints* (Cohn, 2013a). That is, from the perspective of cognition, different modalities like language, music, and visual narratives should share in their processing resources. However, their differences should be motivated by the affordances of the modalities themselves, either with processing of that modality or with how that modality subsequently facilitates cognitive mechanisms.

Thus, because modalities differ in how they structure information, they should vary the most in the front-end processes: how you extract information from images will differ from how you extract information from text or speech (Loschky et al., 2018; Magliano, Higgs, & Clinton, In press; Magliano, Loschky, Clinton, & Larson, 2013). For example, inference generation in visual narratives trigger processes of visual search and attentional selection that operate differently in processing textual discourse (Hutson et al., 2018). Yet, the context of narrative sequencing may guide such processes to be more focal and directed for images in visual narratives compared to general scene perception (Foulsham et al., 2016; Hutson et al., 2018; Laubrock et al., 2018).

Back-end processes are thus posited as operating across domains more than the front-end processes. Behavioral research has shown that verbal working memory processes interact with non-verbal inferential processes for understanding missing information in visual narrative sequences (Magliano et al., 2015), and similar segmentation occurs for verbal and visual narratives (Magliano et al., 2012). Findings from ERPs are even more suggestive. Similarities between the N400s elicited in words, sentences, images, and visual sequences suggest a common mechanism for processing meaningful information across domains (Kutas & Federmeier, 2011). Indeed, semantic processing is modulated even when crossing modalities, such as substituting images for words in sentences (Federmeier & Kutas, 2001; Ganis, Kutas, & Sereno, 1996), or words for images in visual narrative sequences (Manfredi, Cohn, & Kutas, 2017). In addition, N400 effects

are attenuated for individuals with autism in both verbal and visual narratives, suggesting a shared deficit independent of modalities (Coderre et al., 2018).

Additional similarities persist in the brainwaves associated with structural aspects of processing, such as (lateralized) anterior negativities or P600s. Both of these waveforms have been associated with grammatical processing in language (Hagoort, 2017; Kuperberg, 2007) and in music (Koelsch, Gunter, Wittfoth, & Sammler, 2005; Patel, 2003), and have also appeared in manipulations of visual narrative grammar (Cohn et al., 2014; Cohn & Kutas, 2017). Such similarities have motivated proposals for shared processing mechanisms between combinatorial systems (Patel, 2003), despite differences in their representational systems. That is, the brain may draw on similar mechanisms for assigning units to categorical roles and/or organizing them into hierarchic sequences across domains. However, it does so across different representations for modalities: the syntax of sentences uses different constructs than the narrative structure of visual sequences, and these both differ from the fairly asemantic sequencing of music.

Narrative structure itself may also negotiate between domain-specific and domain-general processes. Narrative grammar is posited as a domain-general organizational system for a particular level of information structure (i.e., the sequencing of events). In this capacity, Visual Narrative Grammar has been applied to film (Amini, Riche, Lee, Hurter, & Irani, 2015; Cohn, 2016a; Yarhouse, 2017), motion graphics (Barnes, 2017), discourse (Fallon & Baker, 2016), and computational generation of narrative (Andrews & Baber, 2014; Kim & Monroy-Hernandez, 2015; Martens & Cardona-Rivera, 2016). If narrative structure operates in a domain-general way, it must adapt to the affordances of different modalities (Cohn, 2013b, 2016a). For example, though drawn narratives in comics and moving narratives in film both use sequential visual meaning-making, film involves temporality to its sequencing, along with dynamic movement by characters, and viewpoint of the camera. Drawn visual narratives (other than animation) do not use such movement, and by contrast must produce inferences of motion while also organizing sequences, not in time, but spatially across a layout (Cohn, 2016a).

Nevertheless, domain-general and domain-specific processes may interact in interesting ways. For example, front-end processes related to attentional selection and information extraction are posited to be mechanisms from scene perception that operate in visual narratives differently from textual narratives (Loschky et al., Under review). Meanwhile, the back-end processes of inference generation and mental model updating have been posited as domain-general, and operating similar across different modalities (Cohn & Kutas, 2015; Gernsbacher, 1990; Magliano et al., 2015; Magliano et al., 2013). Yet, inferential situation model construction in visual narratives may sponsor mechanisms unique to the visual-graphic domain, such as visual search functions instigated by the need to find information to infer what was not provided overtly in a visual narrative (Hutson et al., 2018).

Thus, visual narrative processing implicates several domain-general mechanisms involved in both the meaning-making and sequencing of visual images and other domains like language and music. However, these forms differ in how their modalities demand that information be extracted, and how these domain-general mechanisms might manifest in domain-specific ways. Using visual narratives to explore the limits and interface between such general and specific mechanisms can offer a fruitful method for investigating fundamental questions about cognition.

4 Development of visual narrative comprehension

The comprehension of visual narratives has been characterized as involving front-end processes similar to scene perception, and back-end processing balancing both semantic and

narrative levels of representation. These back-end processes appear to draw on more domain-general mechanisms, while the front-end processes tap into aspects of visual perception, but interacting with domain-specific constraints. Given this, we can ask: how might such a system be learned and develop?

If sequential image understanding relied on event and perceptual processing alone, it should “come for free” through basic cognition. This would imply no need for further fluency or development of specialized knowledge, since visual narratives could rely on event cognition, attention and perception, Theory of Mind, and other domain-general mechanisms. Such assumptions of transparency have largely motivated the use of visual narratives in psychological tasks to assess Theory of Mind (Baron-Cohen et al., 1986; Sivaratnam, Cornish, Gray, Howlin, & Rinehart, 2012), temporal cognition (Ingber & Eden, 2011; Weist, 2009), sequential reasoning (Zampini et al., 2017) and many other aspects of cognition.

In contrast to such assumptions, cross-cultural research suggests that not all neurotypical adults have a capacity to construe visual narratives as *sequential* images. For example, researchers in Nepal found that respondents did not construe the continuity constraint that characters in one image were the same as those in other images; rather, they viewed each image as its own scene with different characters (Fussell & Haaland, 1978). This inability to construe referential continuity typically occurs with participants from rural areas without literacy or exposure to Western culture (e.g., Bishop, 1977; Byram & Garforth, 1980; Cook, 1980; Gawne, 2016; San Roque et al., 2012). Such findings imply sequential image comprehension requires exposure to visual narratives, just as fluency in language requires exposure to a linguistic system. This suggests that accessing the domain-general processes involved in visual narrative comprehension is not trivial, and that domain-specific processes—and their interface with back-end processes—require some degree of learning through exposure.

So, what might this learning look like? Developmental research has implied that full sequential image understanding occurs along a trajectory, with recognition of a visual narrative *as a sequence* only occurring between 4 to 6 years old. While dedicated research programs on the development of sequential image understanding have yet to be established, the extant literature suggests incremental “stages” that may align with the sketch of sequential image understanding above, given the assumptions of exposure.

Before the age of 3, children do not seem to comprehend images as a sequence, but will recognize the referential information within images (Trabasso & Nickels, 1992; Trabasso & Stein, 1994). The style of images may modulate this understanding, with objects in a realistic style recognized easier than a cartoony style (Ganea, Pickard, & DeLoache, 2008). In producing visual narratives, 3 year old children mostly just draw an inventory of characters, and even 5-year-olds rarely produce image sequences with distinct juxtaposed panels (Silver, 2002). By the age of 4, children will recognize the event structures depicted by individual images (Trabasso & Nickels, 1992; Trabasso & Stein, 1994).

Prior to the age of 4 children do not seem to make connections across images in a sequence. Specifically, they lack a continuity constraint to construe that referential entities in one image are the same as those in subsequent frames (Bornens, 1990). This may be why children below the age of 4 do poorly at picture arrangement tasks, which give a participant several images and score them relative to how well they construct an expected sequence (Friedman, 1990; Weist, Atanassova, Wysocka, & Pawlak, 1999; Weist, Lyytinen, Wysocka, & Atanassova, 1997). Similarly, young children are fairly inaccurate at choosing a preferred panel to end a narrative sequence (Zampini et al., 2017), and their narration of picture stories typically describes the

contents of images as isolated events without connecting them across a narrative sequence (Berman & Slobin, 1994; Poulsen, Kintsch, Kintsch, & Premack, 1979; Trabasso & Nickels, 1992; Trabasso & Stein, 1994).

This ability to construe continuity and activity across an image sequence begins between the ages of 4 and 5, reaching full comprehension between 5 and 6 (Bornens, 1990). This shift to recognizing image *sequencing* also aligns with growing accuracy in picture arrangements and selection of sequence ending images (Friedman, 1990; Weist et al., 1999; Weist et al., 1997; Zampini et al., 2017), and with narrations of picture stories incorporating more descriptions of sequential events (Berman & Slobin, 1994; Karmiloff-Smith, 1985; Paris & Paris, 2003; Poulsen et al., 1979; Trabasso & Nickels, 1992; Trabasso & Stein, 1994). Children between 5 and 7 years old also shift to drawing stories using a sequence of images, given exposure to comics and picture stories (Wilson & Wilson, 1979, 1982).

Starting around age 5, children also begin inferring contents omitted from a sequence (Schmidt & Paris, 1978; Zampini et al., 2017). In tasks asking participants to infer the contents of a deleted panel, few 5-7-year-olds recognize the missing information, but grow in proficiency through age 14 (Nakazawa & Nakazawa, 1993). This aligns with improvements across these ages for understanding coherence across visual narrative sequences (Bingham, Rembold, & Yussen, 1986). Development of inferential skills continues into older ages, modulated by comic reading expertise, as college students—who read comics the most—are more proficient than older adults (Nakazawa, 2004). This same type of experience-based fluency has been observed for tasks using picture arrangement (Nakazawa, 2004) and for recall and comprehension (Nakazawa, 1997).

Table 2. Estimated developmental trajectory of visual narrative comprehension, along with corresponding aspects of processing.

Age of onset	Comprehension abilities	Cognitive mechanisms
~2-3	Single image referential understanding entities and events	Information extraction (mapping form to meaning) and semantic access
~3-5	Single image referential relations and temporal events	Information extraction (mapping form to meaning) and semantic access
~4-6	Recognition of relationships across panels; continuity and activity constraints	Mental model construction and mapping (form-meaning mappings across images)
~5-12	Recognition of inferential relationships across panels	Mental model updating (inference)

Speculatively, visual narrative comprehension develops in ways that appear to be similar to aspects of processing, as in Table 2. The recognition of referential and event information in individual images at early ages (2-4) may be characterized by the “information extraction” and “semantic access” stages of processing, as they pertain to the mapping of the graphic surface to meanings. Recognition of the continuity and activity constraints around 4-5 years old implies a growing ability to construct a coherent situation model from the disparate parts of the narrative sequence. This begins the ability to create mappings between states of situation models, and then extends to being able to revise a situation model when needing an inference to make sense of otherwise discontinuous information. It is noteworthy that this timeline is similar for development of verbal narratives (Berman & Slobin, 1994; Trabasso & Stein, 1997), again warranting the

possibility of domain-general mechanisms. However, such shared back-end comprehension does not explain the lack of recognizing referential continuity across *images* in early ages, where the modality-specific affordances still need to be negotiated.

This progression characterizes the development of construing meaning from a sequence of narrative images—essentially the semantic level of representation described above. However, as argued, a narrative level of representation also organizes such meaning. Findings on narrative structural development remains less forthcoming, as research has primarily targeted children’s meaning-making abilities for visual narratives, rather than their recognition of the conventions and structures organizing those meanings. Certainly, tasks like picture arrangements and final panel selection can assess the structure of narrative, and prior developmental findings may implicate narrative structure. However, probing such intuitions directly requires more targeted manipulations within such tasks (e.g., Cohn, 2014).

Nevertheless, data from studies of children producing visual narratives imply that fluent populations can create structured narratives as young as 7 years old. Studies in Japan, where people read manga (“comics”) prevalently throughout their lives, indicate that nearly all 6-year-olds can draw coherent visual narratives, often with complex framing changes (Wilson, 1988). With explicit instruction in comic creation, American and Canadian students between 7 and 9 created complex visual stories with clear narrative arcs often with complex framing, although they often required coaching to create coherent visual sequences (Pantaleo, 2013b; Stoermer, 2009). By age 12, some of these students’ examples show sophisticated narrative patterns, such as zoom panels, alternation between characters, point-of-view shots, and narrative “rhythm” (Pantaleo, 2012, 2013a, 2015, 2017), demonstrating proficiency in narrative structure.

It is worth noting that many benchmarks for comprehending static sequential images align with those observed in children’s development of understanding the dynamic sequencing in film. Children younger than 18 months have difficulty construing the sequencing of film, and may view them as separate isolated scenes (Noble, 1975; Pempek et al., 2010; Richards & Cronise, 2000). Sequential understanding appears to fully mature between the ages of 4 and 7 (Collins, Wellman, Keniston, & Westby, 1978; Munk et al., 2012; R. Smith, Anderson, & Fischer, 1985), and be sensitive to frequency of exposure with filmic patterns (Abelman, 1990; Barr, Zack, Garcia, & Muentener, 2008). Also, cross-cultural research has similarly implied that naïve film viewers have difficulty comprehending filmic sequences that require more modification and inferential ties between film shots (Ildirar & Schwan, 2015; Schwan & Ildirar, 2010).

While development of filmic sequencing appears comparable to static sequential images, they differ in their affordances in several ways (Cohn, 2016a; Magliano et al., 2013). First, films usually involve dynamic movement, both of the figures within a scene and of the viewpoint itself (i.e., due to camera movement). In addition, the sequencing unfurls in a single space over time, rather than several images across a spatial layout, as in drawn visual narratives. Although, it is worth noting that in production, film sequencing often begins with creating *storyboards*—static sequential images that guide the subsequent filming and editing of a film. In comprehension, dynamic movement may alleviate some of the demands of basic mapping across shots, because of the attentional advantage of maintaining continuity across shots (T. J. Smith, 2012). Similar reinforcement occurs through a concurrent audio track, which appears to aid naïve film viewers in construing film sequences (Ildirar, Levin, Schwan, & Smith, 2018). This is different than the coherence mappings required of spatial sequencing of static images, where figures must repeat in each frame and thereby require referential continuity to bind them together.

A second affordance that often differs is the nature of the percepts, in that static visual narratives typically use drawings, while film most often uses natural percepts (except with animation). Unlike natural percepts, drawings require a decoding process, particularly across sequencing that demands referential continuity despite changes in postures, viewpoint, etc. This may be why referential continuity with natural percepts remains mostly retained by naïve film viewers for filmic narratives (Ildirar & Schwan, 2015), but is more problematic in drawn form for those unfamiliar with static visual narratives (e.g., Bishop, 1977; Byram & Garforth, 1980; Cook, 1980; Gawne, 2016; San Roque et al., 2012).

In addition, development of aspects of scene perception appear substantially earlier than those implicated for picture understanding. Children discriminate basic percepts even at birth, and can perceive feature variation across objects within their first months (Johnson, 2013). By two years visual attention can be guided by scene context (Helo, van Ommen, Pannasch, Danteny-Dordoigne, & Rämä, 2017), and, indeed, by 15 months infants can detect subtle referential *discontinuity* between objects in sequentially presented scenes (Duh & Wang, 2014). Given that referential continuity across sequential images is not recognized until years later—ages 4 to 6 years old (Bornens, 1990)—and requires exposure, it implies drawn images require proficiency beyond natural percepts.

Altogether, this literature suggests that visual narrative comprehension develops incrementally, given exposure to an external system (i.e., comics and picture stories). Yet, because no dedicated field has yet consolidated around studying visual narratives, much research is still required to better understand this developmental trajectory, how and whether it differs from narrative development in other modalities, and how it interacts with the work on processing discussed above.

5 Conclusions

This review illustrates the complexity underlying visual narrative comprehension across processing, cognition, and development. Overall, visual narratives involve multiple interacting cognitive mechanisms, which may vary in their domain-generality or domain-specificity. Insofar as they overlap with language, visual narratives provide a way to test theories of the uniqueness of linguistic processing in a different non-verbal, highly conventionalized, sequentially presented, meaning-making modality.

Also, like language, the development of visual narratives balances exposure and practice. For visual narratives, exposure introduces additional complications because they typically appear in the context of multimodal interactions, with both text and images contributing meaning. Indeed, for young children, exposure to visual narratives often occurs in the multimodal context of picture stories along with verbal and gestural interactions with a caregiver. Thus, not only must we characterize the processing and development of the verbal and visual elements individually, but also their multimodal interactions (Cohn, 2016b).

Beyond language, research can also investigate how visual narrative development is concurrent with or interacts with other processes often claimed to motivate visual narrative understanding, like sequential reasoning, temporal cognition, event cognition, theory of mind, and general narrative abilities. Indeed, visual narratives are often used in tasks to study the development of Theory of Mind (Baron-Cohen et al., 1986; Sivaratnam et al., 2012), temporal cognition (Ingber & Eden, 2011; Weist, 2009), and sequential reasoning (Zampini et al., 2017), as well as general IQ tests (Kaufman & Lichtenberger, 2006; Wechsler, 1981). However, the age range for visual narrative comprehension (4-6 years old) typically occurs *after* the age of onset for

many of these cognitive abilities, meaning that studying those abilities using visual narratives in tasks with younger children may face significant confounds. These task limitations go largely unrecognized by researchers, who assume visual narratives are transparent and expertise-neutral stimuli.

Altogether, scholarship on the understanding of visual narratives extends across many fields, from psycholinguistics, scene perception and cognitive neuroscience, to cross-cultural and developmental psychology. Such reach should not be surprising, given that meaning-making across drawn images is a basic human ability spanning across history and cultures, and is evident in the earliest of records of human intelligence in cave paintings. Despite this, a consolidated field to study their cognition is only just emerging. Visual narratives thus provide a rich opportunity to study many fundamental questions about the mind.

6 References

- Abelman, R. (1990). You can't get there from here: Children's understanding of time-leaps on television. *Journal of Broadcasting & Electronic Media*, 34(4), 469-476. doi:10.1080/08838159009386755
- 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, NY: ACM.
- 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): 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. doi:<http://dx.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*, 1-21. doi:10.1080/21670811.2017.1279020
- Baron-Cohen, S., Leslie, A. M., & Frith, U. (1986). Mechanical, behavioural and intentional understanding of picture stories in autistic children. *British Journal of Developmental Psychology*, 4(2), 113-125.
- Barr, R., Zack, E., Garcia, A., & Muentener, P. (2008). Infants' Attention and Responsiveness to Television Increases With Prior Exposure and Parental Interaction. *Infancy*, 13(1), 30-56. doi:10.1080/15250000701779378
- Bateman, J. A., & Wildfeuer, J. (2014). A multimodal discourse theory of visual narrative. *Journal of Pragmatics*, 74, 180-208. doi:10.1016/j.pragma.2014.10.001
- Berman, R. A., & Slobin, D. I. (1994). *Relating events in narrative: A crosslinguistic developmental study*. New Jersey: Lawrence Erlbaum Associates.
- Bingham, A. B., Rembold, K. L., & Yussen, S. R. (1986). Developmental change in identifying main ideas in picture stories. *Journal of Applied Developmental Psychology*, 7(4), 325-340. doi:[https://doi.org/10.1016/0193-3973\(86\)90003-1](https://doi.org/10.1016/0193-3973(86)90003-1)
- Bishop, A. (1977). Is a Picture Worth a Thousand Words? *Mathematics Teaching*, 81, 32-35.
- Bornens, M.-T. (1990). Problems brought about by "reading" a sequence of pictures. *Journal of Experimental Child Psychology*, 49(2), 189-226. doi:[http://dx.doi.org/10.1016/0022-0965\(90\)90055-D](http://dx.doi.org/10.1016/0022-0965(90)90055-D)
- Bott, O. (2010). *The processing of events* (Vol. 162). Amsterdam: John Benjamins Publishing Company.
- 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*, n/a-n/a. doi:10.1111/cogs.12461
- 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.
- Chomsky, N. (1965). *Aspects of the Theory of Syntax*. Cambridge, MA: MIT Press.

- Coderre, E. L., Cohn, N., Slipper, 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. (2012). *Structure, meaning, and constituency in visual narrative comprehension*. (Doctoral Dissertation), Tufts University, Medford, MA.
- Cohn, N. (2013a). *The visual language of comics: Introduction to the structure and cognition of sequential images*. London, UK: Bloomsbury.
- Cohn, N. (2013b). Visual narrative structure. *Cognitive Science, 37*(3), 413-452. doi:10.1111/cogs.12016
- Cohn, N. (2014). You're a good structure, Charlie Brown: The distribution of narrative categories in comic strips. *Cognitive Science, 38*(7), 1317-1359. doi:10.1111/cogs.12116
- Cohn, N. (2015). Narrative conjunction's junction function: The interface of narrative grammar and semantics in sequential images. *Journal of Pragmatics, 88*, 105-132. doi:10.1016/j.pragma.2015.09.001
- Cohn, N. (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. doi:10.1016/j.cognition.2015.10.007
- Cohn, N. (Under review). Your brain on comics: A cognitive model of visual narrative comprehension. *Topics in Cognitive Science*.
- 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. doi:<http://dx.doi.org/10.1037/xlm0000290>
- Cohn, N., & Foulsham, T. (In prep). Zooming in on the cognitive neuroscience of visual narrative.
- Cohn, N., Jackendoff, R., Holcomb, P. J., & Kuperberg, G. R. (2014). The grammar of visual narrative: Neural evidence for constituent structure in sequential image comprehension. *Neuropsychologia, 64*, 63-70. doi:10.1016/j.neuropsychologia.2014.09.018
- Cohn, N., & Kutas, M. (2015). Getting a cue before getting a clue: Event-related potentials to inference in visual narrative comprehension. *Neuropsychologia, 77*, 267-278. doi:10.1016/j.neuropsychologia.2015.08.026
- Cohn, N., & 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. doi: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. doi:10.1016/j.brainres.2015.01.018
- Cohn, N., Paczynski, M., Jackendoff, R., Holcomb, P. J., & Kuperberg, G. R. (2012). (Pea)nuts and bolts of visual narrative: Structure and meaning in sequential image comprehension. *Cognitive Psychology, 65*(1), 1-38. doi:10.1016/j.cogpsych.2012.01.003
- Cohn, N., & Wittenberg, E. (2015). Action starring narratives and events: Structure and inference in visual narrative comprehension. *Journal of Cognitive Psychology, 27*(7), 812-828. doi:10.1080/20445911.2015.1051535
- Collins, W. A., Wellman, H., Keniston, A. H., & Westby, S. D. (1978). Age-Related Aspects of Comprehension and Inference from a Televised Dramatic Narrative. *Child Development, 49*(2), 389-399.

- Cook, B. L. (1980). Picture communication in the Papua New Guinea. *Educational Broadcasting International*, 13(2), 78-83.
- Culicover, P. W., & Jackendoff, R. (2005). *Simpler Syntax*. Oxford: Oxford University 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. doi: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. doi:<https://doi.org/10.1016/j.neuropsychologia.2018.09.016>
- Duh, S., & Wang, S.-h. (2014). Infants detect changes in everyday scenes: The role of scene gist. *Cognitive Psychology*, 72, 142-161. doi:<https://doi.org/10.1016/j.cogpsych.2014.03.001>
- 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. doi:10.15012/00000777
- Federmeier, K. D., & Kutas, M. (2001). Meaning and modality: Influences of context, semantic memory organization, and perceptual predictability on picture processing. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, 27(1), 202-224.
- Foulsham, T., & Cohn, N. (In prep). Zooming in on visual narrative comprehension.
- Foulsham, T., Wybrow, D., & Cohn, N. (2016). Reading without words: Eye movements in the comprehension of comic strips. *Applied Cognitive Psychology*, 30, 566-579. doi:10.1002/acp.3229
- Friedman, W. J. (1990). Children's Representations of the Pattern of Daily Activities. *Child Development*, 61(5), 1399-1412. doi:10.1111/j.1467-8624.1990.tb02870.x
- 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.
- Ganea, P. A., Pickard, M. B., & DeLoache, J. S. (2008). Transfer between Picture Books and the Real World by Very Young Children. *Journal of Cognition and Development*, 9(1), 46-66. doi:10.1080/15248370701836592
- Ganis, G., Kutas, M., & Sereno, M. I. (1996). The search for "common sense": An electrophysiological study of the comprehension of words and pictures in reading. *Journal of Cognitive Neuroscience*, 8, 89-106.
- Gawne, L. (2016). *A sketch grammar of Lamjung Yolmo*. Canberra: Asia-Pacific Linguistics.
- Gernsbacher, M. A. (1990). *Language Comprehension as Structure Building*. Hillsdale, NJ: Lawrence Erlbaum.
- Goldberg, A. (1995). *Constructions: A Construction Grammar Approach to Argument Structure*. Chicago, IL: University of Chicago Press.
- Graesser, A. C., Millis, K. K., & Zwaan, R. A. (1997). Discourse Comprehension. *Annual Review of Psychology*, 48, 163-189.
- Hagoort, P. (2017). The core and beyond in the language-ready brain. *Neuroscience & Biobehavioral Reviews*, 81, 194-204. doi:<https://doi.org/10.1016/j.neubiorev.2017.01.048>
- Hagoort, P., Brown, C. M., & Groothusen, J. (1993). The syntactic positive shift (SPS) as an ERP measure of syntactic processing. In S. M. Garnsey (Ed.), *Language and cognitive processes. Special issue: Event-related brain potentials in the study of language* (Vol. 8, pp. 439-483). Hove: Lawrence Erlbaum Associates.

- 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. doi:[https://doi.org/10.1016/S1388-2457\(02\)00161-X](https://doi.org/10.1016/S1388-2457(02)00161-X)
- Helo, A., van Ommen, S., Pannasch, S., Danteny-Dordoigne, L., & Rämä, P. (2017). Influence of semantic consistency and perceptual features on visual attention during scene viewing in toddlers. *Infant Behavior and Development*, *49*, 248-266. doi:<https://doi.org/10.1016/j.infbeh.2017.09.008>
- 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.
- Hutson, J. P., Magliano, J., & Loschky, L. C. (2018). Understanding Moment-to-moment Processing of Visual Narratives. *Cognitive Science*.
- Ildirar, S., Levin, D. T., Schwan, S., & Smith, T. J. (2018). Audio Facilitates the Perception of Cinematic Continuity by First-Time Viewers. *Perception*, *47*(3), 276-295. doi:10.1177/0301006617745782
- Ildirar, S., & Schwan, S. (2015). First-time viewers' comprehension of films: Bridging shot transitions. *British Journal of Psychology*, *106*(1), 133-151. doi:10.1111/bjop.12069
- Ingber, S., & Eden, S. (2011). Enhancing sequential time perception and storytelling ability of deaf and hard of hearing children. *American Annals of the Deaf*, *156*(4), 391-401.
- Johnson, S. P. (2013). Development of the Visual System. In J. L. R. Rubenstein & P. Rakic (Eds.), *Neural Circuit Development and Function in the Brain: Comprehensive Developmental Neuroscience* (Vol. 3, pp. 249-269): Elsevier Inc.
- Karmiloff-Smith, A. (1985). Language and cognitive processes from a developmental perspective. *Language and Cognitive Processes*, *1*(1), 61-85. doi:10.1080/01690968508402071
- Kaufman, A. S., & Lichtenberger, E. O. (2006). *Assessing Adolescent and Adult Intelligence* (3rd ed.). Hoboken: Wiley.
- Khetarpal, K., & Jain, E. (2016). *A preliminary benchmark of four saliency algorithms on comic art*. Paper presented at the Multimedia & Expo Workshops (ICMEW), 2016 IEEE International Conference on.
- Kim, J., & Monroy-Hernandez, A. (2015). Storia: Summarizing Social Media Content based on Narrative Theory using Crowdsourcing. *arXiv preprint arXiv:1509.03026*.
- Kluender, R., & Kutas, M. (1993). Bridging the gap: Evidence from ERPs on the processing of unbound dependencies. *Journal of Cognitive Neuroscience*, *5*(2), 196-214.
- 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. (2007). Neural mechanisms of language comprehension: Challenges to syntax. *Brain Research*, *1146*, 23-49.
- Kuperberg, G. R. (2013). The pro-active comprehender: What event-related potentials tell us about the dynamics of reading comprehension. In B. Miller, L. Cutting, & P. McCardle (Eds.), *Unraveling the Behavioral, Neurobiological, and Genetic Components of Reading Comprehension* (pp. 176-192). Baltimore: Paul Brookes Publishing.
- 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.
- Laubrock, J., Hohenstein, S., & Kümmerer, M. (2018). Attention to comics: Cognitive processing during the reading of graphic literature. In A. Dunst, J. Laubrock, & J. Wildfeuer (Eds.), *Empirical Comics Research: Digital, Multimodal, and Cognitive Methods* (pp. 239-263). New York: Routledge.
- Lauer, T., Cornelissen, T. H., Draschkow, D., Willenbockel, V., & Võ, M. L.-H. (2018). The role of scene summary statistics in object recognition. *Scientific Reports*, *8*(14666), 1-12.
- 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. (Under review). The Scene Perception & Event Comprehension Theory (SPECT) Applied to Visual Narratives. *Topics in Cognitive Science*.
- 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*, 0-0. doi:10.1080/0163853X.2015.1136870
- Magliano, J. P., Kopp, K., McNeerney, 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. doi: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. doi: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 Publishing Co.
- Magliano, J. P., & Zacks, J. M. (2011). The impact of continuity editing in narrative film on event segmentation. *Cognitive Science*, *35*(8), 1489-1517. doi:10.1111/j.1551-6709.2011.01202.x
- Mandler, J. M., & Johnson, N. S. (1977). Remembrance of things parsed: Story structure and recall. *Cognitive Psychology*, *9*, 111-151.
- 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. doi: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, USA, November 15-18, 2016, Proceedings, Los Angeles, CA.

- McCloud, S. (1993). *Understanding Comics: The Invisible Art*. New York, NY: Harper Collins.
- McNamara, D. S., & Magliano, J. (2009). Toward a comprehensive model of comprehension. *Psychology of learning and motivation, 51*, 297-384.
- McPherson, W. B., & Holcomb, P. J. (1999). An electrophysiological investigation of semantic priming with pictures of real objects. *Psychophysiology, 36*(1), 53-65.
- Munk, C., Rey, G. D., Diergarten, A. K., Nieding, G., Schneider, W., & Ohler, P. (2012). Cognitive Processing of Film Cuts Among 4- to 8-Year-Old Children. *European Psychologist, 17*(4), 257-265. doi:10.1027/1016-9040/a000098
- Nakazawa, J. (1997). *Development of manga reading comprehension: Developmental and experimental differences in adults*. Paper presented at the Proceedings of the 8th Annual Conference of Japan Society of Developmental Psychology.
- Nakazawa, J. (2004). Manga (comic) literacy skills as determinant factors of manga story comprehension. *Manga Studies, 5*, 7-25.
- Nakazawa, J., & Nakazawa, S. (1993). Development of *manga* reading comprehension: How do children understand *manga*? In Y. Akashi (Ed.), *Manga and child: How do children understand manga?* (pp. 85-189): Research report of Gendai Jidobunka Kenkyukai.
- Noble, G. (1975). *Children in front of the small screen*. London: Constable.
- Osterhout, L., & Holcomb, P. (1992). Event-related potentials elicited by syntactic anomaly. *Journal of Memory and Language, 31*, 758-806.
- Paczynski, M., Jackendoff, R., & Kuperberg, G. (2014). When Events Change Their Nature: The Neurocognitive Mechanisms Underlying Aspectual Coercion. *Journal of Cognitive Neuroscience, 26*(9), 1905-1917. doi:10.1162/jocn_a_00638
- Pantaleo, S. (2012). Exploring the intertextualities in a grade 7 student's graphic narrative. *L1 Educational Studies in Language and Literature, 12*, Running Issue(Running Issue), 23-55. doi:10.17239/l1esll-2012.04.01
- Pantaleo, S. (2013a). Matters of Design and Visual Literacy: One Middle Years Student's Multimodal Artifact. *Journal of Research in Childhood Education, 27*(3), 351-376. doi:10.1080/02568543.2013.796334
- Pantaleo, S. (2013b). Paneling "Matters" in Elementary Students' Graphic Narratives. *Literacy Research and Instruction, 52*(2), 150-171. doi:10.1080/19388071.2012.754973
- Pantaleo, S. (2015). Exploring the intentionality of design in the graphic narrative of one middle-years student. *Journal of Graphic Novels and Comics, 6*(4), 398-418.
- Pantaleo, S. (2017). The semantic and syntactic qualities of paneling in students' graphic narratives. *Visual Communication*. doi:10.1177/1470357217740393
- Paris, A. H., & Paris, S. G. (2003). Assessing narrative comprehension in young children. *Reading Research Quarterly, 38*(1), 36-76. doi:doi:10.1598/RRQ.38.1.3
- Patel, A. D. (2003). Language, music, syntax and the brain. *Nature Neuroscience, 6*(7), 674-681. doi:10.1038/nm1082
- Pempek, T. A., Kirkorian, H. L., Richards, J. E., Anderson, D. R., Lund, A. F., & Stevens, M. (2010). Video comprehensibility and attention in very young children. *Developmental Psychology, 46*(5), 1283.
- Poulsen, D., Kintsch, E., Kintsch, W., & Premack, D. (1979). Children's comprehension and memory for stories. *Journal of Experimental Child Psychology, 28*(3), 379-403. doi:[https://doi.org/10.1016/0022-0965\(79\)90070-5](https://doi.org/10.1016/0022-0965(79)90070-5)

- Richards, J. E., & Cronise, K. (2000). Extended Visual Fixation in the Early Preschool Years: Look Duration, Heart Rate Changes, and Attentional Inertia. *Child Development, 71*(3), 602-620. doi:doi:10.1111/1467-8624.00170
- Rumelhart, D. E. (1975). Notes on a schema for stories. In D. Bobrow & A. Collins (Eds.), *Representation and understanding* (pp. 211-236). New York, NY: Academic Press.
- San Roque, L., Gawne, L., Hoenigman, D., Miller, J. C., Rumsey, A., Spronck, S., . . . Evans, N. (2012). Getting the story straight: Language fieldwork using a narrative problem-solving task. *Language documentation and conservation, 6*, 135-174.
- Saraceni, M. (2016). Relatedness: Aspects of textual connectivity in comics. In N. Cohn (Ed.), *The Visual Narrative Reader* (pp. 115-129). London: Bloomsbury.
- Schmidt, C. R., & Paris, S. G. (1978). Operativity and Reversibility in Children's Understanding of Pictorial Sequences. *Child Development, 49*(4), 1219-1222. doi:10.2307/1128764
- Schwan, S., & Ildirar, S. (2010). Watching Film for the First Time: How Adult Viewers Interpret Perceptual Discontinuities in Film. *Psychological Science, 21*(7), 970-976. doi:10.1177/0956797610372632
- Silver, L. D. (2002). *Linguistic and pictorial narratives in preschool children: An exploration into the development of symbolic representation*. (Doctoral Dissertation), University of California, Berkeley,
- Simcock, G., Garrity, K., & Barr, R. (2011). The Effect of Narrative Cues on Infants' Imitation From Television and Picture Books. *Child Development, 82*(5), 1607-1619. doi:doi:10.1111/j.1467-8624.2011.01636.x
- Sivaratnam, C. S., Cornish, K., Gray, K. M., Howlin, P., & Rinehart, N. J. (2012). Brief Report: Assessment of the Social-Emotional Profile in Children with Autism Spectrum Disorders using a Novel Comic Strip Task. *Journal of Autism and Developmental Disorders, 42*(11), 2505-2512. doi:10.1007/s10803-012-1498-8
- Smith, R., Anderson, D. R., & Fischer, C. (1985). Young Children's Comprehension of Montage. *Child Development, 56*(4), 962-971.
- Smith, T. J. (2012). The attentional theory of cinematic continuity. *Projections, 6*(1), 1-27. doi:<http://dx.doi.org/10.3167/proj.2012.060102>
- 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.
- Stoermer, M. (2009). *Teaching between the frames: Making comics with seven and eight year old children, a search for craft and pedagogy*. (Doctoral Dissertation), Indiana University, Indiana.
- Szawerna, M. (2017). *Metaphoricity of Conventionalized Diegetic Images in Comics: A Study in Multimodal Cognitive Linguistics*: Peter Lang Publishing.
- Takayama, K., Johan, H., & Nishita, T. (2012, November 21-24, 2012). *Face detection and face recognition of cartoon characters using feature extraction*. Paper presented at the Image, Electronics and Visual Computing Workshop, Kuching, Malaysia.
- Trabasso, T., & Nickels, M. (1992). The development of goal plans of action in the narration of a picture story. *Discourse Processes, 15*, 249-275.
- Trabasso, T., & Stein, N. L. (1994). Using Goal-Plan Knowledge to Merge the Past with the Present and the Future in Narrating Events on Line. In M. M. Haith, J. B. Benson, R. J. Roberts Jr., & B. F. Pennington (Eds.), *The Development of Future-Oriented Processes* (pp. 323-349). Chicago: University of Chicago Press.

- Trabasso, T., & Stein, N. L. (1997). Narrating, representing, and remembering event sequences. In P. van den Broek, P. J. Baur, & T. Bourg (Eds.), *Developmental spans in event comprehension and representation: Bridging fictional and actual events*. New York: Lawrence Erlbaum Associates.
- van Berkum, J. J. A. (2009). The neuropragmatics of “simple” utterance comprehension: An ERP review. In U. Sauerland & K. Yatsushiro (Eds.), *Semantics and Pragmatics: From Experiment to Theory* (pp. 276-316). Basingstoke: Palgrave.
- van Dijk, T., & Kintsch, W. (1983). *Strategies of Discourse Comprehension*. New York: Academic Press.
- Van Petten, C., & Kutas, M. (1991). Influences of semantic and syntactic context on open- and closed-class words. *Memory and Cognition*, *19*, 95-112.
- Van Petten, C., & Luka, B. J. (2012). Prediction during language comprehension: Benefits, costs, and ERP components. *International Journal of Psychophysiology*, *83*(2), 176-190. doi:<http://dx.doi.org/10.1016/j.ijpsycho.2011.09.015>
- Wechsler, D. (1981). *WAIS-R manual: Wechsler adult intelligence scale-revised*: Psychological Corporation.
- Weist, R. M. (2009). Children think and talk about time and space. In P. Łobacz, P. Nowak, & Z. W. (Eds.), *Language, Science, and Culture*. Poznań: Wydawnictwo Naukowe UAM.
- Weist, R. M., Atanassova, M., Wysocka, H., & Pawlak, A. (1999). Spatial and temporal systems in child language and thought: a cross-linguistic study. *First Language*, *19*(57), 267-308. doi:10.1177/014272379901905701
- Weist, R. M., Lyytinen, P., Wysocka, J., & Atanassova, M. (1997). The interaction of language and thought in children's language acquisition: a crosslinguistic study. *Journal of Child Language*, *24*(01), 81-121. doi:doi:null
- 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. (1988). The Artistic Tower of Babel: Inextricable Links Between Culture and Graphic Development. In G. W. Hardiman & T. Zernich (Eds.), *Discerning Art: Concepts and Issues* (pp. 488-506). Champaign, IL: Stipes Publishing Company.
- 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.
- Wilson, B., & Wilson, M. (1979). Children's Story Drawings: Reinventing Worlds. *School Arts*, *78*(8), 6-11.
- Wilson, B., & Wilson, M. (1982). *Teaching Children to Draw*. Englewood Cliffs, NJ: Prentice-Hall.
- Wittenberg, E., Paczynski, M., Wiese, H., Jackendoff, R., & Kuperberg, G. (2014). The difference between “giving a rose” and “giving a kiss”: Sustained neural activity to the light verb construction. *Journal of Memory and Language*, *73*(0), 31-42. doi:<http://dx.doi.org/10.1016/j.jml.2014.02.002>
- Yano, M. (2018). Predictive processing of syntactic information: evidence from event-related brain potentials. *Language, Cognition and Neuroscience*, 1-15. doi: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.

- Zampini, L., Zanchi, P., Suttora, C., Spinelli, M., Fasolo, M., & Salerni, N. (2017). Assessing sequential reasoning skills in typically developing children. *BPA-Applied Psychology Bulletin (Bollettino di Psicologia Applicata)*, 65(279), 44-50.
- Zwaan, R. A., & Radvansky, G. A. (1998). Situation models in language comprehension and memory. *Psychological Bulletin*, 123(2), 162-185.