

The cognitive systems of visual and multimodal narratives

Neil Cohn (neilcohn@visuallanguagelab.com)

Department of Communication and Cognition, Tilburg University, The Netherlands

Emily Coderre (emily.coderre@med.uvm.edu) & Elizabeth O'Donnell

Department of Psychology, University of Vermont, USA

Aidan Osterby (aidan.osterby@gmail.com) & Lester C. Loschky (loschky@ksu.edu)

Department of Psychology, Kansas State University, USA

Keywords: visual narrative; visual language; event perception; scene perception; attention; comics; autism

Introduction

Cognitive research on visual and multimodal narratives has been burgeoning in the last decade. An increasing number of researchers in the psychological sciences have turned to examining sequential images from a variety of subdisciplines, particularly those from fields of linguistics (Cohn, 2013a), discourse studies (Magliano, Larson, Higgs, & Loschky, 2015), the perceptual sciences (Foulsham, Wybrow, & Cohn, 2016; Loschky, Hutson, Smith, Smith, & Magliano, In press), and the cognitive neurosciences (Cohn, Paczynski, Jackendoff, Holcomb, & Kuperberg, 2012).

This broad coverage by different subfields of the cognitive sciences is testament to how complex sequential images can be, especially when they combine in multimodal interactions, such as with text. Indeed, visual narratives have proven to be a good testing-ground for many facets of basic cognition.

The presentations in this symposium highlight several emerging lines of research on visual narratives in the cognitive sciences, spanning across the subfields of psycholinguistics, scene perception, cognitive neuroscience, and clinical psychology.

Changes in attentional breadth while viewing visual narratives: A test of the Scene Perception & Event Comprehension Theory (SPECT)

The first presentation by **Lester C. Loschky, Ryan V. Ringer, Maverick E. Smith, John P. Hutson, Heather R. Bailey, Jeffrey M. Zacks, and Joseph P. Magliano** examines visual narratives from the perspective of scene perception and event cognition. We tested a novel hypothesis of the Scene Perception & Event Comprehension Theory (SPECT) (Loschky, et al., In press) regarding the relationship between front-end processes, which operate during single eye fixations (e.g., attentional selection), and back-end processes, which occur across fixations in memory (e.g., shifting to build a new event model), while viewing visual narratives. Event perception research has shown better memory for objects fixated at event boundaries than event middles. Scene perception research suggests a potential connection between

memory at event boundaries and attentional breadth. During the first two seconds of picture viewing, viewers make long saccades and short fixations (ambient mode), but by 4-6 seconds, they make short saccades and long fixations (focal mode). Viewers' average saccade length measures the extent of their overt visual attention. Integrating these findings, SPECT proposes that visual attention broadens at event boundaries and then narrows in the middles. We tested this hypothesis in two lines of research. First, we used picture stories similar to comics, except one views the narrative sequence one picture at a time. One group of viewers segmented 6 picture stories, and another group viewed them while we tracked their eyes (overt attention). Viewers remained in the ambient mode longer while viewing pictures at normative event boundaries than middles. Second, we had participants view videos of everyday events while performing the gaze-contingent useful field of view visual discrimination task to measure their covert attentional breadth at normative event boundaries and middles. Around event boundaries, viewers' attention expanded, then narrowed toward the middles, consistent with SPECT's hypothesis. Thus, using both overt and covert attention measures, in picture stories and everyday event videos, we supported SPECT's hypothesis in visual narratives.

Complexity in Visual Narrative Grammar

Recent work by **Neil Cohn** on visual narrative has posited that sequential images draw on a Visual Narrative Grammar (VNG) that uses similar principles to linguistic syntax, only at a discourse level of semantics (Cohn, 2013b). Early empirical work on VNG sought to establish the validity of basic constructs such as distributionality of narrative categories, constituent structure, and the independence of narrative structure from semantics. In addition, studies of the electrophysiology of the brain using event-related potentials have implicated similar neural responses to the manipulation of narrative and semantics as to syntax (P600, LAN) and semantics (N400) in sentence processing (Cohn, Jackendoff, Holcomb, & Kuperberg, 2014; Cohn et al., 2012). Here, we focus on a new stage of research which looks at additional hallmarks of syntactic structure in narrative grammar: relationships of units and constituents across distance dependencies, such as those where the same content may be rearranged to yield a different structure. Such work shows

further that sequential image comprehension extends beyond juxtaposed semantic relationships between images, and highlights the “syntactic” nature of narrative grammar.

The online processing of multimodal narratives

While there is a large body of work on traditional text narratives, and growing work on wordless visual narratives, there is relatively little on the comprehension of multimodal narratives, where text and images combine. This presentation by **Aidan Osterby, Neil Cohn, Lester C. Loschky, and Joseph P. Magliano** examines the interaction between sequential images and text. Cohn (2016) proposed a framework in which multimodal relationships are classified based on whether one or both modalities convey the primary meaning, and whether one or both modalities contain a necessary structural ordering, or grammar. The current study explores “co-assertive” cases—when both modalities balance their expression of meaning, and both use a grammatical structure. Here, we present two experiments which examine the self-paced viewing times to participants who were shown comics strips that modulate their structural contributions of text and images—via deletion and reordering of neither, one, or both modalities. Our findings support that text-image interactions balance relationships beyond meaning alone, and such work provides initial support for examining these complex multimodal interactions in systematic ways.

The roles of semantic relatedness and narrative structure in narrative comprehension in individuals with ASD

The final presentation by **Elizabeth O’Donnell, Neil Cohn, and Emily Coderre** examines visual narratives from a different perspective, by examining the comprehension abilities of a different population. Individuals with autism spectrum disorders (ASD) often struggle with narrative comprehension. Such difficulties could arise from impaired semantic processing and/or impaired grammatical processing, both of which have been documented in this population. Most studies investigating semantic processing and grammatical comprehension have used written or spoken sentences or narratives; yet language deficits are also documented in ASD, which may confound results. This makes visual narratives an ideal venue for investigating more global aspects of narrative comprehension in this population. In previous work we have shown that semantic processing is impaired for both linguistic and visual narratives in ASD (Coderre et al., Under Review). It remains to be seen, however, whether narrative comprehension difficulties might also stem from impairments in structural sequencing abilities that are required for narrative comprehension. We investigate semantic and structural (grammatical) processing in non-linguistic narrative comprehension to determine the relative contributions of each to the purported impairments in narrative comprehension in individuals with ASD. To do so we replicate a previous study of sequential image comprehension (Cohn et al., 2012) in a population of adults

with ASD and typically-developing (TD) adults. Stimuli consist of normal sequences (containing both structure and meaning); semantic-only sequences (containing meaning between panels but no narrative structure); structural-only sequences (containing a narrative structure but no semantic relatedness); and scrambled sequences (randomly-ordered panels with neither semantic relatedness nor narrative structure). Semantic processing is evaluated by comparing the effect of sequence type on the N400 component of the event-related potential (ERP), while structural processing is evaluated via the left anterior negativity (LAN) effect. Preliminary results will be discussed with respect to patterns of semantic and structural processing both within and between groups. This work offers a novel insight into mechanisms of narrative comprehension in ASD.

References

- Coderre, E. L., Cohn, N., Slipher, S. K., Chernenok, M., Ledoux, K., & Gordon, B. (Under Review). Visual and linguistic narrative comprehension in autism spectrum disorders: Neural evidence for modality-independent impairments.
- 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. (2016). A multimodal parallel architecture: A cognitive framework for multimodal interactions. *Cognition*, 146, 304-323. doi:10.1016/j.cognition.2015.10.007
- 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., 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
- 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
- Loschky, L. C., Hutson, J. P., Smith, M. E., Smith, T. J., & Magliano, J. (In press). Viewing Static Visual Narratives Through the Lens of the Scene Perception and Event Comprehension Theory (SPECT). In J. Laubrock, J. Wildfeuer, & A. Dunst (Eds.), *Empirical Comics Research: Digital, Multimodal, and Cognitive Methods*. London: Routledge.
- 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