

Neil Cohn\*, Ryan Taylor and Kaitlin Pederson

# A Picture is Worth More Words Over Time: Multimodality and Narrative Structure Across Eight Decades of American Superhero Comics

DOI 10.1515/mc-2017-0003

**Abstract:** The visual narratives of comics involve complex multimodal interactions between written language and the visual language of images, where one or the other may guide the meaning and/or narrative structure. We investigated this interaction in a corpus analysis across eight decades of American superhero comics (1940–2010s). No change across publication date was found for multimodal interactions that weighted meaning towards text or across both text and images, where narrative structures were present across images. However, we found an increase over time of narrative sequences with meaning weighted to the visuals, and an increase of sequences without text at all. These changes coincided with an overall reduction in the number of words per panel, a shift towards panel framing with single characters and close-ups rather than whole scenes, and an increase in shifts between temporal states between panels. These findings suggest that storytelling has shifted towards investing more information in the images, along with an increasing complexity and maturity of the visual narrative structures. This has shifted American comics from being textual stories with illustrations to being visual narratives that use text.

**Keywords:** visual language, multimodality, narrative structure, comics, discourse

## Introduction

Language rarely appears in isolation, and increasing attention has turned towards its interactions with other domains, such as gestures or images (Bateman 2014). Comics make a unique place to study multimodality because they involve complex interactions between written language and a “visual language” of images, with one or the other guiding the meaning and/or narrative structure (McCloud 1993; Cohn 2016). Yet, virtually no works have empirically characterized these complex multimodal interactions in actual visual narratives. Corpus analyses have become a growing tool to investigate aspects of visual narratives like the framing of panel content (Cohn et al. 2012b; Neff 1977; Cohn 2011), visual morphology (Cohn and Ehly 2016; Forceville 2011; Forceville et al. 2010; Abbott and Forceville 2011), and page layout (Pederson and Cohn 2016). However, studies of multimodality related to comics have mostly looked at onomatopoeia (Pratha et al. 2016; Forceville et al. 2010; Guynes 2014) and conceptual metaphors (Forceville 2016; Forceville and Urios-Aparisi 2009). Here, we used corpus analysis to explore complex multimodal interactions and their changes over time in one of the longest contiguous traditions of comics in the world: American superhero comics from the 1940s to the 2010s (Duncan and Smith 2009).

The most prominent theory of multimodality in comics comes from theorist and creator Scott McCloud (1993), who proposed a taxonomy of meaningful relationships between text-image combinations. Like other theories of multimodal relationships (Painter et al. 2012; Martinec and Salway 2005; Royce 2007; Stainbrook 2016; Bateman 2014; Bateman and Wildfeuer 2014a, 2014b), McCloud focuses on the surface semantic relations between text and image. He posited a tradeoff between meaning being focused more in the text (Word-Specific), the images (Picture-Specific), or balanced, including redundant expressions in both modalities (Duo-Specific) and non-intersecting expressions (Parallel), among others. Overall, this theory posited a gradation between the “semantic dominance” of one modality or another – i. e., which modality carries the “weight” of meaning.

---

\*Corresponding author: Neil Cohn, Tilburg center for Cognition and Communication, Tilburg University, 5000 LE Tilburg, Netherlands, E-mail: neilcohn@visuallanguagelab.com

Ryan Taylor, Kaitlin Pederson, Tilburg center for Cognition and Communication, Tilburg University, 5000 LE Tilburg, Netherlands

While McCloud’s (and others’) categories do well to characterize semantic relationships between individual text and images, they cannot account for differences that arise in multimodal image *sequences* (Cohn 2016). Consider Figure 1(a) and 1(b). Both of these sequences convey more information in the text than in the visuals: Figure 1(a) discusses the man’s wishes to marry, and Figure 1(b) shows him describing parts of his rundown house. We can confirm that the text is more semantically weighted than the visuals by using a “deletion test” where we omit the text, as in Figure 1(c) and 1(d). In both cases, the primary gist of the sequence disappears without the written language. The reverse should also be true: meaning would be retained if we only omitted the images (not shown).



**Figure 1:** Two sequences in which verbal information carries the weight of meaning, but which differ in the structure of the visual sequences. A Verb-Assertive interaction is used in (a) while a Verb-Dominant interaction is used in (b). Examples come from *Lady Luck* by Klaus Nordling (1949).

These tests suggest that Figure 1(a) and 1(b) both carry more weight of meaning in the text. Nevertheless, they differ significantly in the structure of their visual sequences. In Figure 1(a), the sequence has a coherent structure to it, even in the absence of text, as in Figure 1(c); clearly, the images progress one to the next with a coherent narrative. By contrast, the image sequence in Figure 1(b) does not carry the sequential order, as evident in Figure 1(d). Rather, these images are like a visual list (here, about different parts of a decrepit house), and would remain coherent as a sequence even if the images were rearranged (unlike Figure 1(a/c)). These sequential properties of multimodal relations go unaccounted for in descriptions of surface semantic relations alone (McCloud 1993; Stainbrook 2016; Bateman and Wildfeuer 2014a, 2014b; Tasić and Stamenković 2015): namely, where verbally weighted semantics differ in their contributions of narrative structure.

This difference in sequencing has been emphasized by recent work arguing that sequential images can be guided by a “visual narrative grammar.” This system organizes images into coherent sequences analogous to the way that syntactic structure organizes words into coherent sentences (Cohn 2013b). Like syntax, narrative grammar assigns categorical roles to panels and arranges them into hierarchic constituents, but does so at a discourse level of meaning because, as the saying goes, an individual picture “is worth a thousand words.” Importantly, this approach maintains an unambiguous separation between semantics and narrative, which interface together: The narrative grammar provides a coherent organization to semantic information (i. e., meaningful elements like events, objects, their properties, etc.). Psychological research has shown that manipulation of this narrative grammar elicits different brainwave responses than the manipulation of semantic structure (Cohn et al. 2012a, 2014), and these brainwave responses appear similar to those evoked by manipulations of grammar and meaning in sentence processing (Kaan 2007). Such findings provide evidence that semantics and narrative grammar operate on separate processing streams, and that the neurocognitive resources for the “visual languages” used in comics appear to be shared with that of sentences (Cohn et al. 2012a, 2014; Magliano et al. 2015).

Figure 1 contrasts a sequence that uses this narrative grammar in the visuals (1a) and one that does not (1b). Experimentation has shown that readers have a harder time comprehending wordless sequences composed of random panels (Cohn et al. 2012a) or scrambled panels of a sequence (Foulsham et al. 2016; Gernsbacher et al. 1990; Osaka et al. 2014) than semantically coherent sequences that also use a narrative grammar. In addition, processing differs between sequences that use narrative structure and/or semantic associations across images (Cohn et al. 2012a). This contrast is essential for understanding the sequences in Figure 1: though the text holds more semantic weight in both, Figure 1(a) uses a narrative grammar to bind the relations between images, while Figure 1(b) does not. Without the text, Figure 1(b) appears to use a sequence of semantically associated images (aspects of a house) with no guiding narrative structure. We could test this by rearranging the panels in the sequence, which should not disrupt the overall coherence in this case, indicating a lack of narrative structure.

The opposite pattern arises in Figure 2(a) and 2(b). In these cases, the visuals carry more semantic weight than the text. In Figure 2(a), a man (wearing a dress) captures a thief escaping a sewer, while in Figure 2(b) a woman trips and then captures hooligans robed in white sheets. Here, deleting the text in Figure 2(c) and 2(d) retains the gist despite the absence of written language. One can imagine the opposite deletion, omitting the images but retaining the text. The text here is not negligible, but it plays a secondary role to the meaning in the visuals. Both 2a and 2b use narrative structures – they maintain coherent sequencing across panels. What differs here is the grammatical contribution of the *text*: Figure 2(a) uses fully grammatical text (full sentences) while Figure 2(b) does not (predominantly only single word utterances like *Whooo ...* and *Gha-aa!*). Thus, in reverse of the sequences in Figure 1, those in Figure 2 give semantic weight to the visuals, and use narrative structure across images, but differ in the structure of the text.

Several interactive types emerge between text and images in a multimodal visual narrative sequence which may or may not involve grammatical structures, and may or may not invest more meaning in one domain or another. Cohn (2016) outlines three basic interactions between modalities. First, an **Autonomous** relation – which is not actually an interaction – is where only one modality is present. Writing in the absence of images would thus be *Verb-Autonomous*, while a sequence of images without any text would be



**Figure 2:** Two sequences in which visual information carries the weight of meaning, but which differ in the grammatical content of the verbal text. A Vis-Assertive interaction is used in (a) while a Vis-Dominant interaction is used in (b). Examples come from *Lady Luck* by Klaus Nordling (1949).

*Vis-Autonomous*. Second, a **Dominant** relationship balances the presence of multiple modalities, but at least one does not use a grammatical structure. For example, a *Verb-Dominant* relation would have syntactic text with an individual image, or with a sequence of images with no coherent connecting structure (as in Figure 1(b)). In contrast, a *Vis-Dominant* relation would have a sequence of images with a well-formed narrative structure, but with text lacking syntax – such as onomatopoeia – which do not exist within a sentence context (as in Figure 2(b)).

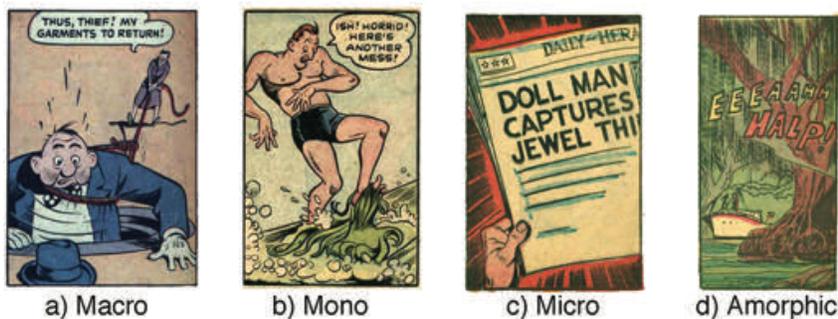
Finally, an **Assertive** relationship involves multiple modalities where all of them use a grammar. *Verb-Assertive* relations weight meaning more towards the text, though the visuals would still contribute meaning in a narrative sequence (as in Figure 1(a)). The same structural relations where the visuals carry more semantic weight would thus be *Vis-Assertive* (as in Figure 2(a)). If meaning is

balanced across both grammatical structures, the sequence would be *Co-Assertive*. Figure 2(a) is Vis-Assertive and not Co-Assertive because the primary gist is retained even when text is deleted (Figure 2(c)). This would not be possible if it was Co-Assertive – both modalities would be necessary for the meaning of the sequence. Altogether, these multimodal types balance the contributions between modality, grammar (syntax, narrative), and which modality holds more weight for the overall semantics of an interaction.

While empirical studies of multimodal interactions have not yet looked at comics – let alone American comics specifically – corpus analyses have examined other facets of sequential structure, such as McCloud’s (1993) analysis of the “transitions” between panels. These linear semantic relationships are similar to the coherence relations posited by theories of discourse (e. g., Zwaan and Radvansky 1998) which have been analyzed in visual narratives in the context of film (Magliano et al. 2001; Magliano and Zacks 2011). McCloud used his taxonomy of semantic changes between panels to analyze various comics around the world. He found that American comics use transitions related to the conveying of actions (~65%), followed by some specifying changes in characters (~20 %) and scenes (~14 %). While substantial evidence has shown panel transitions are insufficient for the comprehension of sequential images, and that a narrative grammar is also necessary (Cohn et al. 2014, 2012a), readers do appear sensitive to changes in meaningful information across sequential images (Cohn and Kutas 2015; Cohn and Kutas, Under Review, Magliano and Zacks 2011). In addition, such changes may be useful for characterizing aspects of the semantic and narrative structure of sequential images (Cohn and Bender 2017).

Additional research has examined how panels frame information in the narrative sequence. Panels differ in how much they depict “active” information, that is, information relevant to the narrative sequence (Cohn 2013a). **Macro** panels depict multiple interacting entities (e. g., characters, Figure 3(a)) within a scene, while **monos** show only single entities (Figure 3(b)), and **micros** frame less than a single entity (i. e., with a close up, Figure 3(c)). **Amorphic** panels show no active entities, such as in depictions of environmental information (i. e., the outside of a building or non-character scene information, Figure 3(d)). This variation in **attentional framing** of panel content is important for the narrative because providing less information – as in monos, micros, and amorphic panels – often requires additional inference of the broader scene (Schwan and Ildirar 2010), and may involve additional narrative structures (Cohn 2015, Cohn and Kutas Under Review). Furthermore, micros provide refined viewpoints characteristic of narrative modifiers, such as zooms that focus on information in prior panels (Cohn 2015, 2013a). Thus, like semantic relations between panels, attentional framing can be suggestive of complex aspects of narrative structure.

Corpus analyses have studied these attentional framing categories in American comics. An initial study (Cohn 2011) examining various American comics from 1986 to 2003 found more macros (59%) than monos (32%) or micros (6%). Subsequent research (Cohn et al. 2012b) found that American mainstream comics (i. e., superhero genre) from 1992 to 2005 again used more macros (47%) than monos (39%), micros (9%), or amorphic panels (4%). These proportions did not differ from panels used in comics from the American



**Figure 3:** Examples of attentional framing categories. (a–b) come from *Lady Luck* by Klaus Nordling (1949), and (c–d) come from *Doll Man* by Bill Quackenbush (1950).

Independent genre, though they did differ from those in Japanese manga, which used more monos than macros and increased numbers of micros and amorphous panels.

Based on these theoretical and empirical precedents, we investigated multimodal interactions over time in American superhero comics. First, we hypothesized that written language would decrease in semantic weight over time, along with an increase in the visuals conveying more meaning and more structure. We based this hypothesis on general observations about the development of storytelling across mainstream American comics (Moore 2003; Verano 2006; Mazur and Danner 2014; Duncan et al. 2015), and indeed our other corpus work has shown changes in other structures, like page layout (Pederson and Cohn 2016).

In addition, the mass importation of Japanese manga into the American comic market in the 1990s and 2000s (Goldberg 2010; Brienza 2016) has been said to influence the storytelling of American artists (McCloud 1996). This influx of manga has been cited as contributing to a “decompressed” storytelling – a comic industry term referring to a style in the past two decades where narratives extend across more panels and pages, rather than remaining as a terse narrative sequence (Mazur and Danner 2014; Burgas 2007; Cicci 2015). Additional research has shown substantial changes to aspects of film structure across the last 100 years (Cutting 2015; Cutting et al. 2010), and thus drawn visual narratives (often the basis for films in storyboards) may also have undergone structural changes in this same time period. All of these reasons forecast the potential for multimodality and narrative structures to have changed over time.

Second, if meaning shifts more to the visuals, we hypothesized that changes might also occur to the structure of the narrative sequences. This would be indicated both by the inherent status of narrative structure coded in multimodal types, and implied by changes to the framing of panels and/or semantic relationships across panels (Cohn 2013a, 2015). Such findings would be consistent with observations that complexity emerges in the structures of one modality when shifting away from the semantic dominance of another modality, be it from verbal to gestural (Goldin-Meadow 2006; Goldin-Meadow et al. 1996), or verbal to visual-graphic (Wilkins 1997/2016; Green 2014). Thus, changes in multimodal and narrative structures across time may provide insight on changes in a developing visual language.

## Methods

### Materials

We analyzed forty superhero comics from the 1940s to the 2010s, using the same collection of books as in other corpus work (Pederson and Cohn 2016). We focused on “superhero comics” from the United States of America, defined as those from the “mainstream” genre, typically from publishers like Marvel Comics and DC Comics. This genre often features protagonist(s) imbued with heroic or fantastical powers that allow them to fight for the good of mankind, while often concealing their own identities (Coogan 2006; Hatfield et al. 2013). These comics are often characterized by color printing, episodic issues, and are typically drawn with a style exaggerating the strength and sexuality of human forms, often in dynamic situations with a great deal of action (Duncan et al. 2015). This “style” is characterized in Visual Language Theory as the “Kirbyan” dialect of American Visual Language (Cohn 2013a).

Five books were chosen per decade from 1941 to 2014. We selected one or two books from the beginning, middle, and end of each decade. Most books were published by Marvel (17) and DC Comics (11), along with additional comics from the 1940s, 1950s, and 1960s now listed as Public Domain from online archives ([www.comicbookplus.com](http://www.comicbookplus.com)), including *Lady Luck* and *Doll Man*, as in Figures 1–3. Our sampling aimed for well-established comic series and characters, from a range of creators (both

“influential” and not). Most works reflected non-significant episodes of ongoing titles, and only one work could be called noteworthy within the broader canon of superhero comics (*Batman: The Killing Joke*). We hoped that such criteria resulted in a suitably generalizable sample. All works analyzed are listed in the Appendix.

We analyzed each comic panel-by-panel, either in their entirety or up to the first 25 pages, for the various features listed in the Areas of Analysis. We excluded side stories unrelated to the main character of the book, which occurred in some older comics. Altogether, our analysis across 40 comics examined 925 pages with 4,648 panels. All coded data is included in the Visual Language Research Corpus (VLRC): <http://www.visuallanguagelab.com/vlrc>.

## Areas of analysis

Each book was coded across four dimensions: Multimodal interactions, attentional framing, semantic relationships, and number of words per panel; year of publication was used as an indicator of a book's time period. Two researchers independently coded each book, and were trained in the identification of, and criteria for, all areas of analysis (described below). All coders completed pre-coding assessments where they exceeded the threshold of 90 % coding accuracy. Coders tended to be consistent, and we found an inter-rater reliability of Kappa = 0.517,  $p < 0.001$ , for multimodal types and Kappa = 0.656,  $p < 0.001$  for attentional categories. Semantic relations had an inter-rater reliability of Kappa = 0.863,  $p < 0.001$  for character changes, Kappa = 0.645,  $p < 0.001$  for spatial location changes, and Kappa = 0.659,  $p < 0.001$  for time changes.

### Multimodality

Our analysis of multimodal relationships followed the decision tree outlined in Cohn (2016). First, coders asked whether a panel used one or multiple modalities. If only one, the panel was categorized as *Autonomous* (*Vis-*, *Verb-*), but if multiple modalities were present, they then asked about semantic dominance: Does one modality control the meaning more than another? This factor was decided by imagining the omission of each modality and assessing whether the overall gist was retained (as in Figures 1 and 2). If the gist of the panel was maintained during omission, the retained modality inferred as semantically dominant (*Vis-*, *Verb-*). If the gist was not retained with omission, meaning was shared between modalities (*Co-*). Finally, coders asked whether grammatical structures appeared in both modalities (*-Assertive*) or only one (*-Dominant*). Coders assessed the narrative grammar in the visual modality using diagnostic tests outlined in previous research, such as tests of substitution, deletion, and rearrangement (Cohn 2013b, 2014, 2015). Thus, the presence of narrative structure is coded inherently within the criteria for multimodal interactions.

These interactive states assess the relative contributions of narrative structure and meaning. However, Verb-Dominant panels could contain a single sentence or several sentences as long as the text remains more semantically weighted than the non-narrative images. Though we may hypothesize that more words would lead to more semantic weight, this assumption cannot be taken for granted. Thus, to further investigate the density of text in these books – and whether it aligns with semantic weight – coders also counted the number of words total per panel.

### Semantic relations

We coded changes in semantic elements between panels along dimensions discussed in theories of visual narrative and discourse (Zwaan and Radvansky 1998; Magliano and Zacks 2011; McCloud 1993):

characters, spatial location, and time. Semantic shifts were considered to be non-mutually exclusive (i. e., panel relationships could have multiple coherence changes) and non-exhaustive (i. e., coherence changes could be both full and partial). Our coding followed protocols used to analyze previous experimental stimuli (Cohn and Bender 2017). Complete shifts in characters between panels were coded as a “1”, partial changes as “0.5” (i. e., one or more characters held constant but others added or omitted), and no change as “0”. We coded full shifts in spatial location as “1”, partial changes (such as changes within a common space, such as moving from one room to another in the same building) as “0.5”, and no changes in location as “0.” Temporal changes were designated with a “1” if time was interpreted as passing between panels, or “0” if time was ambiguous between panels.

### Attentional framing structure

Finally, we analyzed the “attentional framing structure” of panels. While counting the numbers of words per panel allowed for an assessment of the density of information in text, panel framing types similarly measured the amount of information conveyed by panels in a visual sequence. Our coding focused on the number of entities per panel, in line with prior theoretical (Cohn 2013a) and corpus research (Cohn et al. 2012b). A *macro* depicted multiple interacting entities (typically, characters, as in Figure 3(a)), while a *mono* framed only a single entity (Figure 3(b)). *Micros* depicted less than a single entity (Figure 3(c)), usually with a close up, and *amorphous panels* depicted non-active entities (Figure 3(d)), such as the exterior of a building or environmental scene information other than acting characters. Panels not conforming to these categories were coded as ambiguous (ex. a fully black panel or a panel with only text).

### Data analysis

For each classification under analysis, we calculated the mean rate of appearance per panel per book by the individual coders, and our final analysis averaged coders’ scores. We conducted analysis of variance (ANOVAs) for comparisons between categorical types across decades of publication date, setting multi-modal type, attentional framing, or semantic structure categories as within-group factors (see above) and decades of publication date as a between-groups factor. Because we were not interested in the incremental changes between decades, we did not statistically analyze pairwise differences between each decade. Rather, we analyzed polynomial contrasts to estimate the overall linear trend of a particular dimension across decades. Finally, targeted correlations investigated additional relationships between specific elements.

## Results

### Basic page properties

As reported in our previous analysis of this corpus (Pederson and Cohn 2016), pages used a mean of 5.04 panels per page per book (range: 3.6–7.4). Analysis across decades of publication date revealed linear trends implying a decrease towards fewer panels being used per page over time (Table 1). On average this shift went from 6.5 panels per page (1940s) to 4.2 panels per page (2010s).

**Table 1:** F-values and R-squared values for linear polynomial trends of various structures across decade of publication date.

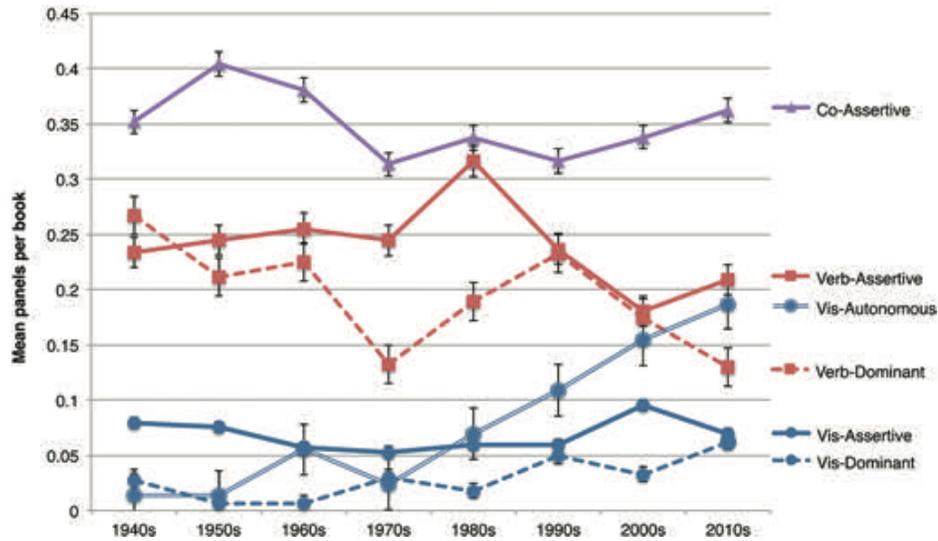
Visual narrative structures	Linear trend	
	F-Value	R <sup>2</sup>
<i>Page properties</i>		
Panels per page	15.998***	0.296
<i>Multimodality</i>		
Verb-Dominant	2.9 <sup>^</sup>	0.071
Verb-Assertive	1.07	0.027
Co-Assertive	0.31	0.008
Vis-Assertive	0.03	0.001
Vis-Dominant	12.58**	0.249
Vis-Autonomous	17.28***	0.313
Number of words	8.25**	0.178
<i>Semantic relations</i>		
Temporal changes	11.55**	0.233
Character changes	1.88	0.047
Spatial location changes	11.36**	0.23
<i>Attentional framing</i>		
Macro	2.67	0.066
Mono	8.97**	0.191
Micro	8.68**	0.186
Amorphic	0.417	0.011

Note: \*\*\* $p < .001$ , \*\* $p < .01$ , \* $p < .05$ , <sup>^</sup> $p < .1$ .

## Multimodality

Our first analysis looked at the multimodal interactions between semantics (Verb-, Vis-, Co-) and grammar (-Autonomous, -Dominant, -Assertive) across publication dates. We excluded Verb-Autonomous panels from analyses because they appeared in less than 1% of all panels across books, and did not appear in any books in the 2000s or 2010s. Overall, we found a main effect of multimodal type,  $F(5,160) = 65.3$ ,  $p < 0.001$ , but no interaction between multimodal type and publication date,  $F(35,160) = 0.962$ ,  $p = 0.536$ . This suggested that, though the individual multimodal types differed, there was little change in the relation between those differences across publication date. This is apparent in Figure 4, where the overall trends across decades had no uniform relational characteristics. Overall, Co-Assertive interactions appeared more than all other types, followed by Verb-Assertive and Verb-Dominant interactions. In contrast, Vis-Assertive and Vis-Dominant interactions were used the least of all types. Vis-Autonomous panels began as used the least, but became used in intermediate proportions (as discussed below). It is noteworthy that within relations of semantic dominance (Vis-, Verb-), Assertive relations were almost always greater than Dominant ones. This implies a preference for a balance in grammatical structures, rather than one form being more structurally rich.

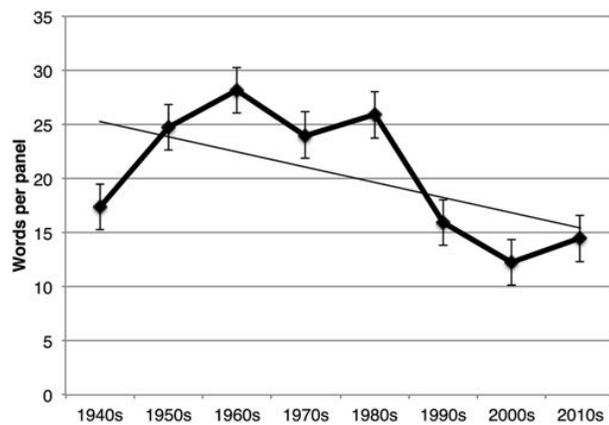
Our subsequent analysis looked across publication date for each of the types of multimodal interactions. We found no significant trends for Verb-Assertive, Co-Assertive, or Vis-Assertive relationships (all  $ps > 0.162$ ). Linear and quadratic trends approaching significance suggested a weak decrease in frequency for Verb-Dominant interactions across publication date (all  $ps = 0.097$ ). Finally, significant linear trends for Vis-Dominant interactions, and Vis-Autonomous interactions suggested that these types increased in frequency across publication dates (Table 1). The relative stability of Assertive types over time implies that storytelling has predominantly used relations between text and image where they both use narrative structures, but where meaning is balanced or weighted towards the text.



**Figure 4:** Changes in types of multimodal interactions across decade of publication date for American superhero comics. Solid lines depict Assertive relationships, dotted lines depict Dominant relationships, and double bar lines depict Autonomous relationships. Color and marker indicate modality (Co: purple/triangle; Verb: square/red; Vis: blue/circle). Error bars depict standard error.

Because our multimodal interactions could not inform us about how verbose the text may be, we also directly examined the number of words per panel. We found linear trends suggesting that the number of words decreased across decades (Table 1). As depicted in Figure 5, the average words per panel in the 2010s ended up only slightly less than in the 1940s. However, a large uptick appeared in words per panel from the 1950s to the 1980s, declining again in the 1990s.

Additional correlations compared the relationship between multimodal interactions with number of words per panel. We found positive correlations between words per panel and the rates that a book used Verb-Assertive,  $r(38) = 0.327$ ,  $p < 0.05$ , and Co-Assertive interactions,  $r(38) = 0.312$ ,  $p = 0.05$ , which suggested that these types were used more in books with greater numbers of words per panel. In contrast, we observed negative correlations for words per panel with the rates that books used Vis-Assertive,  $r(38) = -0.324$ ,  $p < 0.05$ , Vis-Dominant,  $r(38) = -0.531$ ,  $p < 0.001$ , and Vis-Autonomous interactions,  $r(38) =$



**Figure 5:** Number of words per panel used in American superhero comics across the decades. Error bars depict standard error.

$-0.554, p < 0.001$ . This suggested that the number of words per panel in a book decreased as the visual modality carried more meaning. Overall, these findings suggest that, as multimodal relationships conveyed more information through the visuals, the text became less verbose. Thus, the density of words in a panel correlates with the semantic weight of a multimodal interaction.

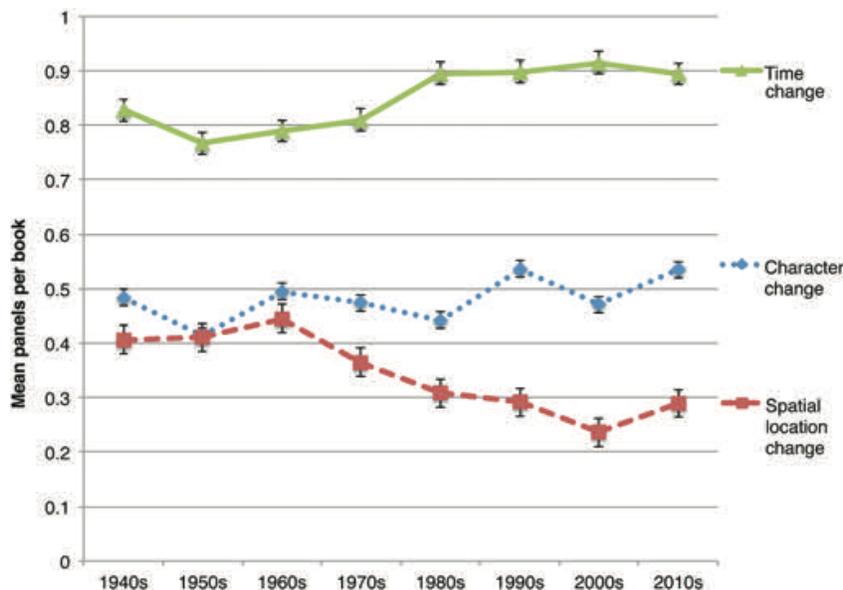
## Semantic relations

Analysis of semantic relations between panels (changes in characters, spatial location, and time) showed both a main effect of transition type,  $F(2,64) = 398.1, p < 0.001$ , and an interaction between transition and decade of publication,  $F(14,64) = 3.4, p < 0.001$ . As depicted in Figure 6, shifts in temporal states exceeded both other types of semantic changes, while character shifts were used more than those between spatial locations.

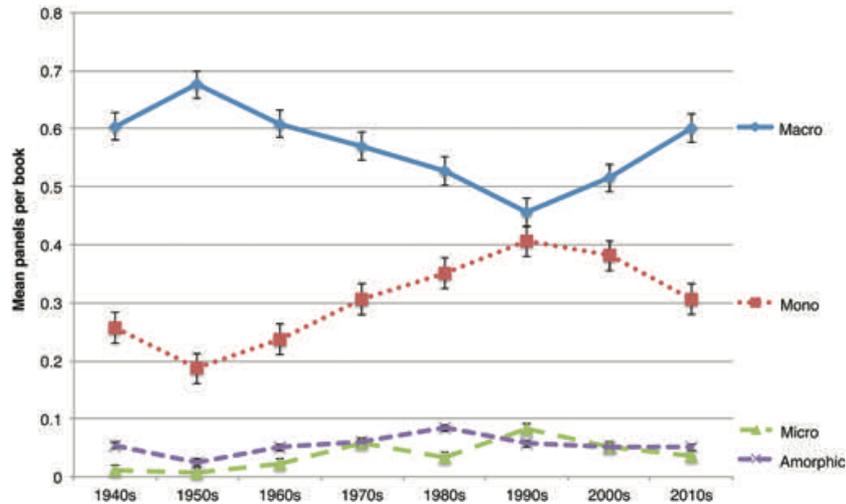
Further analysis revealed linear trends for both shifts between temporal states and spatial locations (Table 1), but not for changes between characters (all  $p > 0.178$ ). This implied that shifts in temporal states have grown more frequent with more recent publications, and changes in spatial location have become less frequent.

## Attentional framing structure

Our overall analysis of attentional framing (macros, monos, micros, amorphic panels) showed a main effect across all categories,  $F(3,96) = 237.3, p < 0.001$ , and an interaction between framing and decade of publication date,  $F(21,96) = 1.7, p < 0.05$ . These findings suggest both that attentional categories differed from each other, and that these differences changed with publication date over time. Consistent with previous studies of American comics (Cohn 2011; Cohn et al. 2012b), macros were used more than monos in all decades, with far fewer proportions of both micros and amorphic panels (Figure 7).



**Figure 6:** Different types of framing for panels used in American superhero comics across the decades. Error bars depict standard error.



**Figure 7:** Different types of framing for panels used in American superhero comics across the decades. Error bars depict standard error.

Polynomial contrasts revealed linear trends across decades for monos and micros (Table 1). No significant trends across decades were found for macros (all  $p$ s > 0.110) or amorphic panels (all  $p$ s > 0.502). These findings across decades help clarify disparities observed in previous studies of attentional framing types in American comics. Cohn (2011) found that American comics used a far greater proportion of macros than monos in a corpus with publication dates between the 1980s to the early 2000s. However, a follow up study found that American mainstream comics used only slightly more macros than monos, but with a corpus more focused within the first decade of the 2000s (Cohn et al. 2012b). These differences can now be attributed to publication dates of the analyzed comics, as reflected in our current data.

Though monos increased over time, and macros showed no significant decrease on their own across publication date, monos and macros displayed an inverse trend from each other (as implied by visual inspection of the data). That is, as monos increased, macros decreased. We therefore analyzed the *difference* between these types across decades (macros minus monos), where we observed significant linear trends (all  $F$ s > 3.7, all  $p$ s < 0.05, all  $R^2 > 0.131$ ), suggesting a tradeoff between macros and monos over time. Such a tradeoff is further suggested by a strong negative correlation between macros and monos in general,  $r(38) = -0.734$ ,  $p < 0.001$ , regardless of publication date.

## Discussion

This study examined multimodal interactions, attentional framing, and semantic relationships between panels across eight decades of American superhero comics. Overall, we found that multimodal interactions have increased the number of sequences with visual narrative structure, coinciding with fewer words per panel. In addition, panels have developed to depicting only a single character (or fewer), with more temporal shifts between panels, and fewer shifts between spatial locations. Altogether, these findings support the idea that American superhero comics have shifted away from conveying multimodal information through text accompanied by illustrative images, to those using more complex sequential visual narrative structures that coalesce as an equal partner with text.

## Changes in storytelling

First, the changes in multimodality, semantic relations, and framing of panels all suggest alterations to visual narrative structures (Cohn 2013a, 2015). Our findings imply that scenes are taking place more in one location (fewer spatial changes, more time changes) than switching between locations. However, the relative stability in character changes across books implies that panels with focal framing (monos, micros) remain more on individual characters who change in time, rather than switching between characters within a scene (Cohn 2015).

These observations align with ideas that storytelling in American superhero comics has become “decompressed” in the last two decades (Mazur and Danner 2014), with a narrative style that extends scenes across more panels, rather than “compressing” meaning into a more terse narrative sequence. By focusing on more narrative moments of a scene, it would necessitate more panels to convey an elaborated meaning. In fan and industry discourse, decompression has been described as a “cinematic” style of storytelling, in part because of association with “widescreen” layouts – a series of vertically stacked panels where panels extend the width of a page. Such an association is not unfounded: corpus work has shown that widescreen panels have become increasingly prominent since the mid-1980s (Pederson and Cohn 2016), which aligns with the onset of our observed increases in time changes.

“Decompression” is also involved with shifting meaning from text (and more words) to images. Simply, words can convey more information in less space than images. This also concurs with our finding that fewer panels appear per page across the decades, meaning that pages are literally becoming less compact (Pederson and Cohn 2016). Thus, more changes between temporal states and fewer changes in spatial location, along with more focus on the parts of scenes (monos, micros) and a shift toward more visual narrative structures, implies the “decompression” towards visual scenes becoming longer.

Decompression in American mainstream comics has largely been attributed to an influence of manga since the 1980s and 1990s (Mazur and Danner 2014; Cicci 2015). Starting in the late 1980s, manga began to be imported into the American comic market (Goldberg 2010; Brienza 2016), which escalated to wide popularity in the 1990s and 2000s. Such books have been widely recognized as being “influential” on American artists, particularly in the 1990s (McCloud 1996). However, manga have been observed to use fewer shifts between actions, and more shifts between characters and elements of the environment than American comics (Cohn et al. 2012b; McCloud 1993). In addition, manga use more monos than macros, and greater proportions of micros and amorphic panels overall (Cohn et al. 2012b).

Thus, if manga were to influence American comics on these dimensions, we would predict an increase in character changes and a decrease in time changes starting in the 1980s. Yet, we observe the opposite pattern: a steady proportion of character changes and an increase of time changes. In addition, the greatest change in attentional framing seemed to occur between the 1940s and the 1980s, and in semantic relations in the 1970s, *prior* to the mass importation of manga. Indeed, following the 1990s, these trends reversed slightly, perhaps signaling a backlash against this changing of the “American Visual Language” used in superhero comics (Cohn 2013a), be it perceived against manga or other factors. This reversal may also be consistent with the anecdotal disparagement of 1990s comics and their style by contemporary fans and critics, with framing types rebounding in the aftermath.

Thus, on its face, our data suggests that manga may not have influenced to American superhero comics’ storytelling as claimed. However, an alternative interpretation might be that manga’s influence has been selective. True influence of manga would increase tightly framed panels (monos, micros) along with character changes, which would necessitate concomitant changes in narrative and inferential machinery (Cohn 2015; Schwan and Ildirar 2010). Rather than significantly change the structures used to create and comprehend their visual narratives, American authors may have instead taken the surface observation of “drawn out scenes” and adapted it to the conventions of their own system. That is, rather than “drawing out scenes” by changing viewpoints between a scene’s component characters and environment (as in manga), American superhero comics “decompress” by adding more temporal states. Thus, American comics

do not directly borrow manga's narrative structures, but rather adapt their perceived end-state (longer scenes) in a way that is consistent with the existing machinery of American Visual Language narrative structure. Such misappropriation may be one reason for the critical or mixed reception of decompressed storytelling (e. g., Burgas 2007).

Another possibility is that changes in storytelling relate to the relative roles that pencilers and writers play in the creation of comics. American superhero comics are often created by a team of people each contributing to the singular composite multimodal product. Our observed changes in prominence for the visuals relative to the text may align with an interpretation that pencilers have exerted greater influence in the creative process over time. An alternative interpretation may be that writers and pencilers have perhaps worked more towards an integrative product. In creating older comics, pencilers sometimes would create the visuals of a comic independently (given varying degrees of direction from writers), after which a writer would add their own text – which may or may not have created a cohesive message with the visuals (Verano 2006; Petersen 2011). More recent cooperative methods may have resulted in writers and pencilers working towards a unified product, which better balances their contributions.

Nevertheless, attributing multimodal change solely to roles on a creative team remains complicated. The perceived value of pencilers versus writers has fluctuated over time (ex. pencilers exerted more influence in the 1990s, and writers more in the 2000s). Furthermore, writers and pencilers may play a greater or lesser role in determining what gets shown and how, depending on the creative team. Such a complex dynamic has occurred even since earlier American comics (Petersen 2011). Thus, attributing changes in storytelling to relative roles in a creative team alone seems insufficient.

It is also noteworthy that Hollywood movies have reduced the number of characters per shot over time, along with shorter shot duration, over the past 100 years (Cutting 2015; Cutting et al. 2010). Thus, there may have also been an influence related to film on the structure of the visual language used in comics. However, directionality of such influence – film to comics, comics to film, or mutually developing (Lacassin 1972) – is difficult to assess from existing data. This is especially true given that films often begin as drawn, static storyboards.

Formatting and technological reasons may have also influenced changes in the structure of these works (Verano 2006). Comic *books* began by retaining holdover formatting from comic strips of the earlier era, but shifted towards longer stories in the 1960s, facilitated by moves away from 64 page anthologies of short stories towards single issue full stories (Moore 2003). In addition, comics began being drawn on physically larger pages in the 1960s, thus allowing for content to “breathe” more on a page, i. e., take up more space with greater details (Moore 2003). The greater length would have facilitated “decompression” of meaning invested in the visuals rather than being compacted into a shorter space where the physically smaller words would have been more efficient. If page count allowed by formatting influences decompression, we might therefore predict further changes with contemporary movements towards book-formatted works, where authors have fewer limits on story length.

Within these contexts, another possibility is that particularly influential comics artists or comic books drove change over time. Our corpus does indeed include some widely accepted influential creators (e. g., Jack Kirby, Neal Adams), though we did not necessarily find evidence that they were outliers from their peers in these particular books and on these dimensions. (Their influence may be more measurable on other facets of this system not coded here.) In addition, our observed shifts across publication date appear fairly gradual from decade-to-decade, implying that there was not necessarily a sudden innovation, but that, rather, each generation built on their predecessors (be they standout influencers or others).

## Maturation of visual languages

Overall, these gradual changes in multimodal, narrative, semantic, and framing structures suggest an evolving maturity of the visual language used in American superhero comics over time. In essence, text provided a historical “bootstrapping” by which the visual language subsequently emerged with greater

complexity and autonomy. In the 1940s, the “Kirbyan” American Visual Language found in American superhero comics was still fairly new and developing, a growth perhaps facilitated by abandoning the terse formatting of strips of the earlier era (Verano 2006; Duncan and Smith 2009; Moore 2003). Carrying over from earlier formatting, authors may have relied more on the verbal text than the visuals to carry the semantic weight, as the graphic system was still developing. Indeed, our selection of books from the 1940s showed less than 5% of Vis-Dominant or Vis-Autonomous panels, and less than 10% of Vis-Assertive panels. This meant few sequences with visual narrative structure carried the semantic weight of the meaning. In addition, a system reliant on less space (strips or short stories) would, by necessity, need to show more information per unit, as evident in the greater proportion of macros overall. Together, this would make comics of this era seem more like “illustrated stories” whereby the text guided the meaning (and narrative), while the visuals remained illustrative and supplementary (and often redundant in content).

Over time, this relationship shifted. Across subsequent generations, the visual language progressed in complexity, emerging more as a primary means by which to communicate. Note that this process may have been cyclical: as each generation used a somewhat more complex visual narrative grammar with greater prominence within the multimodal system, it served as the stimulus for the next generation. That cohort then expanded on the previous one, providing yet even greater complexity for the subsequent generation. Such a process would help explain the gradual alterations from one decade to the next, even if influenced by outside visual languages (like manga), media (like film), or technology (like page formatting).

The end result of this process is a shift from using an “emerging” visual language reliant on the multimodal communicative power of text (in the 1940s), and towards incorporation of a “mature” visual language with greater conventionality and codified structures (albeit still changing – just like any other language). With a more developed system, it could freely convey information beyond redundancy or illustrations of the text, allowing for dually contributing modalities. A similar interpretation is hinted at by the changes in page layout across this same corpus: structures became more conventionalized and systematic across the eras (Pederson and Cohn 2016). Here we posit a similar interpretation to the narrative structure, as reflected in its presence in multimodal interactions, the semantic relations between panels, and the framing of panels themselves. Confirmation of such changes in the structures of the visual language would require additional study beyond our observations here.

These changes may also provide insights into how visual languages develop, beyond just this one population. This system has, overall, shifted from one dominated by text, with more simple structures (macro framing, short scenes, etc.), to one with more complex structures and autonomy for the visuals (mono/micro framing, longer scenes, more narrative structures). These trends imply that as verbal language is relied on less, the narrative structure takes on greater complexity. This is consistent with observations of emerging structure observed in gestural (Goldin-Meadow 2006; Goldin-Meadow et al. 1996) and other visual-graphic (Wilkins 1997/2016; Green 2014) communication systems when speech is reduced or absent. In this context, multimodal “bootstrapping” of the visual by the verbal could provide one model by which visual languages develop. Given that visual languages likely emerge as secondary to verbal expression within a multimodal system, we might expect them to play a reduced role with simpler structures until they are used across enough generations to develop more complexity. However, to confirm such an interpretation, we would need longitudinal analyses of other multimodal visual narrative systems, such as comics from other countries.

## Implications for cognition

Finally, changes in the structure of the visual language and its multimodal interactions with written language thereby imply changes in the cognitive structures that guide their production and understanding. In essence, readers and creators of one generation of comics engage with a different system than

subsequent generations: the visual language of 1950s superhero comics is different from that of the 2000s. Thus, the cognitive structures stored in memory that enable comprehension and production of such systems must also change. These data imply that this changing knowledge is not just monomodal, but also reflects knowledge of multimodal interactions.

In addition, such changes to readers' cognitive structures can be informative for characterizing how readers of one generation may perceive the structures used in works by other generations. For example, such change might motivate derogatory statements about the storytelling styles of other generations, like that certain artists/authors using a changed system are "less talented" or "less proficient in storytelling" – as in critiques of decompressed storytelling (Burgas 2007). Interpreted with regard to our data, such statements may reflect the discomfort or unfamiliarity experienced by one generation in understanding the structures of an evolved system. Effectively, such statements would be the multimodal and visual language version of statements like "kids today are ruining the language." Thus, such corpus analyses within cognitive framework provide a way to understand why readers may have preferences for different types of comics.

If readers and creators of different works bring to bear different structures on visual narrative understanding, this implies that comprehension of multimodal sequential images is not uniform across individuals. Rather, it is conditioned by individuals' reading habits. Just as differences in cognitive processing can be measured between readership of culturally different visual languages (Cohn and Kutas, Under Review), so, too, could we experimentally test whether changes in the structure of multimodality and/or visual narratives leads to changes in processing preferences across generations.

Altogether, our data provide initial evidence that multimodal storytelling has changed in American superhero comics over the past eight decades. Though balanced and text dominant (Co-/Verb-Assertive) multimodal interactions remain prevalent, the past 80 years have seen an increase in storytelling dominated by visual narrative structures, with decreasing focus on text carrying the weight of meaning. In essence, American comics have shifted from being a textual story with illustrations to a visual narrative that uses text.

**Acknowledgement:** Gerardo Soto-Becerra is thanked for additional assistance in data gathering.

## Appendix. Works analyzed

Our corpus analysis used the following books, listed chronologically by publication date.

1. Berold, B. and Eisner, W. (1940). *The Flame*. 3: 1–20. Fox Comics.
2. Kirby, J. and Wellman, M. W. (1941). *Captain Marvel Adventures*. 1: 1–31. Fawcett Comics.
3. Binder, J. (1942). *Captain Midnight*. 1: 1–34. Fawcett Comics.
4. Vagoda, B. and Weisbecker, C. (1943). *Black Hood*. 9: 1–23. MLJ Magazine
5. Nordling, K. (1949). *Lady Luck*. 86: 1–32. Quality Comics.
6. Quackenbush, B and Eisner, W. (1950). *Doll Man*. 30: 1–32. Comic Favorites, Inc.
7. Anderson, M and Siegel, J. (1951). *Lars of Mars*. 11: 1–16 & 28–31. Approved Comics.
8. Ferstadt, L, Fago, A., & Fox, V. (1955). *Blue Beetle*. 18: 1–25. Charlton Comics.
9. Cole, J. and Woolfolk, B. (1956). *Plastic Man*. 64: 1–31. Comic Magazines.
10. Plastino, A. and Bernstein, R. (1959). *Action Comics*. 252: 1–27. DC Comics.
11. Springer, F. and Kastle, H. (1962). *Brain Boy*. 2: 1–28. Dell Comics/Western Publishing.
12. Wood, W. and Lee, S. (1964). *Daredevil*. 5: 1–20. Marvel Comics.
13. Ditko, S. and Gill, J. (1965). *Captain Atom*. 78: 1–20. Charlton Comics.
14. Ditko, S. and Glanzman, D.C. (1967). *The Blue Beetle*. 3: 1–20. Charlton Comics.
15. Steranko, J. (1968). *Nick Fury Agent of S.H.I.E.L.D.* 1: 1–20. Marvel Comics.
16. Adams, N. and O'Neil, D. (1970). *Green Lantern*. 76: 1–23. DC Comics.

16. Novick, I. and Bates, C. (1973). *The Flash*. 211: 1–16. DC Comics.
17. Dillion, D. and Pasko, M. (1975). *Justice League of America*. 122: 1–18. DC Comics.
19. Tuska, G. and Mantlo, B. (1977). *The Invincible Iron Man*. 100: 1–19. Marvel Comics.
20. Rogers, M. and Englehart, S. (1978). *Detective Comics*. 475: 1–17. DC Comics.
21. Colan, G. and Mantlo, B. (1981). *The Avengers*. 210: 1–22. Marvel Comics.
22. Colan, G. and Thomas, R. (1982). *Wonder Woman*. 289: 1–26. DC Comics.
23. Byrne, J. (1984). *Fantastic Four*. 269: 1–22. DC Comics.
24. Bolland, B. and Moore, A. (1988). *Batman: The Killing Joke*. 1: 1–20. DC Comics.
25. Dwyer, K. and Gruenwald, M. (1989). *Captain America*. 358: 1–22. Marvel Comics.
26. McFarlane, T. and Michelinie, D. (1990). *The Amazing Spider-Man*. 328: 1–23. Marvel Comics.
27. Lyle, T. and Potts, C. (1993). *Venom: Funeral Pyre*. 2: 1–22. Marvel Comics.
28. Kubert, A. and David, P. (1997). *The Incredible Hulk*. 454: 1–25. Marvel Comics.
29. Yu, L. F. and Ellis, W. (1998). *Wolverine*. 121: 1–21. Marvel Comics.
30. Quesada, J. and Smith, K. (1999). *Daredevil*. 8: 1–22. Marvel Comics.
31. Dillon, S. and Ennis, G. (2001). *The Punisher*. 6: 1–22. Marvel Comics.
32. Sciver, E. V. and Johns, G. (2005). *Green Lantern: Rebirth*. 5: 1–21. DC Comics.
33. Garney, R. and Straczynski, J. M. (2007). *Amazing Spider-Man*. 539: 1–23. Marvel Comics.
34. Romita Jr., J. and Hudlin, R. (2008). *Black Panther*. 35: 1–22. Marvel Comics.
35. Medina, P. and Way, D. (2009). *Deadpool*. 11: 1–22. Marvel Comics.
35. Gleason, P. and Tomasi, P. (2010). *Green Lantern Corps*. 42: 1–24. DC Comics.
36. Finch, D. and Jenkins, P. (2011). *Batman The Dark Knight*. 2: 1–20. DC Comics.
38. Pacheco, C., Diaz, P., and Gillen, K. (2012). *Uncanny X-men*. 10: 1–20. Marvel Comics.
39. Guedes, R., Fawkes, R. & Lemire, J. (2013). *Constantine*. 1: 1–21. DC Comics.
40. Gerads, M. and Edmondson, N. (2014). *The Punisher*. 5: 1–20. Marvel Comics.

## References

- Abbott, M., and Forceville, C. (2011). Visual representation of emotion in manga: loss of control is loss of hands in *Azumanga Daioh* Volume 4. *Language and Literature*, 20:91–112.
- Bateman, J. A. (2014). *Text and Image: A Critical Introduction to the Visual/Verbal Divide*. New York: Routledge.
- Bateman, J. A., and Wildfeuer, J. (2014a). Defining units of analysis for the systematic analysis of comics: a discourse-based approach. *Studies in Comics*, 5:373–403.
- Bateman, J. A., and Wildfeuer, J. (2014b). A multimodal discourse theory of visual narrative. *Journal of Pragmatics*, 74:180–208.
- Brienza, C. (2016). *Manga in America: transnational book publishing and the domestication of Japanese comics*. London: Bloomsbury Academic.
- Burgas, G. (2007). Compressed storytelling versus decompressed storytelling: pros and cons [Online]. *Comic Book Resources*. Available: <http://www.cbr.com/compressed-storytelling-versus-decompressed-storytelling-pros-and-cons/> [Accessed 2/27/2017].
- Cicci, M. (2015). *Turning the page: Fandoms, multimodality, the transformation of the “comic book” superhero*. Doctoral Dissertation, Wayne State University.
- Cohn, N. (2011). A different kind of cultural frame: an analysis of panels in American comics and Japanese manga. *Image & Narrative*, 12:120–134.
- 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*, 37:413–452.
- Cohn, N. (2014). The architecture of visual narrative comprehension: the interaction of narrative structure and page layout in understanding comics. *Frontiers in Psychology*, 5:1–9.
- Cohn, N. (2015). Narrative conjunction’s junction function: the interface of narrative grammar and semantics in sequential images. *Journal of Pragmatics*, 88:105–132.
- Cohn, N. (2016). [A multimodal parallel architecture: A cognitive framework for multimodal interactions](#). *Cognition*, 146:304–323.
- Cohn, N., and Bender, P. (2017). Drawing the line between constituent structure and coherence relations in visual narratives. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, 43:289–301.

- Cohn, N., and Ehly, S. (2016). The vocabulary of manga: visual morphology in dialects of Japanese visual language. *Journal of Pragmatics*, 92:17–29.
- Cohn, N., Jackendoff, R., Holcomb, P. J., and Kuperberg, G. R. (2014). [The grammar of visual narrative: neural evidence for constituent structure in sequential image comprehension](#). *Neuropsychologia*, 64:63–70.
- Cohn, N., and Kutas, M. (2015). [Getting a cue before getting a clue: event-related potentials to inference in visual narrative comprehension](#). *Neuropsychologia*, 77:267–278.
- Cohn, N., and Kutas, M. (Under Review). What's your neural function, visual narrative conjunction? Grammar, meaning, and fluency in sequential image processing.
- Cohn, N., Paczynski, M., Jackendoff, R., Holcomb, P. J., and Kuperberg, G. R. (2012a). (Pea)nuts and bolts of visual narrative: structure and meaning in sequential image comprehension. *Cognitive Psychology*, 65:1–38.
- Cohn, N., Taylor-Weiner, A., and Grossman, S. (2012b). Framing attention in Japanese and American comics: cross-cultural differences in attentional structure. *Frontiers in Psychology – Cultural Psychology*, 3:1–12.
- Coogan, P. (2006). *Superhero: The Secret Origin of a Genre*. Austin, TX: MonkeyBrain.
- Cutting, J. E. (2015). The framing of characters in popular movies. *Art & Perception*, 3:191–212.
- Cutting, J. E., DeLong, J. E., and Nothelfer, C. E. (2010). [Attention and the evolution of hollywood film](#). *Psychological Science*, 21:432–439.
- Duncan, R., and Smith, M. J. (2009). *The Power of Comics*. New York: Continuum Books.
- Duncan, R., Smith, M. J., and Levitz, P. (2015). *The Power of Comics*. New York: Continuum Books.
- Forceville, C. (2011). Pictorial runes in *Tintin and the Picaros*. *Journal of Pragmatics*, 43:875–890.
- Forceville, C. (2016). Conceptual metaphor theory, blending theory, and other cognitivist perspectives on comics . In: *The Visual Narrative Reader*, N. Cohn (Ed.). Bloomsbury: London.
- Forceville, C., and Urios-Aparisi, E. (2009). *Multimodal Metaphor*. New York: Mouton De Gruyter.
- Forceville, C., Veale, T., and Feytaerts, K. (2010). Balloonics: The visuals of balloons in comics. In: *The Rise and Reason of Comics and Graphic Literature: Critical Essays on the Form*, J. Goggin and D. Hassler-Forest (Eds.). Jefferson: McFarland & Company, Inc.
- Foulsham, T., Wybrow, D., and Cohn, N. (2016). Reading without words: eye movements in the comprehension of comic strips. *Applied Cognitive Psychology*, 30:566–579.
- Gernsbacher, M. A., Varner, K. R., and Faust, M. (1990). Investigating differences in general comprehension skill. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 16:430–445.
- Goldberg, W. (2010). The manga phenomenon in America. In: *Manga: An Anthology of Global and Cultural Perspectives*, T. Johnson-Woods (Ed.). New York: Continuum Books.
- Goldin-Meadow, S. (2006). Talking and thinking with our hands. *Current Directions in Psychological Science*, 15:34–39.
- Goldin-Meadow, S., McNeill, D., and Singleton, J. (1996). [Silence is liberating: removing the handcuffs on grammatical expression in the manual modality](#). *Psychological Review*, 103:34–55.
- Green, J. (2014). *Drawn from the Ground: Sound, Sign and Inscription in Central Australian Sand Stories*. Cambridge, UK: Cambridge University Press.
- Guynes, S. A. (2014). Four-color sound: a peircean semiotics of comic book onomatopoeia. *The Public Journal of Semiotics*, 6:58–72.
- Hatfield, C., Heer, J., and Worcester, K. (2013). *The Superhero Reader*. Jackson: University Press of Mississippi.
- Kaan, E. (2007). Event-related potentials and language processing: a brief overview. *Language and Linguistics Compass*, 1:571–591.
- Lacassin, F. (1972). The comic strip and film language. *Film Quarterly*, 26:11–23.
- Magliano, J. P., Larson, A. M., Higgs, K., and 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.
- Magliano, J. P., Miller, J., and Zwaan, R. A. (2001). Indexing space and time in film understanding. *Applied Cognitive Psychology*, 15:533–545.
- Magliano, J. P., and Zacks, J. M. (2011). The impact of continuity editing in narrative film on event segmentation. *Cognitive Science*, 35:1489–1517.
- Martinec, R., and Salway, A. (2005). A system for image–text relations in new (and old) media. *Visual Communication*, 4:337–371.
- Mazur, D., and Danner, A. (2014). *Comics: A Global History, 1968 to the Present*. London: Thames & Hudson.
- McCloud, S. (1993). *Understanding Comics: The Invisible Art*. New York, NY: Harper Collins.
- McCloud, S. 1996. *Understanding manga*. *Wizard Magazine*, 56, 44–48.
- Moore, S. (2003). In the old days, it woulda been eight pages. In: *A Thousand Flowers: Comics, Pop Culture and the World Outside*, Brady, M. (Ed.).
- Neff, W. A. (1977). *The Pictorial and Linguistic Features of Comic Book Formulas*. Doctoral Dissertation, University of Denver.
- Osaka, M., Yaoi, K., Minamoto, T., and Osaka, N. (2014). Serial changes of humor comprehension for four-frame comic Manga: An fMRI study. *Scientific Reports*:4.
- Painter, C., Martin, J. R., and Unsworth, L. (2012). *Reading Visual Narratives: Image Analysis of Children's Picture Books*. London: Equi-nox.

- Pederson, K., and Cohn, N. (2016). The changing pages of comics: Page layouts across eight decades of American superhero comics. *Studies in Comics*, 7:7–28.
- Petersen, R. S. (2011). *Comics, Manga, and Graphic Novels: A History of Graphic Narratives*. Santa Barbara, CA: ABC-CLIO.
- Pratha, N. K., Avunjian, N., and Cohn, N. (2016). Pow, punch, pika, and chu: The structure of sound effects in genres of American comics and Japanese manga. *Multimodal Communication*, 5:93–109.
- Royce, T. D. (2007). Intersemiotic complementarity: a framework for multimodal discourse analysis. In: *New Directions in the Analysis of Multimodal Discourse*, T. D. Royce, and W. L. Bowcher (Eds.), 63–109. Mahwah, NJ: Lawrence Erlbaum Associates.
- Schwan, S., and Ildirar, S. (2010). [Watching film for the first time: how adult viewers interpret perceptual discontinuities in film](#). *Psychological Science*, 21:970–976.
- 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: *The Visual Narrative Reader*, N. Cohn (Ed.). London: Bloomsbury.
- Tasić, M., and Stamenković, D. (2015). The interplay of words and images in expressing multimodal metaphors in comics. *Procedia – Social and Behavioral Sciences*, 212:117–122.
- Verano, F. (2006). Spectacular consumption: visuality, production, and the consumption of the comics page. *International Journal of Comic Art*, 8:378–387.
- Wilkins, D. P. (1997/2016). Alternative representations of space: arrernte narratives in sand. In: *The Visual Narrative Reader*, N. Cohn (Ed.). London: Bloomsbury.
- Zwaan, R. A., and Radvansky, G. A. (1998). Situation models in language comprehension and memory. *Psychological Bulletin*, 123:162–185.