Visual Language Theory and the Scientific Study of Comics

Neil Cohn

Tilburg University

Email: neilcohn@visuallanguagelab.com

www.visuallanguagelab.com

This is an uncorrected pre-print version of this paper.
Please consult the final version for all citations:

Cohn, Neil. 2018. Visual Language Theory and the scientific study of comics. In Wildfeuer, Janina, Alexander Dunst, Jochen Laubrock (Ed.). *Empirical Comics Research: Digital, Multimodal, and Cognitive Methods*. (pp. 305-328) London: Routledge.

Abstract

The past decades have seen the rapid growth of empirical and experimental research on comics and visual narratives. In seeking to understand the cognition of how comics communicate, Visual Language Theory (VLT) argues that the structure of (sequential) images is analogous to that of verbal language, and that these ***visual languages*** are structured and processed in similar ways to other linguistic forms. While these visual languages appear prominently in comics of the world, all aspects of graphic and drawn information fall under this broad paradigm, including diverse contexts like emoji, Australian aboriginal sand drawings, instruction manuals, and cave paintings. In addition, VLT's methods draw from that of the cognitive and language sciences. Specifically, theoretical modeling has been balanced with corpus analysis and psychological experimentation using both behavioral and neurocognitive measures. This paper will provide an overview of the assumptions and basic structures of visual language, grounded in the growing corpus and experimental literature. It will cover the nature of visual lexical items, the narrative grammar of sequential images, and the compositional structure of page layouts. Throughout, VLT emphasizes that these components operate as parallel yet interfacing structures, which manifest in varying 'visual languages' of the world that temper a comprehender's fluency for such structures. Altogether, this review will highlight the effectiveness of VLT as a model for the scientific study of how graphic information communicates.

1 Introduction

The ability to draw is a basic human ability, and sequential images are prevalent across human history and cultures, in contexts from cave paintings and wall tapestries to sand drawings and comics. This fundamental aspect of human expression yields several important questions, such as how are such individual and sequential images structured, and how do people process and comprehend them? ***Visual Language Theory*** (VLT) offers a framework to answer these questions with the principles and methods of the linguistic and cognitive sciences (Cohn 2013b). This chapter reviews the basic assumptions and constructs of VLT, emphasizing its multi-pronged methods of research.

1.1 What is language?

A language constitutes a set of patterns in a person's mind/brain. To the extent that the patterns in one person's brain (their idiolect) align with other people's, they share a common 'language'. Because people are separated by regional, cultural, and other factors, such patterns may differ between groups of people: there is no universal language, but rather diverse languages manifest across the world. This basic idea holds both for spoken and signed languages, but also for ***visual languages***—the systems guiding individuals' drawing abilities.

Not all drawings constitute visual languages, though. A language requires an interaction between three primary components (Cohn 2013b): *meaning*, *modality*, and *grammar*. First, it combines the mapping of ***meaning*** to a ***modality***. Spoken languages express meaning using the modality of phonology (sound), while visual languages use graphic structure (drawn lines). Systematic mappings between a modality and meanings create a stored ***lexicon***. However, meaningful expressions alone—systematic or unsystematic—are not enough to become a language. Rather, those expressions must be ordered using a ***grammar***. In visual languages, the grammar organizes meaningful images into coherent sequences. Only full languages use this combination of three structures: a modality, which expresses meaning manifested into a lexicon, and a grammar, which orders those meanings into sequences.

These structures mutually interface in a ***parallel architecture*** (Cohn 2016, Jackendoff 2002). That is, the structure of visual languages used in comics does not cascade hierarchically down from a page to a panel to component parts in a singularly divisible structure. Rather, independent components mutually interact to form the perception of a holistic experience. This becomes overt in the mutual contributions of a visual language with written language, which coalesce into a broader multimodal whole (Cohn 2016).

1.2 Comics, visual language, and methodologies

For our purposes, visual languages appear prevalently in what are broadly labeled as 'comics' around the world—though certainly not exclusively. Visual languages arise in several different cultural contexts. Thus, it is important to emphasize the separation between visual language and 'comics'. Comics themselves are not defined as a (visual) language. Rather, just as novels or magazines may be *written in* a written language, comics are drawn in a visual language (often in combination with a written language). It would be unusual to say that 'novels are English', and this is equally true of the misnomer that 'comics are a visual language.' Similarly, the notion of a

'comics medium' simply describes the aggregate use of visual and written languages, in the absence of recognizing this split between the context and its language(s).

Moreover, different types of comics may be characteristic of different visual languages. Japanese Visual Language (JVL) is highly associated with manga, while 'Kirbyan' American Visual Language characterizes superhero comics from the United States (Cohn 2013b)—i.e., the 'drawing and storytelling styles' most associated with those contexts. However, manga and superhero comics also are not expressly defined by their stereotypical visual languages (i.e., manga ≠ JVL). Ultimately, the definition of 'comics' rests with a web of sociocultural ideas, including their cultural context, genres, and possibly the different visual languages they use.

Given this separation, scholars should question whether they are interested in visual language (the structure and processing of sequential images) and/or their multimodal interaction with written language, or whether they are interested in comics (the socio-cultural context that most often uses those languages). Questions related to the latter typically fall within the humanities (literary studies, comics studies, art history, etc.), while questions about the former belong to the cognitive sciences (linguistics, psychology, computer science, etc.). This division reflects the multiple ways that any other language is studied. Research on language pervades across disciplines, motivated by the specific types of questions being asked.

Research on visual language can thus be analyzed using various methods from the linguistic and cognitive sciences, again depending on the questions being asked. In line with linguistics, ***theoretical modeling*** can characterize the internal structures of a system (ex. what are the structures characterizing page layouts?). ***Corpus analyses*** investigate how structures manifest in actual works through annotation (ex. how might page layouts differ across cultures?). Finally, ***experimentation*** can be used to study the psychological instantiation of such structures, and how they are processed (ex. how do people navigate through page layouts?).

While each method offers insights into aspects of visual language, ultimately, they must combine for a full picture. Indeed, certain pitfalls face research confined to one methodology alone. For example, theoretical research that does not consider experimental results may lack psychological validity, or without corpus data, theories may not reflect phenomena found in actual visual languages. Also, claiming 'empirical testability' of a theory does grant it validity without providing actual evidence. Meanwhile, both experimentation and corpus studies in the absence of theory may produce *ad hoc* analyses of fairly superficial constructs. Simply having a tool of analysis (e.g., an annotator or an eye-tracker) does not guarantee insight without a theoretical framework to guide what is being investigated. While a research program may begin with any one method, ultimately, all three should feed into each other, and all are required to understand how a broader system works.

VLT incorporates all three methods into a comprehensive analysis of visual language structure and comprehension. Given this, we now review basic aspects of visual languages: their visual vocabulary, narrative grammar, and external compositional structure. In each, we will review their basic theoretical structure, their cross-cultural variation, and their processing.

2 Visual vocabulary

2.1 Theory

The vocabulary of visual languages arises in how people draw. A ***lexical item*** is a mapping between form and meaning stored in the long-term memory of a user of a language. Spoken languages map phonology to meaning, codified as words (*dog*, *love*), morphemes (*un-*, *-ity*), phrase level constructions (*What's X doing Y?*), or even idioms (*kick the bucket* to mean DIE). In visual languages, patterned mappings arise between graphic structure (lines and shapes) and meaning. Visual lexical items may be component parts of an image (how someone draws a hand or face), parts that combine with others (a motion line and a mover, or a lightbulb above the head), patterned whole images, or even sequencing patterns (Cohn 2013b).

Lexical items divide into two broad categories. It is easy to create new ***open-class*** items, which in visual form are typically iconic, since it is easy to invent new ways to draw animals, locations, objects, etc. It is harder to invent novel ***closed-class*** items. These often constitute the 'symbology' in comics, such as motion lines to show movement, lightbulbs above the head for inspiration, or hearts replacing eyes for lust. Such forms are highly conventionalized, and thus it is harder to invent novel closed-class items.

Closed-class visual lexical items are often interesting because of their combinatorial properties. *Bound morphemes* cannot stand alone, and must combine with a *stem* (Cohn 2013b, Forceville 2011). For example, a speech balloon must connect to a speaker, just as lightbulbs or stars which sometimes appear above characters' heads cannot float independently with the same meaning.

Visual languages use similar strategies for combinations as other languages. Affixes attach a bound morpheme before (prefix), after (suffix), inside (infix), or around (circumfix) a stem. Visual affixation can cover a range of spatial locations, but 'upfixes' (affixes that are up) are a class of elements that hover above characters' heads (like stars, hearts, or lightbulbs). Speech balloons also affix to a speaker, while motion lines attach to moving objects, among many others. Morphemes can also be substituted entirely via suppletion (*go* → *went*) or internally in umlaut (*sing* → *sang*). Graphically, substitution occurs when people fight in a cloud with arms and legs sticking out, or when dotted lines show invisibility. 'Eye-umlauts' replace eyes for hearts (lust), spirals (dizziness), dollar signs (desire for money), etc. Finally, reduplication is used to repeat forms (*salad-salad*; *tick-tock*). This occurs visually when all or part of a figure takes different postures in the same image to indicate movement, or when lines repeat with an offset to depict shaking or the double-vision of another character.

Visual morphology derives meaning through these combinations in different ways. Sometimes, a closed-class morpheme itself conveys the meaning, like hearts which mean LOVE no matter whether placed above the head or substituted for eyes. In other cases, context matters, such as stars, which have no intrinsic sense, and mean different things when above the head (dizzy, pain), substituted for eyes (desire for fame), or floating near a body part (pain). Other morphemes depict basic conceptual understandings, like motion lines as depictions of movement paths, or dotted lines depicting the folk understanding of a path of vision (Cohn 2013b).

Visual morphology might also invoke deeper conceptual metaphors (Lakoff and Johnson 1980, Forceville 2016) where one domain maps to another. For example, an upfix for gears above the head to mean thinking involves the metaphor that the MIND IS A MACHINE (Lakoff and Johnson 1980), while a steam upfix (or steam coming out of ears) invokes the idea that ANGER IS HOT FLUID IN A PRESSURIZED CONTAINER (Forceville 2005). Metaphoric understanding is a basic aspect of human languages across domains, and it subtly manifests across many uses of visual morphology (Forceville 2005, Shinohara and Matsunaka 2009).

2.2 Cross-cultural variation

Cross-cultural variation in visual morphology is often recognizable—unless you know that bubbles from the nose mean sleepiness in Japanese Visual Language, it may seem unusual. Nevertheless, only a few empirical studies have annotated visual morphology. Case studies have examined the range of morphemes in specific comics (Forceville 2011) often using metaphorical reference (Abbott and Forceville 2011, Forceville 2005), while others use a small corpus to compare certain constructs (Forceville, Veale, and Feyaerts 2010, Shinohara and Matsunaka 2009)

Recent work has also used visual morphology to characterize consistency and variation within visual languages. Cohn and Ehly (2016) compared closed-class morphemes in Japanese shonen (boys') manga and shojo (girls') manga. Overall, most morphemes appeared in both genres, suggesting a generalized Japanese Visual Language across both. However, shonen manga used more morphemes related to action (zoom lines), anger (pointed teeth, giant yelling head), and moodiness (gloom lines), while shojo manga used more morphemes related to embarrassment (blush), overwhelming emotion (chibi superdeformation—spontaneous transformation to a hyper cartoony style), and backgrounds related to emotion (flower petals, sparkles). These proportional differences suggested that the systems used in shonen and shojo may constitute 'dialects' within a broader shared JVL.

Visual vocabularies have also been detailed in visual languages outside of comics, such as those in Australian sand drawings (Green 2014, Wilkins 1997/2016), or historical works like Mayan pottery (Wichmann and Nielsen 2016) or the Bayeux Tapestry (Díaz Vera 2013). Such research usefully characterizes diversity across systems, and hints at regularities of various conventions, like depicting paths (motion lines), speech, and some metaphors. This work thus foreshadows more extensive typological research within and across cultures.

2.3 Processing

The processing of visual morphology has mostly focused on a few morphemes like carriers (speech balloons and thought bubbles), motion lines, and upfixes. For example, the omission of motion lines reduces the comprehensibility of depicted actions (Cohn and Maher 2015, Ito, Seno, and Yamanaka 2010), and aids in the perception that objects are moving (e.g., Kawabe and Miura 2006). This understanding goes beyond simple perceptual processing (Ito, Seno, and Yamanaka 2010), and is modulated by people's frequency reading comics (Cohn and Maher 2015). Upfixes meanwhile are constrained to appear above characters' heads and to 'agree' with characters' facial expressions (Cohn, Murthy, and Foulsham 2016). Since violation of these constraints is considered

worse in both conventional *and* novel upfixes, it suggests that they involve an abstract schema, not just individually memorized tokens.

Understanding of visual morphemes is modulated by age (Nakazawa 2016), frequency of reading comics (Cohn and Maher 2015, Cohn, Murthy, and Foulsham 2016), and the frequency that those morphemes appear in comics (Newton 1985). Nevertheless, this varies. Speech balloons and thought bubbles are understood by around 4 years old (Wellman, Hollander, and Schult 1996), but jagged lines to mean loudness are understood by even preliterate children (Yannicopoulou 2004). Understanding the meaning of motion lines remains poor around age 6 but proficient by age 12 (Friedman and Stevenson 1975, Gross et al. 1991, Carello, Rosenblum, and Grosofsky 1986, Mori 1995, Nakazawa 2016), accompanying a shift from interpreting them as physical elements (like wind) to symbolic (Gross et al. 1991). This differs from repeating body parts to show movement, which even young children understand (Friedman and Stevenson 1975).

3 Narrative structure

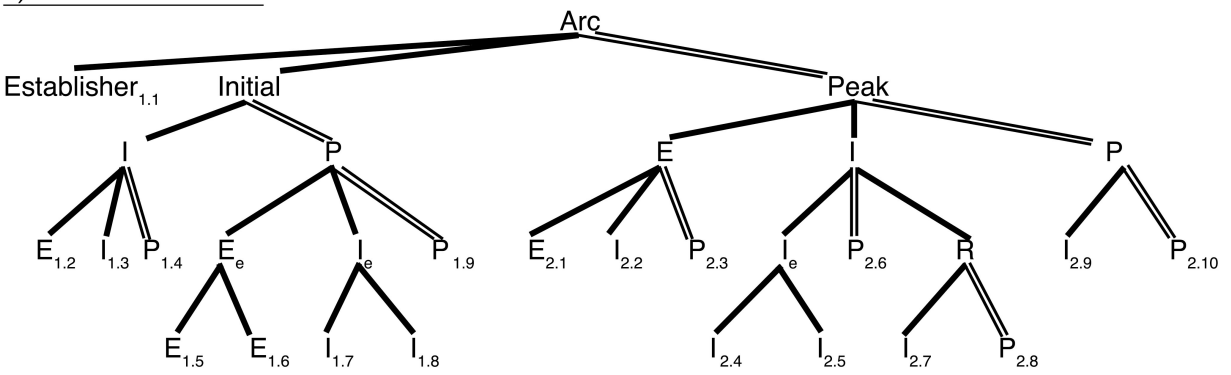
3.1 Theory

Beyond individual images and their parts, readers of comics must also comprehend *sequences* of images. Early theories emphasized the meaningful changes between juxtaposed images, as in "panel transitions" between referential, spatial, and temporal information between panels (McCloud 1993), which have subsequently been operationalized (Saraceni 2016, Stainbrook 2016) and expanded on (Bateman and Wildfeuer 2014). However, comprehension of sequential images cannot rely on meaningful connections alone (Cohn 2010, 2013c). ***Visual Narrative Grammar*** (VNG) argues that the semantic information in panels maps to categorical roles within a constituent structure, similar to how words play grammatical roles within constituents in syntax (Cohn 2013c).

A sequence canonically starts with an *Establisher (E)*, which passively sets up the entities or situation. An *Initial (I)* then marks the start of a sequence's events, prototypically a preparatory action. The sequence climaxes in the *Peak (P)*, where events manifest and/or come together, followed possibly by a *Release (R)*, which dissolves the interaction's tension through an aftermath or resolution. These basic narrative categories proceed in this order in the canonical narrative schema (Cohn 2013c, 2014b), possibly supplemented by elaborative categories (Cohn 2013b). Narrative categories are defined by an interaction between their bottom-up semantic content (e.g., Initials are typically preparations), and by their top-down context in a sequence (e.g., Initials canonically follow Establishers and precede Peaks) (Cohn 2013c, 2014b).

Narrative categories apply beyond just individual panels, and can also characterize groupings of panels. Across the two pages in Figure 1, several small narrative phases use parts of a canonical sequence, which build to progressively higher levels. The opening Establisher sets up the entire sequence (the samurai rabbit, Usagi, defending the panda lord from attacking ninja). Various sub-sequences of actions then successively set-up (Initials) and climax (Peaks) in the actions of the fight scene. For example, the second page constitutes the entire main Peak clause. This opens with Usagi getting dust thrown in his eyes (Establisher), who then jumps to avoid getting slashed and falls down (Initial), only to get saved by the young lord, who stabs the main ninja (Peak). Each of these subsequences gives the overall sequence a richer structure.

a) Narrative Structure



b) External Compositional Structure

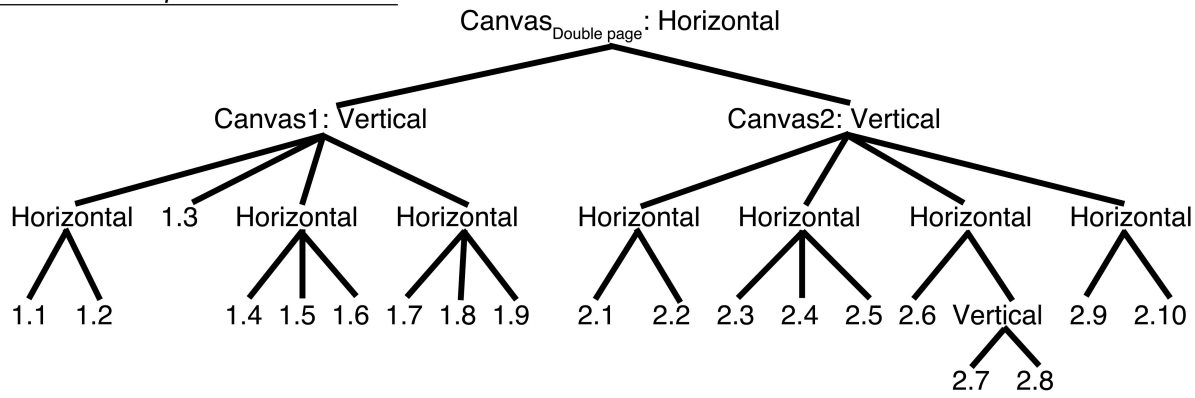
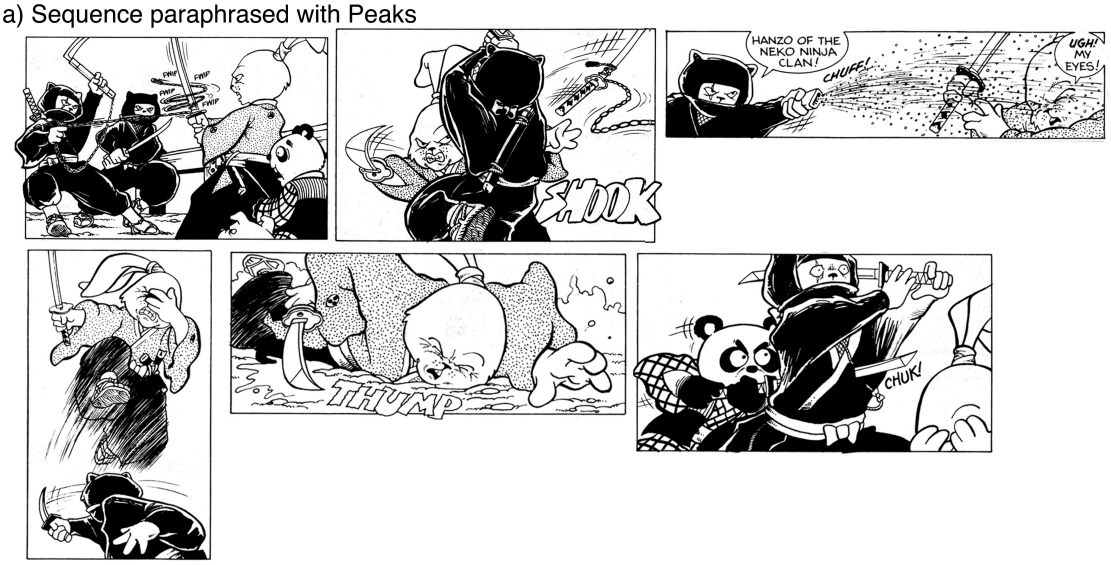


Figure 1. Two pages analyzed for their narrative structure and external compositional structure. Usagi Yojimbo™ © 1987 Stan Sakai. Courtesy of Fantagraphics Books (www.fantagraphics.com).

The hierarchic aspects of a sequence can be illustrated by paraphrasing. Peak panels motivate their constituents (notated by double-bar lines), meaning that other panels expand on the Peak, which drives the narrative. Because of this, sequences can be paraphrased by dropping out non-Peak panels. Figure 2a paraphrases Figure 1 with only its Peak panels. Here, these panels no longer act as Peaks, but 'move up' into roles higher in the structure. The same significant narrative moments maintain as the broader sequence, albeit terser. This coherence is not achieved when paraphrasing with only Initials, as in Figure 2b. This illustrates that Initials have a supportive relationship for the primary Peak panels.

a) Sequence paraphrased with Peaks



b) Sequence paraphrased with Initials

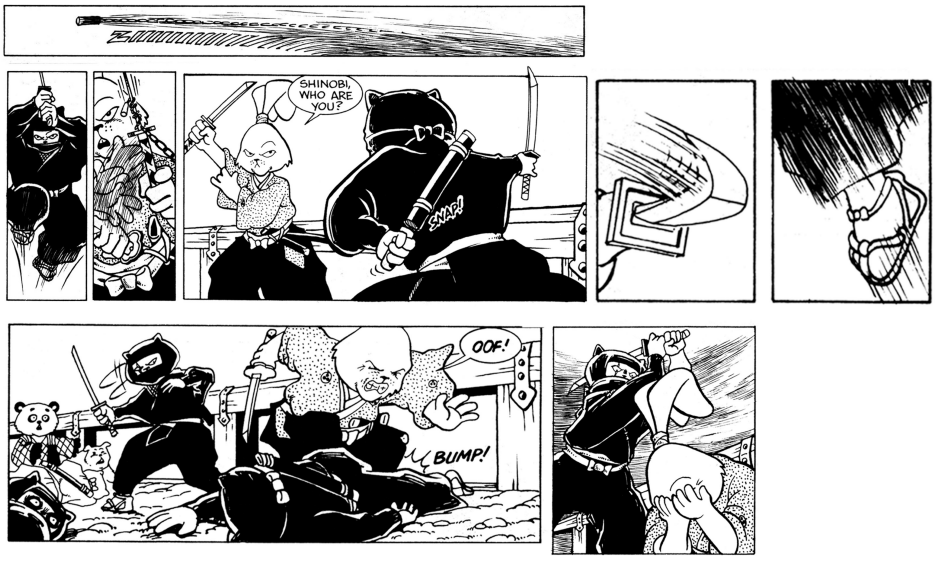


Figure 2. Sequences created through 'paraphrasing' the narrative of the pages in Figure 1, by a) dropping out all but Peak panels, and b) dropping out all by Initial panels. Usagi Yojimbo™ © 1987 Stan Sakai. Courtesy of Fantagraphics Books (www.fantagraphics.com).

Modifiers can also elaborate the canonical schema. Consider panels 1.5 and 1.6 (also 1.7/1.8 and 2.4/2.5) in Figure 1, where each panel contains a different character in the same narrative state—a single panel could replace these pairs showing the whole scene. This framing focuses on each character individually, but demands an inference that they belong to the same spatial location, since it is not overtly depicted. Because both panels play the same narrative role, they are *conjoined* within a constituent of that category. This particular construction is an ***E(nvironmental)-Conjunction***, meaning a *conjunction* (repetition of categories in a constituent of that type) that indexes a semantically inferred *environment*. The subscript "e" notates this mapping, which suggests connections to a semantic/spatial structure (Cohn 2015).

E-Conjunction is only one type of semantic interface for narrative conjunction. Other conjunctions involve aspects of characters, actions, and semantic associative fields (Cohn 2015). All repeat a common narrative category (conjunction) but with different semantic mappings. Other narrative modifiers zoom onto information in another panel (Refiners) or show the same information in another panel but from a different viewpoint (Perspective Shift).

The hierarchy of narrative structure extends to ever larger structures. An 'Arc' reflects a maximal node that plays no role in a larger sequence. While this example illustrates just two pages as an 'Arc', they may also play a role in a larger story. The narrative structures can thus recursively build to the level of whole plotlines.

3.2 Cross-cultural variation

Though basic constructs of VNG may pervade across cultures, their manifestations may differ. McCloud (1993) first observed that American and European comics differed in their panel-to-panel semantic changes from Japanese manga with regard to transitions between actions, characters, and environmental information. Subsequent analyses demonstrated that manga use more panels showing individual characters (monos), zoomed in panels (micros), and environmental information (amorphics) than American comics of both mainstream and indy genres (Cohn 2013b), which generally depicted whole scenes (macros).

Recent research has targeted specific narrative constructs in VNG. For example, the 'basic narrative progression' characterizes a series of panels with multiple interacting characters (macros) with a narrative change between each panel. Corpus analysis of over 200 books (~24,000 panels) suggests that this pattern occurs more in European and older American comics, but less in recent American comics, and much less in Asian books (Cohn in press). This is contrasted by findings of Environmental-Conjunction, which occurs more in Asian books than those from Europe or America. This proportional difference also appears to be a tradeoff: the more a system uses E-Conjunction, the less it uses the basic narrative progression. As illustrated in Figure 3, such analysis not only implies that different narrative grammars are being used by different visual languages, but it also hints at regularities across systems for how narrative patterns interact.

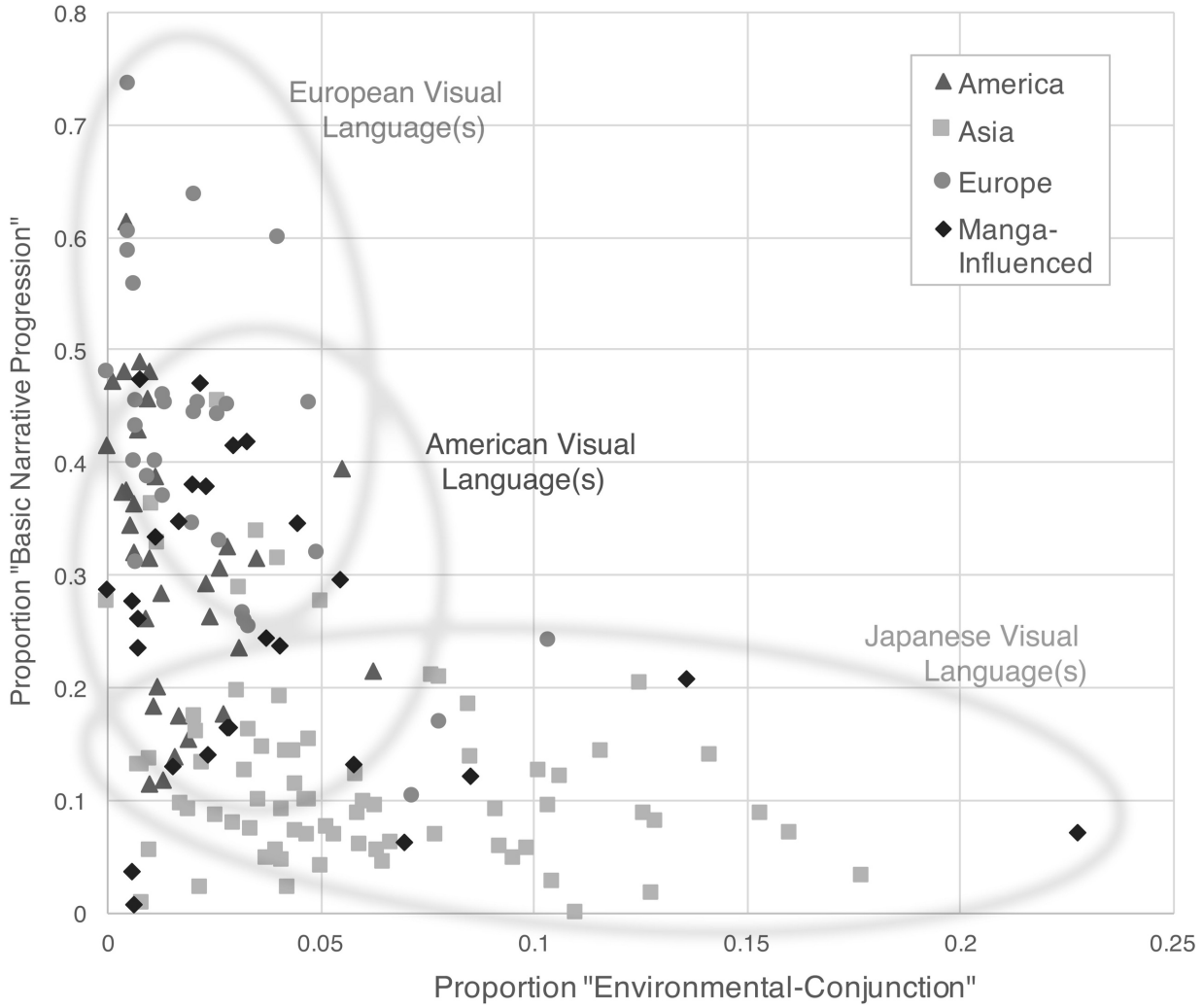


Figure 3. Scatterplot of different comics' frequency using the basic narrative progression and Environmental-Conjunction. 'Manga-influenced' works are those imitative of manga, but created by English speakers outside Japan.

3.3 Processing

The growing experimental literature has contributed to an emerging understanding of the processing of sequential images. Research measuring the electrical activity of the brain has confirmed that comprehension of visual sequences involves *both* semantic and grammatical processing streams (Cohn et al. 2014, Cohn et al. 2012). Different brain responses are evoked by the manipulation of semantics and narrative grammar in sequential images (Cohn et al. 2014, Cohn et al. 2012), and these neural responses resemble those observed in the processing of semantics and syntax in language (Kaan 2007).

A comprehender first ***accesses*** the relevant semantic information of an image. Semantic retrieval is focused on the morpho-semantic information that may be relevant to the sequence (Foulsham, Wybrow, and Cohn 2016, Cohn and Maher 2015, Cohn and Paczynski 2013). This access demands the greatest cost when a sequence begins, as no context has established a scene, as evident from

slower reading times (Cohn and Wittenberg 2015) and larger brainwave responses related to semantic processing (Cohn et al. 2012, West and Holcomb 2002)—i.e., the "N400 effect" (Kutas and Federmeier 2011). Discontinuity generates costs for retrieving basic semantic information, but in a coherent narrative, semantic processing is attenuated across each panel of a sequence (Cohn et al. 2012). In addition, eye-movements are more focal to content in panels within coherent narratives than scrambled ones (Foulsham, Wybrow, and Cohn 2016). That is, a sequence's meaning—and the content that cues it—gets easier to access as one progresses from image to image in a coherent narrative.

Readers use this basic semantic information to construct a mental model of the scene and its events (Zwaan and Radvansky 1998). A default expectation of *continuity* persists throughout a sequence: the elements in one panel should appear in subsequent panels. Discontinuity creates costs for ***updating*** this mental model when faced with shifts in referential (characters, objects) or spatial information (Magliano and Zacks 2011, Cohn and Kutas 2017, 2015), or inferential processing for omitted events (Cohn and Kutas 2015, Cohn and Wittenberg 2015, Magliano et al. 2015). Like the access of basic semantic information, this updating process is ongoing throughout a sequence (Cohn and Kutas 2015).

The morphological cues in panels provide bottom-up information about expected narrative categories, tempered by the top-down context of a narrative schema. For example, a panel depicting a person reaching back their arm to punch will be mapped to a narrative Initial on the basis of this prototypical event structure. Identification of narrative categories then allows for ***predictions*** of upcoming structure based on positioning within a narrative schema. Thus, if an Initial is identified, a subsequent Peak will be predicted, because Peaks follow Initials in the canonical narrative schema. This prediction can be confirmed if a subsequent image fulfills that expectation as a Peak. If it is not, an additional updating process may be required on the basis of this *structural* discontinuity (Cohn 2013b, Cohn and Kutas 2015). Further combinatorial processing guides the construction of hierarchic constituents and grammatical constructs, independent of the semantic updating processes (Cohn et al. 2014, Cohn and Kutas 2017).

These mechanisms of *access*, *prediction*, and *updating* occur for both the semantic and grammatical streams as ongoing and iterative operations throughout the unit-by-unit processing of sequential images. These processes are also modulated by fluency in a visual language. Several studies show that frequency of comic reading correlates with both behavioral and neurocognitive measures of processing sequential images (Cohn and Kutas 2015, Cohn and Maher 2015, Cohn et al. 2012). Indeed, fluency appears necessary for even basic processing, such as recognizing that characters are the same referential entities across images (e.g., Byram and Garforth 1980, Fussell and Haaland 1978). Proficiency may also extend beyond general 'visual language fluency' to expertise with culturally specific visual languages. Recent research has shown that the neurocognitive processes for Environmental-Conjunction were modulated by participants' experience reading Japanese manga while growing up (Cohn and Kutas 2017). Manga, as discussed above, use more E-Conjunction than Western comics. Frequent manga readers relied more on combinatorial processing than updating, while infrequent readers use more updating processes. Such results imply that sequential image understanding relies not only on experience with visual narratives in general, but with culturally specific visual languages.

4 Page layout: External compositional structure

4.1 Theory

Separate from their meaning, panels are arranged into a layout that serves as the ***external compositional structure*** (ECS) of a canvas (i.e., a page, webpage, wall, etc.). That is, it is the compositional structure 'external' to the panel (i.e., their spatial relations to each other) rather than the ***internal compositional structure*** (ICS) of spatial relations of elements inside a panel. ECS itself refers to the spatial relationships between panels, and remains independent of the meaningful elements within images. For example, a six-panel sequence might be arranged horizontally, vertically, as a 3x2 grid, as a 2x3 grid, or other layouts. Unless the layout actually changed the reading order, the same general meaning would persist.

Indeed, structures of ECS have been identified in isolation from content (Bateman et al. 2016, Cohn 2013a). The standard feature of layouts is a ***grid***—panels ordered in horizontal rows stacked vertically—similar to the layout of words in alphabetic writing systems. This arrangement implies a ***Z-path*** of navigation, moving horizontally between panels within rows, before moving vertically between them. A ***pure grid*** (Figure 4a) occurs when all panels maintain contiguous vertical and horizontal borders. A ***horizontal stagger*** (4c) misaligns the vertical borders, but maintains contiguity of a row. In contrast, a ***vertical stagger*** (4b) misaligns the horizontal borders, but not the vertical ones, thereby disrupting the contiguity of a row. If a vertical stagger is pushed far enough, it becomes a ***blockage*** arrangement (4d). Blockage occurs when a large vertical panel spans a column of vertically stacked panels. In addition, an ***inset panel*** (4f) occurs when one panel is enclosed inside a larger, ***dominant*** panel. Other features of layout further play with arrangements of panels, their distance between each other—proximity through ***separation*** (4g) or ***overlap*** (4h)—their shapes, and their borders (or lack thereof).

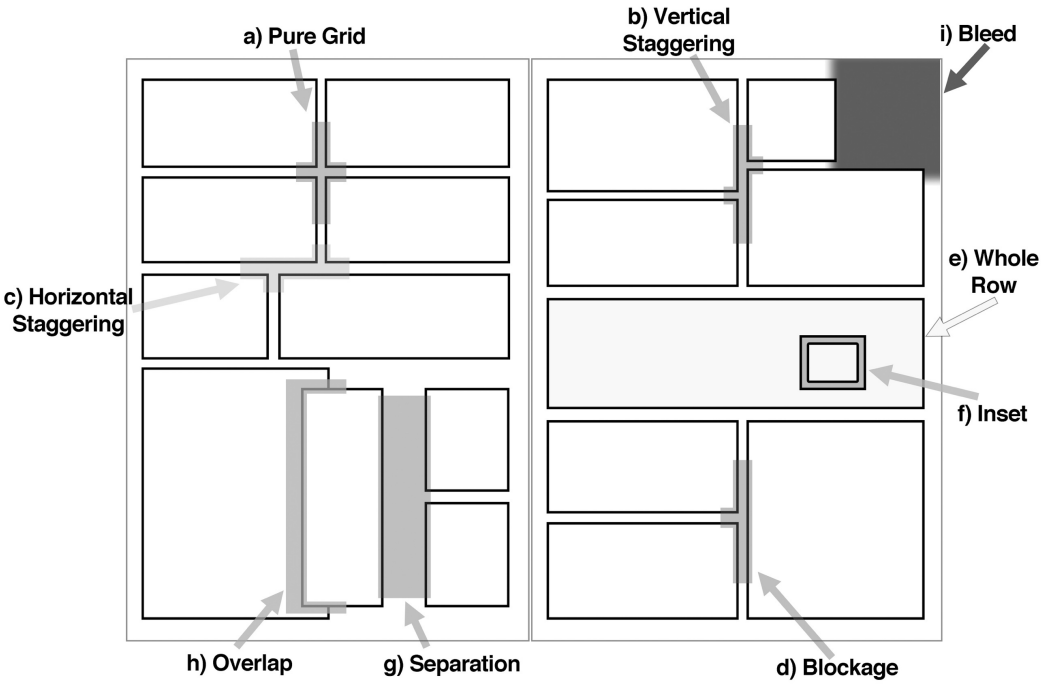


Figure 4. Basic features of the external compositional structure of a page layout.

Surface ECS features index a more complex underlying structure. Basic arrangements divide canvases into horizontal and vertical constituents (rows and columns) which embed inside each other. For example, page 2 of Figure 1 uses four horizontal constituents (rows), with an embedded vertical constituent (column) within the third row. This creates a hierarchic structure, which is suggested both by experimental research (see below) and by computational models (Cao, Chan, and Lau 2012, Tanaka et al. 2007).

As discussed above, ECS and narrative are separate structures, but they interface with each other. As yet, the preferences of such alignment are unclear. In Figure 1, both pages end with Peaks and begin with Establishers, preserving a canonical schema at some level within each page unit. In addition, the three conjunctions occur within rows—these narrative constituents align with the ECS constituents, and are not divided. Nevertheless, little empirical data has informed whether such patterns are systematic, or whether misalignment would affect a reader. While preferred interfaces between narrative and ECS most certainly exist (ex. Establishers as first panel of a page), only limited work has speculated on them (Cohn 2014a).

Layout and narrative/semantics interfacing also arises in the relationship between internal and external compositional structure, i.e., how the composition within a panel (ICS) interacts with the arrangement of panels (ECS). In Figure 1, actions within panels flow in concert with their reading order. In panel 2.3, the beads are thrown left-to-right with the reading path; in panel 2.6 Usagi jumps upward towards panel 2.7, where his motion continues downward towards 2.8, where landing retains a vertical motion. That is, between 2.6-2.8, Usagi's actions imply an arcing path which aligns with the physical reading of the reverse-blockage layout. We might hypothesize that depicted movement within panels that aligns with the reading path (ex. Figure 1) might facilitate comprehension more than content mismatching the reading path (ex. a person flying upward with a downward vertical reading path). These surface alignments are among the qualities that critics and scholars fixate on the most about the visual language used in comics, though they have received little empirical examination.

4.2 Cross-cultural variation

Corpus analyses have begun analyzing the ECS of comic pages within and between cultures. The first analysis examined how ECS has changed in American superhero comics from the 1940s to the 2010s (Pederson and Cohn 2016). Overall, Z-path layouts appeared to decrease over time, while vertical arrangements increased. This manifested in fewer grid-type arrangements, primarily horizontal staggering, with a rise in verticality from stacked 'widescreen' panels spanning the width of a page. Layouts also used more variant non-grid features, especially overlapping and bleeding panels. Overall, ECS in superhero comics has become both more decorative and more systematic, treating the page as a whole canvas.

Research has also identified trends in layouts across cultures (Cohn et al. Under Review). European books (Swedish, French) rely more on a canonical Z-path, with higher proportions of horizontal staggering. Asian pages (Japanese shonen manga, Hong Kong manhua) used substantially more vertical columns, particularly blockage, and ***bleeds*** where borderless panels extend beyond the edge of a page. This verticality could possibly be an influence of the Japanese

and Chinese writing systems, which can also arrange words in vertical columns (an N-path), despite the Z-path remaining their dominant ECS arrangement of panels. These arrangements appear different than the vertical columns in American superhero comics, which manifest largely as stacked widescreen panels—essentially 'rows' comprising only a single panel. American superhero comics also showed more variability in ECS features, consistent with interpretations of decorative page layouts.

4.3 Processing

Experimental research on ECS has focused on two primary measurements: reader choices for panel orders, and observed eye-movements across pages. This research has supported a separation between ECS and meaning. Readers' eye-movements suggest that they generally do not explore various options before committing to a reading path, implying that choices are not based on image content. These reading paths are generally consistent with the basic mechanisms of reading text—i.e., smooth reading paths with eye regressions largely to spatially adjacent panels (Foulsham, Wybrow, and Cohn 2016, Omori, Ishii, and Kurata 2004). In addition, readers choose clear preferences for navigating through page layouts, even in empty panels without content (Cohn 2013a, Cohn and Campbell 2015). Finally, comprehension differs little for altered layouts of the same content (Foulsham, Wybrow, and Cohn 2016, Omori, Ishii, and Kurata 2004).

The basic Z-path appears to be a default for navigating ECS, but various arrangements exert pressure towards alternate routes. For example, the further a vertical stagger approaches a blockage layout, the more likely that readers will choose a vertical path (Cohn and Campbell 2015). Such reading orders appear to be conditioned by a preference rule system (Cohn 2013a), whereby competing constraints pressure a reader to navigate in different ways. These constraints balance the more rule-bound Z-path with the perceptual features of contiguity, proximity, connectedness, and other Gestalt principles (e.g., Wertheimer 1923). Overall though, readers follow a basic principle of ***assemblage***, which motivates readers to (1) prefer grouped areas to ungrouped areas, (2) prefer smooth paths to broken paths, and (3) prefer not to jump over panels or (4) not to leave gaps. These overall preferences work to build the hierarchic groupings of an ECS.

Despite the broad agreement about the hierarchic structure of page layouts (Bateman et al. 2016, Cao, Chan, and Lau 2012, Cohn 2013a), the cross-cultural corpus research mentioned above poses questions to current theories of layout, particularly the findings that European layouts use more horizontal staggering. Imagine a pure grid with six panels: two rows of three panels each. This ECS should involve two horizontal constituents (rows) embedded within a single vertical constituent (one column). This same structure should appear for a horizontal stagger of two rows. However, because European comics use more horizontal staggering than American superhero comics or Asian books (manga, manhua), it suggests that European creators store something in memory that is different about how layouts could (should?) be constructed. That is, they are storing something beyond simply the general hierarchic structure, which does not distinguish between pure grids and horizontal staggering.

Such results raise questions about the generalizability of navigation by readers of different visual languages. Would readers of European comics have greater sensitivity to the salience of horizontal staggering? Certainly, such salience is suggested by the relative prominence of a theoretical construct of the 'strip' (i.e., row) in European theories of page layout (e.g., Groensteen 2007,

Chavanne 2015). Furthermore, because manga and manhua use more blockage and vertical constituents, would their readers more easily navigate downward in blockage arrangements? Would readers of books that hardly use blockage have greater difficulty following this vertical, non-Z-path route? While general visual language fluency seems to modulate some aspects of readers' navigations through ECS (Cohn 2013a, Cohn and Campbell 2015, Nakazawa 2002), these types of cross-cultural influences have yet to be examined in depth.

5 Conclusion and Questions

This chapter has summarized the basic principles and constructs of Visual Language Theory. Though these substructures remain distinct, several themes extend throughout. Below, I discuss how such themes might frame additional research into visual languages across disciplines.

5.1 Structural comparisons

The constructs of VLT are based on a linguistic analysis of the structure of visual and multimodal information. There are many important questions for future empirical research on basic cognition, as hinted above, particularly related to how visual languages overlap with other linguistic and cognitive systems. In particular, does the brain draw from general processing mechanisms in the comprehension of visual languages? VLT predicts such overlap, and both behavioral and neurocognitive research has supported this (Cohn et al. 2014, e.g., , Magliano et al. 2015), though additional targeted studies need to further tease apart these complex relationships.

Questions also arise regarding how structure or processing in visual languages might be compared with that in other behavioral domains (Cohn 2013b, Magliano et al. 2013). For example, how might narrative structure differ between modalities (drawings, film, discourse) or how might ECS vary across formats (comics, magazines, text)? VNG specifically has already been applied to several domains, including film (Amini et al. 2015, Yarhouse 2017), motion graphics (Barnes 2017), health communication (Sontag and Barnes 2017), discourse (Kosara 2017, Versluis 2017), and others. It would be fruitful for further empirical work to examine the extent to which different formats engage similar structures (within and between cultures), and how structures adapt to the affordances of different modalities (e.g., how might the affordances of film push narrative structure to differ from graphic visual languages?).

Finally, beyond basic research, these constructs in VLT can provide rich descriptions for describing the phenomenon that occur in visual narratives. Full corpus analyses can be insightful for quantifying the properties of individual authors and populations of comics. However, VLT can also provide qualitative scholarship more grounded descriptions than approaches which have been shown to be less psychologically viable (e.g., panel transitions).

5.2 Proficiency

Another common theme in VLT is the role of proficiency. Researchers have long observed that various aspects of visual language understanding develop with age and experience reading visual narratives (Nakazawa 2016), and indeed the understanding of sequential images requires cultrual exposure (e.g., Byram and Garforth 1980, Fussell and Haaland 1978). This means that 'comics'

are not as universally understandable as has been otherwise characterized (McCloud 1993). These questions of expertise extend across fluency in general and for culturally specific visual languages. Such findings complicate the presumed accessibility of 'comics' as a communicative tool, which can impact the growing advocacy of using them as teaching tools or cross-cultural communicative materials (Hosler and Boomer 2011, Nalu 2011, Short, Randolph-Seng, and McKenny 2013, Topkaya 2016) , or 'expertise neutral' experimental stimuli (Sivaratnam et al. 2012, Baron-Cohen, Leslie, and Frith 1986). Empirical work is necessary to further explore the trajectory of development in understanding visual languages, as well as their effectiveness in such educational and humanitarian contexts, including their limitations regarding variable proficiencies.

Because visual languages vary in their structure and how they are processed, it raises important concerns not just for basic cognition, but also for the treatment of these forms in society. For example, people may prefer works that conform to the structures of visual languages they are most fluent in—be they specific to certain cultures, time periods, genres, etc. Such preferences may form biases about different types of works, their qualities, and the sociocultural identities that they may facilitate. Exploring these preferences would be a fruitful line of research bridging the intersection of the cognitive and the social aspects of visual languages.

5.3 Diversity/Uniformity

Finally, a tension pervades this work between the abstract structures that may extend across all visual languages, and how they arise in diverse contexts. Empirically driven work analyzing different systems provides an effective qualitative way to characterize both the specific structures that arise in idiolects of individuals (patterns of a particular person) and in cultural visual languages (reflecting patterns across groups of individuals). Corpus research using VLT have already been effective in characterizing many aspects of the visual languages found in comics, and, as argued above, provide more grounded theories than approaches without experimental support. Such efforts can align with interests of the digital humanities to explore the properties of literary works with empirical methods.

Corpus research will no doubt benefit from the growing work in the computer sciences on automated and machine learning analyses of comics (Cao, Chan, and Lau 2012, Guérin et al. 2017, Rigaud et al. 2016), in addition to more complex annotations from human coders. In turn, data science methods can effectively be used to analyze such annotation data by detecting emergent patterns in structure (Laraudogoitia 2008). VLT can inform such efforts, as corpus research is only as good as the constructs being annotated.

In addition, it is worth considering the broader implications of cross-cultural research done so far within VLT. Consistent findings of patterned regularities across and within cultures suggest the presence of diverse visual languages, and comprehension appears to be modulated by familiarity with those patterns (Cohn and Kutas 2017). What is conventionally referred to as a 'comics medium' arises in the presumed similarity between social objects created using several *different* visual languages—often interacting with written languages—arising in diverse contexts (comics, manga, bande dessinée, etc.). Such works ('comics') are defined as separate from those using similar structures, but arising in other social constructs (illustrated picture books, cartoons, animation, doodles, etc.).

To the extent that such visual languages are indeed diverse systems—as the empirical work suggests—it calls into question the utility (and accuracy) of referring to a monolithic notion of a 'comics medium' at all. Such findings further reinforce the necessity of splitting the notion of 'visual languages' from that of 'comics', and thereby pushing scholars to ask just what is their object of inquiry within a 'comics studies.'

5.4 Methodology

Finally, aside from the actual constructs of VLT, the methodology used in this paradigm of combining theoretical, corpus, and experimental research remains important for the future of examining the visual languages used in comics. As discussed at the outset, effective experimental and corpus research should test predictions made by theories, rather than broad exploratory or *ad hoc* analyses motivated by a tool (e.g., an annotator, an eye-tracker). In addition, theoretical work should account for empirical findings to ground their claims. Maintaining these standards—regardless of theoretical paradigms—should guide effective research into the future, and ensure that the study of the visual languages found in comics is a *scientific* one.

References

Abbott, Michael, and Charles Forceville. 2011. "Visual representation of emotion in manga: Loss of control is Loss of hands in *Azumanga Daioh* Volume 4." *Language and Literature* 20 (2):91-112. doi: 10.1177/0963947011402182.

Amini, Fereshteh, Nathalie Henry Riche, Bongshin Lee, Christophe Hurter, and Pourang Irani. 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*, 1459-1468. New York, NY: ACM.

Barnes, Spencer. 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, Simon, Alan M Leslie, and Uta Frith. 1986. "Mechanical, behavioural and intentional understanding of picture stories in autistic children." *British Journal of developmental psychology* 4 (2):113-125.

Bateman, John A., Francisco O. D. Veloso, Janina Wildfeuer, Felix HiuLaam Cheung, and Nancy Songdan Guo. 2016. "An Open Multilevel Classification Scheme for the Visual Layout of Comics and Graphic Novels: Motivation and Design." *Digital Scholarship in the Humanities*:1-35. doi: 10.1093/llc/fqw024.

Bateman, John A., and Janina Wildfeuer. 2014. "A multimodal discourse theory of visual narrative." *Journal of Pragmatics* 74:180-208. doi: 10.1016/j.pragma.2014.10.001.

Byram, M.L., and Chris Garforth. 1980. "Research and testing non-formal education materials: a multi-media extension project in Botswana." *Educational Broadcasting International* 13 (4):190-194.

Cao, Ying, Antoni B. Chan, and Rynson W. H. Lau. 2012. "Automatic stylistic manga layout." *ACM Transactions on Graphics* 31 (6):1-10. doi: 10.1145/2366145.2366160.

Carello, Claudia, Lawrence D. Rosenblum, and A. Grosofsky. 1986. "Static depiction of movement." *Perception* 15 (1):41-58.

Chavanne, Renaud. 2015. "The Composition of Comics." *European Comic Art* 8 (1):111-144.

Cohn, Neil. 2010. "The limits of time and transitions: Challenges to theories of sequential image comprehension." *Studies in Comics* 1 (1):127-147.

Cohn, Neil. 2013a. "Navigating comics: An empirical and theoretical approach to strategies of reading comic page layouts." *Frontiers in Psychology—Cognitive Science* 4:1-15. doi: 10.3389/fpsyg.2013.00186.

Cohn, Neil. 2013b. *The visual language of comics: Introduction to the structure and cognition of sequential images*. London, UK: Bloomsbury.

Cohn, Neil. 2013c. "Visual narrative structure." *Cognitive Science* 37 (3):413-452. doi: 10.1111/cogs.12016.

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

Cohn, Neil. 2014b. "You're a good structure, Charlie Brown: The distribution of narrative categories in comic strips." *Cognitive Science* 38 (7):1317-1359. doi: 10.1111/cogs.12116.

Cohn, Neil. 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, Neil. 2016. "A multimodal parallel architecture: A cognitive framework for multimodal interactions." *Cognition* 146:304-323. doi: 10.1016/j.cognition.2015.10.007.

Cohn, Neil. In press. "Structural complexity in visual narratives: Theory, brains, and cross-cultural diversity." In *Narrative Complexity and Media: Experiential and Cognitive Interfaces*, edited by Marina Grishakova and Maria Poulaki. Lincoln: University of Nebraska Press.

Cohn, Neil, Jessika Axnér, Michaela Diercks, Rebecca Yeh, and Kaitlin Pederson. Under Review. "The cultural pages of comics: Cross-cultural variation in page layouts."

Cohn, Neil, and Hannah Campbell. 2015. "Navigating comics II: Constraints on the reading order of page layouts." *Applied Cognitive Psychology* 29 (2):193-199. doi: 10.1002/acp.3086.

Cohn, Neil, and Sean Ehly. 2016. "The vocabulary of manga: Visual morphology in dialects of Japanese Visual Language." *Journal of Pragmatics* 92:17-29. doi: 10.1016/j.pragma.2015.11.008.

Cohn, Neil, Ray Jackendoff, Phillip J. Holcomb, and Gina R. Kuperberg. 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, Neil, and Marta Kutas. 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, Neil, and Marta Kutas. 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, Neil, and Stephen Maher. 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, Neil, Beena Murthy, and Tom Foulsham. 2016. "Meaning above the head: Combinatorial constraints on the visual vocabulary of comics." *Journal of Cognitive Psychology* 28 (5):559–574. doi: 10.1080/20445911.2016.1179314.

Cohn, Neil, and Martin Paczynski. 2013. "Prediction, events, and the advantage of Agents: The processing of semantic roles in visual narrative." *Cognitive Psychology* 67 (3):73-97. doi: 10.1016/j.cogpsych.2013.07.002.

Cohn, Neil, Martin Paczynski, Ray Jackendoff, Phillip J. Holcomb, and Gina R. Kuperberg. 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, Neil, and Eva Wittenberg. 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.

Díaz Vera, Javier E. 2013. "Woven emotions: Visual representations of emotions in medieval English textiles." *Review of Cognitive Linguistics* 11 (2):269-284. doi: 10.1075/rcl.11.2.04dia.

Forceville, Charles. 2005. "Visual representations of the idealized cognitive model of anger in the Asterix album *La Zizanie*." *Journal of Pragmatics* 37 (1):69-88.

Forceville, Charles. 2011. "Pictorial runes in *Tintin and the Picaros*." *Journal of Pragmatics* 43 (3):875-890.
Forceville, Charles. 2016. "Conceptual Metaphor Theory, Blending Theory, and other Cognitivist Perspectives on Comics " In *The Visual Narrative Reader*, edited by Neil Cohn, 89-114. London: Bloomsbury.
Forceville, Charles, Tony Veale, and Kurt Feyaerts. 2010. "Balloonics: The Visuals of Balloons in Comics." In *The Rise and Reason of Comics and Graphic Literature: Critical Essays on the Form*, edited by Joyce Goggin and Dan Hassler-Forest, 56-73. Jefferson: McFarland & Company, Inc.
Foulsham, Tom, Dean Wybrow, and Neil Cohn. 2016. "Reading without words: Eye movements in the comprehension of comic strips." *Applied Cognitive Psychology* 30:566-579. doi: 10.1002/acp.3229.
Friedman, Sarah L., and Marguerite B. Stevenson. 1975. "Developmental Changes in the Understanding of Implied Motion in Two-dimensional Pictures." *Child Development* 46:773-778.
Fussell, Diana, and Ane Haaland. 1978. "Communicating with Pictures in Nepal: Results of Practical Study Used in Visual Education." *Educational Broadcasting International* 11 (1):25-31.
Green, Jennifer. 2014. *Drawn from the ground: Sound, sign and inscription in Central Australian sand stories*. Cambridge, UK: Cambridge University Press.
Groensteen, Thierry. 2007. *The System of Comics*. Translated by Bart Beaty and Nick Nguyen. Jackson: University of Mississippi Press.
Gross, Dana, Nelson Soken, Karl S. Rosengren, Anne D. Pick, Bradford H. Pillow, and Patricia Melendez. 1991. "Children's Understanding of Action Lines and the Static Representation of Speed of Locomotion." *Child Development* 62:1124-1141.
Guérin, Clément, Christophe Rigaud, Karell Bertet, and Arnaud Revel. 2017. "An ontology-based framework for the automated analysis and interpretation of comic books' images." *Information Sciences* 378:109-130. doi: http://dx.doi.org/10.1016/j.ins.2016.10.032.
Hosler, Jay, and K. B. Boomer. 2011. "Are Comic Books an Effective Way to Engage Nonmajors in Learning and Appreciating Science?" *CBE-Life Sciences Education* 10 (3):309-317.
Ito, Hiroyuki, Takeharu Seno, and Miyuki Yamanaka. 2010. "Motion impressions enhanced by converging motion lines." *Perception* 39 (11):1555-1561.
Jackendoff, Ray. 2002. *Foundations of Language: Brain, Meaning, Grammar, Evolution*. Oxford: Oxford University Press.
Kaan, Edith. 2007. "Event-Related Potentials and Language Processing: A Brief Overview." *Language and Linguistics Compass* 1 (6):571-591. doi: 10.1111/j.1749-818X.2007.00037.x.
Kawabe, Takahiro, and Kayo Miura. 2006. "Representation of dynamic events triggered by motion lines and static human postures." *Experimental Brain Research* 175 (2):372-375. doi: 10.1007/s00221-006-0673-6.
Kosara, Robert. 2017. "An Argument Structure for Data Stories." Proceedings of the Eurographics Conference on Visualization (EuroVis) 2017, Barcelona, Spain.
Kutas, Marta, and Kara D. Federmeier. 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.

Lakoff, George, and Mark Johnson. 1980. *Metaphors we live by*. Chicago, IL: University of Chicago Press.

Laraudogoitia, Jon Pérez. 2008. "The Comic as a Binary Language: An Hypothesis on Comic Structure." *Journal of Quantitative Linguistics* 15 (2):111-135.

Magliano, Joseph P., Adam M. Larson, Karyn Higgs, and Lester C. Loschky. 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, Joseph P., Lester C. Loschky, James A. Clinton, and Adam M. Larson. 2013. "Is Reading the Same as Viewing? An Exploration of the Similarities and Differences Between Processing Text- and Visually Based Narratives." In *Unraveling the Behavioral, Neurobiological, and Genetic Components of Reading Comprehension*, edited by B. Miller, L. Cutting and P. McCardle, 78-90. Baltimore, MD: Brookes Publishing Co.

Magliano, Joseph P., and Jeffrey M. Zacks. 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.

McCloud, Scott. 1993. *Understanding Comics: The Invisible Art*. New York, NY: Harper Collins.

Mori, Kazuhiko. 1995. "The influence of action lines on pictorial movement perception in pre-school children." *Japanese Psychological Research* 27 (3):183-187.

Nakazawa, Jun. 2002. "Analysis of manga (comic) reading processes: Manga literacy and eye movement during Manga reading." *Manga Studies* 5:39-49.

Nakazawa, Jun. 2016. "Manga literacy and manga comprehension in Japanese children." In *The Visual Narrative Reader*, edited by Neil Cohn, 157-184. London: Bloomsbury.

Nalu, Amber. 2011. "Comics as a cognitive training medium for expert decision making." Doctoral dissertation, Human Factors Psychology, Old Dominion University.

Newton, Douglas P. 1985. "Children's Perception of Pictorial Metaphor." *Educational Psychology* 5 (2):179-185. doi: 10.1080/0144341850050207.

Omori, Takahide, Taku Ishii, and Keiko Kurata. 2004. "Eye catchers in comics: Controlling eye movements in reading pictorial and textual media." 28th International Congress of Psychology, Beijing, China.

Pederson, Kaitlin, and Neil Cohn. 2016. "The changing pages of comics: Page layouts across eight decades of American superhero comics." *Studies in Comics* 7 (1):7-28. doi: 10.1386/stic.7.1.7_1.

Rigaud, Christophe, Thanh-Nam Le, J. C. Burie, Jean-Marc Ogier, Shoya Ishimaru, Motoi Iwata, and Koichi Kise. 2016. "Semi-automatic Text and Graphics Extraction of Manga Using Eye Tracking Information." 2016 12th IAPR Workshop on Document Analysis Systems (DAS), 11-14 April 2016.

Saraceni, Mario. 2016. "Relatedness: Aspects of textual connectivity in comics." In *The Visual Narrative Reader*, edited by Neil Cohn, 115-129. London: Bloomsbury.

Shinohara, Kazuko, and Yoshihiro Matsunaka. 2009. "Pictorial metaphors of emotion in Japanese comics." In *Multimodal Metaphor*, edited by Charles Forceville and Eduardo Urios-Aparisi, 265-293. New York: Mouton De Gruyter.

Short, Jeremy C., Brandon Randolph-Seng, and Aaron F. McKenny. 2013. "Graphic Presentation: An Empirical Examination of the Graphic Novel Approach to Communicate Business Concepts." *Business Communication Quarterly* 76 (3):273-303. doi: 10.1177/1080569913482574.

Sivaratnam, Carmel S., Kim Cornish, Kylie M. Gray, Patricia Howlin, and Nicole J. Rinehart. 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.

Sontag, Jennah M., and Spencer R. Barnes. 2017. "The visual framing of graphics when used in preventative health digital news packages: exploring the use of a narrative structure as the message infrastructure." *Journal of Visual Communication in Medicine* 40 (3):109-119. doi: 10.1080/17453054.2017.1366824.

Stainbrook, Eric 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*, edited by Neil Cohn, 129-154. London: Bloomsbury.

Tanaka, Takamasa, Kenji Shoji, Fubito Toyama, and Juichi Miyamichi. 2007. "Layout Analysis of Tree-Structured Scene Frames in Comic Images." International Joint Conference on Artificial Intelligence, Hyderabad, India, January 6-12, 2007.

Topkaya, Yavuz. 2016. "The Impact of Instructional Comics on the Cognitive and Affective Learning about Environmental Problems." *Egitim ve Bilim* 41 (187):199-219.

Versluis, Christine. 2017. *Why this now? A genre analytic approach to aphasic/non-aphasic interactive events*. Utrecht: LOT.

Wellman, Henry M., Michelle Hollander, and Carolyn A. Schult. 1996. "Young children's understanding of thought bubbles and thoughts." *Child Development* 67:768-788.

Wertheimer, Max. 1923. "Untersuchungen zur Lehre von der Gestalt (Laws of Organization in Perceptual Forms)." *Psychol. Forsch.* 4:301-350.

West, W. Caroline, and Phil Holcomb. 2002. "Event-related potentials during discourse-level semantic integration of complex pictures." *Cognitive Brain Research* 13:363-375.

Wichmann, Søren, and Jesper Nielsen. 2016. "Sequential Text-Image Pairing among the Classic Maya." In *The Visual Narrative Reader*, edited by Neil Cohn, 282-313. London: Bloomsbury.

Wilkins, David P. 1997/2016. "Alternative Representations of Space: Arrernte Narratives in Sand." In *The Visual Narrative Reader*, edited by Neil Cohn, 252-281. London: Bloomsbury. Original edition, 1997. Proceedings of the CLS Opening Academic Year '97 '98, edited by M. Biemans and J. van de Weijer, 133-164. Nijmegen: Nijmegen/Tilburg Center for Language Studies.

Yannicopoulou, Angela. 2004. "Visual Aspects of Written Texts: Preschoolers View Comics." *L1- Educational Studies in Language and Literature* 4:169-181.

Yarhouse, R. Brad. 2017. "Examining Film Engagement Through the Visual Language of Comics." *International Journal of Comic Art* 19 (1):384-404.

Zwaan, Rolf A., and Gabriel A. Radvansky. 1998. "Situation models in language comprehension and memory." *Psychological Bulletin* 123 (2):162-185.