Structural complexity in visual narratives:
Theory, brains, and cross-cultural diversity

Neil Cohn

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

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

Cohn, Neil. 2019. Structural complexity in visual narratives: Theory, brains, and cross-cultural diversity. In Grishakova, Marina and Maria Poulaki (Ed.). *Narrative Complexity and Media: Experiential and Cognitive Interfaces*. (pp. 174-199). Lincoln*:* University of Nebraska Press

Introduction

Narratives are a fundamental aspect of human communication, and extend across several domains of expression. While the structure of verbal narratives has been studied extensively, only recently has cognitive science turned towards investigating the structure of drawn visual narratives. Yet, drawn sequential images date back at least to cave paintings, and appear throughout human history, most popularly in contemporary society in comics. This chapter will survey a research program investigating the structure and cognition of visual narratives across three intersecting domains: narrative structure, cross-cultural variation, and cognitive psychology/neuroscience. Though targeted at research on visual narratives, this approach is proposed as ultimately applying across domains to various types of narrative systems, including verbal and filmed discourse (Cohn 2013b, Magliano, Higgs, and Clinton This volume).

Visual Narrative Grammar

The most prominent beliefs about sequential image understanding hold that readers connect the linear meaningful relations between images. The most popular codification of this approach emerged in theorist and artist Scott McCloud's (1993) proposal of "transitions" between *panels*—the image-units of a visual sequence—with relations like changes in actions, moments, characters, or scenes. Such an approach is similar to models from psycholinguistics which argue that a reader monitors for various types of semantic information throughout a discourse—verbal or visual— elements like time, space, or characters (Magliano and Zacks 2011, Zwaan and Radvansky 1998). When a reader encounters a change in any of these dimensions, a processing cost is incurred for incorporating that information into an altered mental model of the scene (Zwaan and Radvansky 1998). Such a process likely extends across domains, given modality-specific affordances (Magliano, Higgs, and Clinton This volume).

Visual Narrative Grammar (VNG) has argued that, in addition to updating of semantic information, sequential image comprehension requires a *narrative structure* in order to be understood. This narrative grammar assigns panels to categorical roles, and then organizes them into hierarchic constituents at a discourse level of meaning, analogous to how syntactic categories organize words into constituents at a sentence level (Cohn 2013b). That is, because images typically contain more information than individual words, the semantics of the units are closer to whole sentences and thereby operate at a discourse level of *meaning*. However, despite this variance in the semantic structures, the principles that combine these panels remain similar to those between the sentence (syntax) and narrative levels. In essence, narrative structure acts as a "macro-level syntax" that belongs to the surface "textbase" that a comprehender accesses in order to subsequently construct a semantic mental model (van Dijk and Kintsch 1983).

VNG is not the first approach to narrative structure that has drawn on an analogy with syntactic structure. Other grammatical approaches have proposed formal structures for verbal stories (e.g., Mandler and Johnson 1977) and for film (e.g., Carroll 1980, Metz 1974). However, VNG differs from these precedents in that it is based on, and integrated into, contemporary theories of linguistics using construction grammar (Culicover and Jackendoff 2005), not traditional Chomskyan phrase structure grammars (e.g., Chomsky 1965). In a construction grammar, sequencing does not arise out of inserting memorized lexical structures (like words) into rules for phrases, but rather these "rules" exist as schemas stored into long-term memory as lexical items unto themselves, along with interface-rules specifying privileged mappings to semantics (Goldberg 1995, Jackendoff 2002).

While novel narrative schemas can exist, VNG uses several primary sequencing patterns (Table 1): *A canonical narrative schema, a conjunction schema, and a head-modifier schema*. We will address each of these structures in turn.

Table 1. Basic constructional patterns in Visual Narrative Grammar

a) Canonical narrative schema:	$[_{\text{Phase X}}$ (Establisher) – (Initial) – Peak – (Release)]
b) Conjunction schema:	$[_{\text{Phase X}} X_1 - X_2 - \ldots X_n]$
c) Head-modifier schema:	$[_{\text{Phase X}}$ (Modifier) – X – (Modifier)]

First, VNG argues that the semantic cues within the image content of panels can map to narrative categories, which are organized into a canonical narrative schema. This schema is similar to, albeit more operationalized than, traditional notions of narrative arcs (e.g., Freytag 1894, see Cohn 2013b for review). These basic narrative categories include:

Establisher (E) – sets up an interaction without acting upon it, often as a passive state
Initial (I) – initiates the tension of the narrative arc, prototypically a preparatory action and/or a source of a path
Peak (P) – marks the height of narrative tension and point of maximal event structure, prototypically a completed action and/or goal of a path, but also often an interrupted action
Release (R) – releases the tension of the interaction, prototypically the coda or aftermath of an action

These descriptions of narrative roles outline their prototypical correspondences to meaning—i.e., how *semantic* content (the visual cues within images) may influence a panel's *structural* role in a sequence (examples below). Nevertheless, identification of a narrative category uses both a panel's bottom-up content and its top-down context in a global sequence (Cohn 2013b, 2014). Syntactic categories are assigned in a similar way: though syntactic categories (like nouns, verbs) prototypically correspond to the semantics (like objects, events) of words (Jackendoff 1990), they also rely on context within a sentence. For example, the word "dance" (semantically, an event) can play a role either as a noun (*the dance*) or a verb (*they dance*) depending on context.

The basic narrative schema in VNG, as in Table 1a, thus places these narrative categories into a narrative constituent in this particular order. In actualization, not all constituents must contain all categories, meaning that most elements are non-obligatory (as notated by parentheses). Only Peaks are marked as obligatory, because they motivate a sequence as its "head." However, Peaks too can be omitted under specific constrained, inference generating contexts (Cohn and Kutas 2015, Magliano et al. 2015, Magliano et al. 2016).

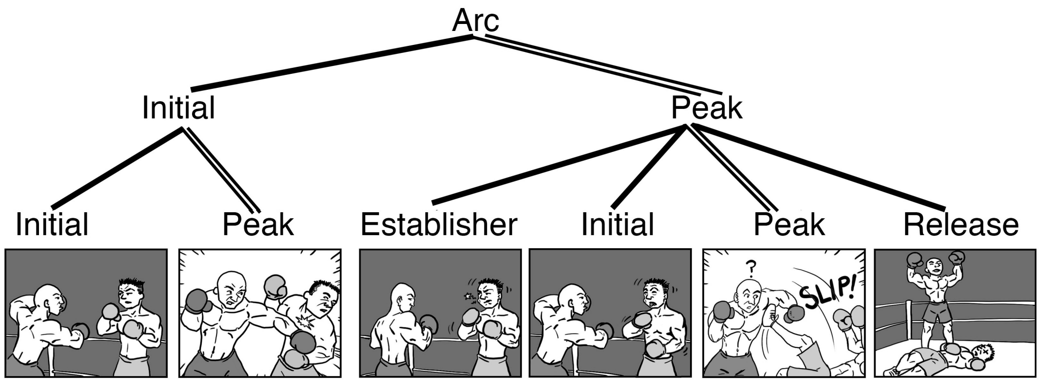

Figure 32. A narrative sequence with two narrative constituents.

Consider Figure 32. It begins with a boxer, who reaches back to punch an adversary. This preparatory action is prototypical of an Initial. The full punch occurs in the next panel, a Peak in relation to that Initial. An Establisher then resets the actions in panel 3, by setting up a new situation with the boxers passively standing facing off again. Another Initial in panel 4 again shows a preparatory action. The subsequent Peak in the penultimate panel does not depict a completed action (as in the second panel), but rather shows an interruption of the boxer's action: he slips. The final panel Release shows the coda of this action, with the victor standing over his opponent.

A first layer of complexity in sequential images can be captured by VNG because narrative categories apply both to individual images and to constituents of images. If taken as a surface string, the narrative roles in this sequence (*I-P-E-I-P-R*) do not conform to the canonical narrative schema. However, combining its segments into constituents allows for groupings that introduce complexity into a narrative. These constituents then also play narrative roles relative to each other at a level above that of individual panels (i.e., groupings of panels). Figure 32 uses two constituents: the first two panels form an Initial that together set up a Peak constituent of the remaining four panels. Internally, each constituent maintains the canonical narrative schema, with Peaks forming the "heads" which motivate the primary meaning of their superordinate constituent (i.e., each grouping is an expansion of its Peak, indicated by double-barred lines). Thus, individual panels and whole constituents both take on narrative roles. This recursion also can extend further upward, since the principles guiding short sequences also apply to higher "plot" level narrative structures. An "Arc" is simply a maximal node, or a constituent that plays no role in a larger sequence.

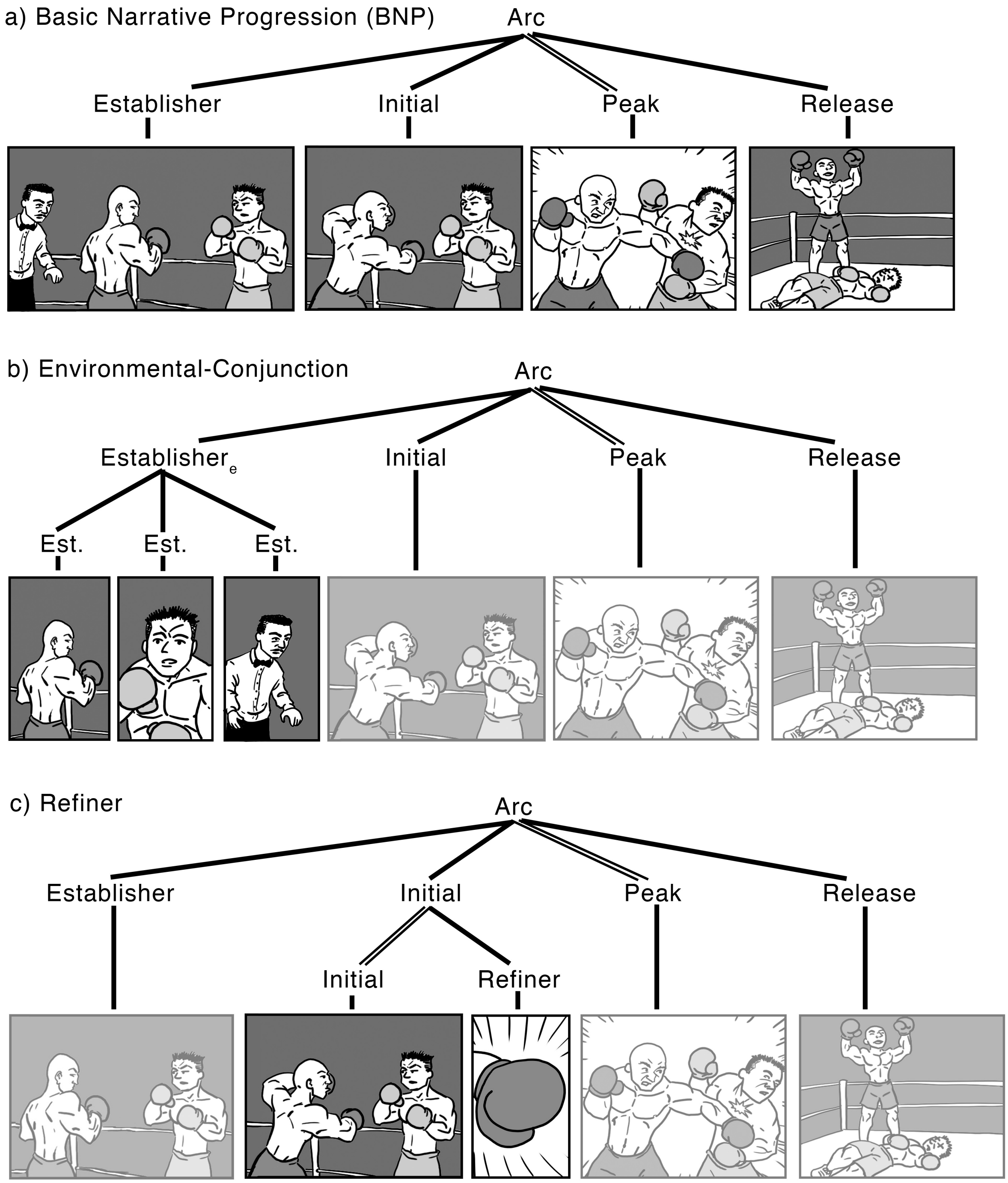

Figure 33. Different basic narrative schemas within Visual Narrative Grammar

Other basic constructs in VNG capture further complexity of sequences by elaborating on the canonical narrative schema. For example, *conjunction* (Table 1b) allows categories to repeat within a constituent of the same category (Cohn 2013b, 2015), similar to how syntactic conjunction repeats grammatical categories (like multiple nouns in a noun phrase: *The butcher, baker, and candlestick maker*). Thus, narrative conjunction also repeats narrative categories within a common constituent. For example, Figure 33b uses conjoined panels at the outset of the sequence. Because these opening panels in (33b) depict the same information as the single first panel in (33a), it suggests that they all function to introduce the scene as Establishers (and indeed, the three panels in 33b could be substituted for the single panel in 33a). However, because each character is now individuated, nothing shows us that they occupy the same spatial location (as in the panel in Figure 33a). Thus, this spatial environment must now be inferred across these panels. Thus, this type of construction is called *Environmental-Conjunction* or "*E-Conjunction*" since it is a conjunction (repetition of panels playing the same narrative role) that leads to an inference of a broader spatial environment (notated with subscript "e").

Scene construction is only one type of meaningful relation possible using conjunction. The conjunction schema only specifies that narrative categories repeat within a constituent, and thus it allows for various *semantic* mappings to this *narrative* schema (Cohn 2013b, 2015). Figure 33 depicts several semantic correspondences for a three-panel conjunction constituent, all Initials. The left tier of Figure 33 shows different three-panel sequences which can operate as conjoined Initials of (a) *actions* or events (*A-Conjunction*), (b) characters within an *environment* (*E-Conjunction*), (c) parts of a single *entity* or character (*N-Conjunction*), or (d) disparate *semantically* associated elements (*S-Conjunction*).

Together, the conjoined images in the left tier are semantically equivalent to (and may create inferences of) the single panels in the right tier. In other words, the single, non-conjoined image in the right tier can substitute for the three conjoined images in the left tier (and this substitution serves as a diagnostic test for assessing conjoined panels). In addition, because conjunction is recursive, these forms can also embed within each other. For example, the three-panels in Figure 33c could replace the first panel in 3b, which would create an N-Conjunction constituent embedded within an E-Conjunction constituent. Thus, the same *narrative* conjunction can have several types of *semantic* mappings.

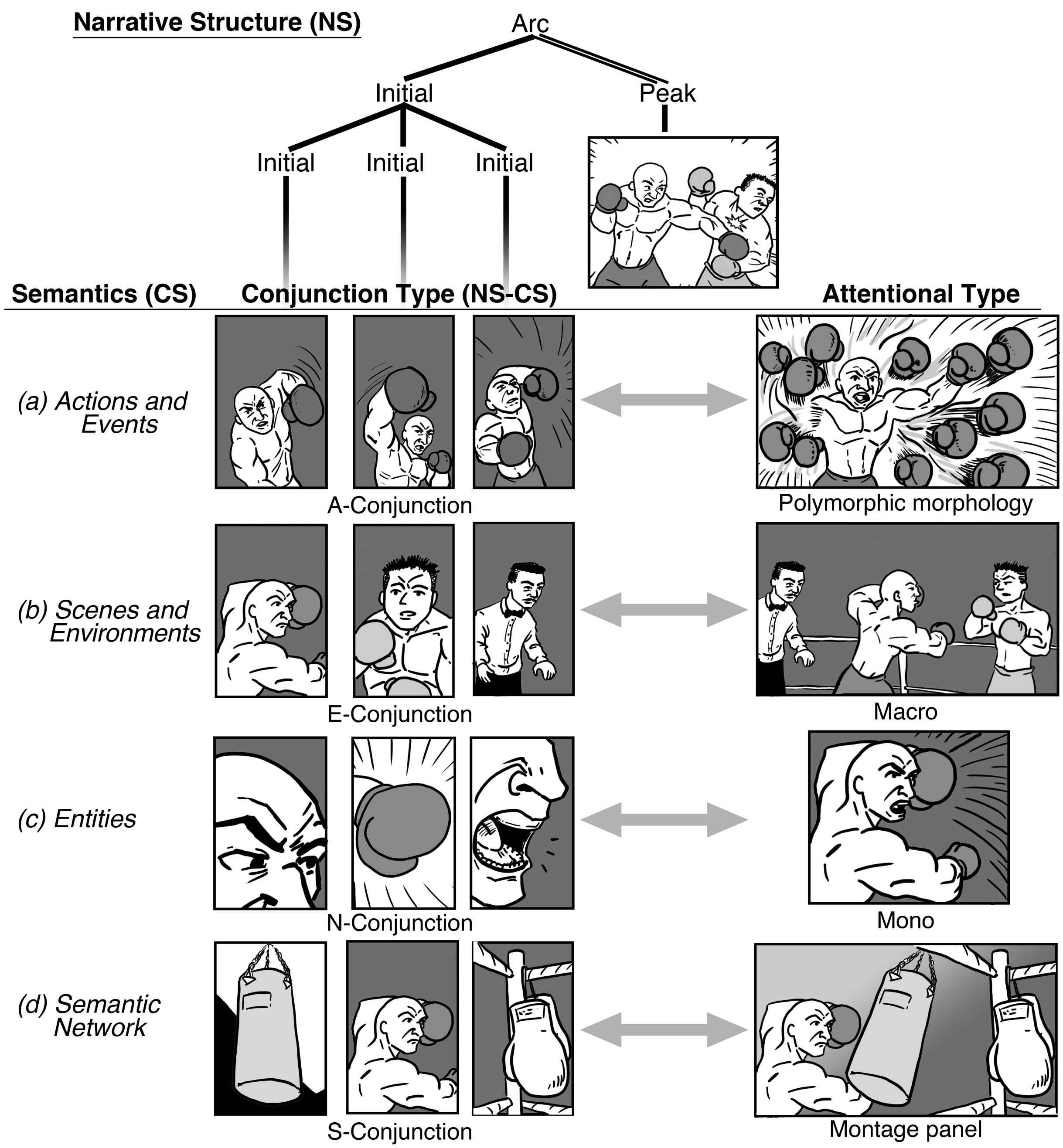

Figure 34. Different ways that narrative conjunction uses repeated narrative categories (here, Initials) to show various semantic information (actions, characters in a scene, parts of an individual, or semantically associated elements). This information could also be framed by a single image (right tier).

VNG can also characterize complexity that occurs when panels modify other panels. The third panel in Figure 33c depicts the same information as its preceding panel, only zooming-in on the puncher's fist. This is a *Refiner* (Cohn 2013a, 2015), which modifies the information in another "head" panel (again, double bar lines) using a narrower viewpoint of the same information. These panels both play the same role in overall narrative arc (in Figure 33c, as an Initial). Unlike conjunction, the narrative role here is not distributed across units, but rather the Refiner modifies the head panel with added focus, while the head retains its wider viewpoint and more fundamental role in the sequence (as in 2a). Refiners can go either before or after their head. Because Refiners modify their head, they should be able to be deleted without recourse on the sequence (unlike in conjunction). If a head is deleted, the modifier takes on the role of the head (ex. in Figure 33c, the Refiner would become the Initial). This is analogous in language to phrases like *I'll take* *<u>the white</u>*, where the adjective *white* takes on the role of a noun in the otherwise more complete *white wine*. Structurally, these constructs are similar in that both are grammatical modifiers which take the role of their heads when those heads are deleted, despite involving different types of, and levels of, semantics.

Both conjunction and Refiners are basic schemas within VNG (Table 1) that can expand sequences into more complicated structures, often as regularized patterns. One such pattern arises in Figure 35a, where the *Alternation* between two characters results in an "A-B-A-B" pattern. This surface pattern is composed of sets of conjoined panels (Cohn 2013a), such that each pairing forms a constituent using E-Conjunction (i.e., [A-B]-[A-B]). In Figure 35a, the first pairing creates an Establisher constituent and the second pairing forms an Initial constituent. Again, this can be confirmed because each pairing of panels can be substituted by the single panels in Figure 33a (Cohn 2013b, 2015). This pattern is a subtype of the "crosscutting" or "multitracking" (Bateman and Schmidt 2012, Bordwell and Thompson 1997) found in films, as well as in drawn visual narratives (Cohn 2013a).

Another pattern uses both conjunction and Refiners. In Figure 35b, the Refiner is separated from the "head" it modifies by an intervening panel. This intervening panel unites with the head using E-Conjunction, forcing the Refiner to connect across a distance. This pattern is called *Refiner Projection*, because the Refiner is "projected" away from its head (Cohn 2013a). Thus, the Refiner must connect across a distance rather than with its juxtaposed images.

The canonical narrative schema, conjunction schema, and head-modifier schema constitute three core narrative patterns, and are stored in long-term memory as "constructions" in line with abstract principles of combination found at the syntactic level as detailed in linguistic theories of construction grammar (Culicover and Jackendoff 2005, Jackendoff 2002). However, as in human languages, narratives do not just use basic schemas. Rather, thousands of constructions appear in regularized ways that employ or depart from those canonical patterns in spoken language. Thus, VNG allows for conventionalized constructional patterns that may or may not use these basic schemas, so long as they are systematic within and/or across authors.

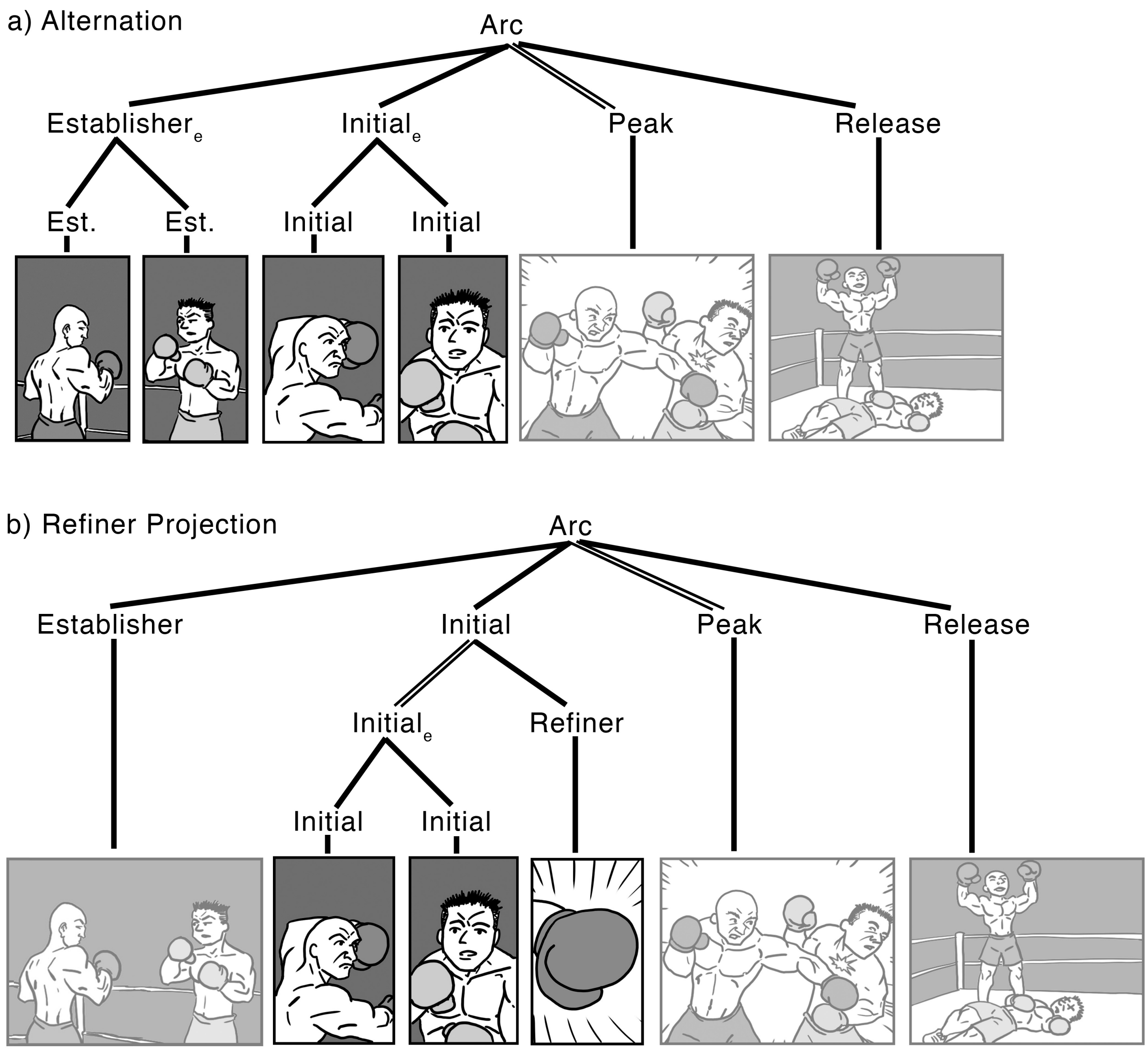

Figure 35. Narrative patterns created by the interaction between narrative schemas within Visual Narrative Grammar

In sum, VNG posits a basic narrative schema composed of categorical roles, which can expand into hierarchic constituents because of its recursive nature. This narrative schema can also be expanded by basic modifying schemas (Conjunction, Refiners) which alter and enrich the framing of attention on a scene. Together, these schemas allow for complex patterning like Alternation and Refiner Projection. Note that in all patterns presented in Figures 32-35, the basic narrative arc persists at the top level of all structures. Formal complexity merely expands from this basic structure. Thus, narrative structure does not simply use a uniform process of updating semantic relations across units of a (visual) discourse (Magliano, Higgs, and Clinton This volume), but rather stitches together structured patterns in complex hierarchic embedding which frame and organize that meaning.

Given the formal complexity that can arise in the structure of a narrative, this may also have consequences on processing and comprehension. Greater complexity in the structure of a narrative grammar—for example, center-embedded clauses, conjunction, Refiners, etc.—may in turn yield more challenging comprehension (e.g., Magliano, Higgs, and Clinton This volume). Such comprehension differences are foreshadowed by the growing empirical work on visual narrative processing (see below), but would reflect similar findings at the syntactic level, where greater structural complexity demands processing costs.

Typological complexity

VNG is embedded within Visual Language Theory which posits that graphic systems manifest in different "visual languages" used by cultures around the world (Cohn 2013a). Just like verbal languages differ around the globe, so do visual languages, and this hypothesis of cross-cultural diversity includes their grammars. Thus, Mainstream American superhero comics are drawn in an "American Visual Language" (specifically the "Kirbyan" dialect), which are predicted to differ from the "Japanese Visual Language" used to create manga. These visual languages combine with writing to create a larger multimodal discourse (Cohn 2016b). In that VNG outlines a narrative grammar in the structure of visual languages, we can now investigate: 1) Do all systems of sequential images use the same narrative grammar, or is there cross-cultural diversity? 2) Are there general trends for complexity in narrative grammars that transcend diversity in individual cultures?

The first question relates to variation between the systems used by different cultures' comics, and indeed claims to this extent go back several decades. McCloud (1993) observed that differences between panel-to-panel semantic relations arose between American comics and Japanese manga, but not between comics from America and Europe. In particular, McCloud noted greater amounts of character changes and fewer temporal changes in Japanese manga than European and American comics. Similarly, we also found differences between the structure of manga and comics from the United States with regard to how a scene is framed (Cohn 2011, Cohn, Taylor-Weiner, and Grossman 2012) which can have consequences on narrative structure, as in E-Conjunction or Refiners.

The second question pertains not to cross-cultural differences, but to the overall trends for structural patterning across narrative systems. We can characterize our aforementioned narrative patterns into three "levels" of complexity. First, the simplest structures depict a full scene in each panel (a "macro" viewpoint) with narrative state changes between each panel, as in Figure 33a and Figure 32. This *Basic Narrative Progression* (BNP) should be the most similar to the iconic perception of viewing events and makes no modulation on framing or narrative pacing. It thus constitutes Level 1 of narrative complexity. Level 2 modifies this structure, and as in conjunction and Refiners. These patterns still use basic schemas within the narrative grammar, but they add complexity beyond a simple progression of a fully depicted scene. Finally, Level 3 consists of patterns that combine these modifiers: Alternation and Refiner Projection. These patterns arise from the interaction between modifying structures, and thus use more complexity than the basic schemas alone.

If narrative patterns arise systematically for different "visual languages" across the world, it provides evidence that authors store structures in memory for a narrative grammar. In VNG, all three levels of complexity described above are, in fact, patterns encoded in memory, be they the basic schemas (Table 1) or constructions derived from them. This position contrasts with theories of sequential images focusing on the semantic relationships between images, where an ongoing inductive process updates meaning without using stored schematic structures in memory (Bateman and Wildfeuer 2014, Magliano and Zacks 2011). Under this semantic panel-to-panel view, narrative arises because of "storytelling choices" (e.g., McCloud 2006), and differences between cultures must amount to broader "cultural" factors operating on those choices (McCloud 1993). However, if narrative patterns vary in a consistent way across different cultures, then it would provide evidence for a narrative grammar operating in the minds of these authors (which would thereby be accessed by readers of those works).

To examine these issues, I drew a sample from the Visual Language Research Corpus (http://www.visuallanguagelab.com/vlrc) examining these patterns in 10,521 panels across 90 comics from various cultures (see Table 2). I here report on a sample of 10 books each from three populations of comics from three parts of the world: the United States (mainstream, indy, OEL manga), Europe (France, Germany, Sweden), and Asia (Japan, China, Korea). This data collapses across most distinctions of genres and demographics that will be explored in other publications. The exception to this is variance in books from the United States, which contrasts both publishing genres (mainstream, indy) and Original English Language (OEL) manga. OEL manga are comics created by English speakers, but ostensibly drawn using the Japanese Visual Language from manga. Inclusion of this sample provides a way to investigate whether the structures used by these authors may be closer to their culture (United States) or their adopted visual language (Japanese Visual Language). It is also worth noting, many authors from the contemporary American indy genre are also influenced by manga, though they may not claim that label overtly (Mazur and Danner 2014).

Table 2. Data from books analyzed from the Visual Language Research Corpus, 10 from each group.

Group	Dates	Total Panels	Panels/page
Swedish	1980–2011	893	6.91
French	1992–2014	1607	7.42
German	1987–2009	1525	5.63
American Mainstream	1990–2014	1016	4.75
American Indy	2002–2014	1336	5.36
OEL Manga	1991–2006	1101	4.57
Japanese	2003–2014	960	5.28
Korean	1987–2010	837	3.90
Chinese	2002–2015	1246	4.90
Total		**10,521**	**5.411964028**

It should also be noted that this corpus analysis of 90 comics is the largest reported in any study of the structure of drawn visual narratives. Nevertheless, data collection on this larger project work is ongoing, and we consider these data as merely preliminary. These reported data were coded by 8 independent researchers, with repeated annotations for ~70% of the books. Before contributing to the corpus, all coders needed to reach a threshold of over 85% agreement with other annotations performed on preliminary practice examples.

Results from this corpus analysis are provided in Figure 36. In general, the overall proportion of patterns decreased across levels (note the differences in scale in Figure 36a-c), with the Basic Narrative Progression constituting the most used of all patterns. This reinforced that this pattern provides a fundamental base for most visual narratives. As evident in Figure 36a, narrative systems in Europe used the BNP more than those from the United States, which in turn used it more than books from Asia and/or those imitative of Japanese manga (Japanese, Chinese, Korean, OEL manga). The lower proportion of the BNP for Japanese manga and those imitative of Japanese manga appeared to coincide with increases in Level 2 patterns (Figure 36b) and Level 3 patterns (Figure 36c) for these same books. Within Level 2, E-Conjunction was used more than Refiners (36b), and within Level 3, Alternation was used more than Refiner Projection (36c).

The overall similarity between both American Mainstream books and European comics echo McCloud's (1993) observation of similarities between semantic panel relations in American and European comics, despite fairly independent narrative traditions (Mazur and Danner 2014). These works were characterized by greater quantities of the Basic Narrative Progression and fewer complex patterns from Level 2 and Level 3. Exceptions arise in the greater usage of Alternation by German comics.

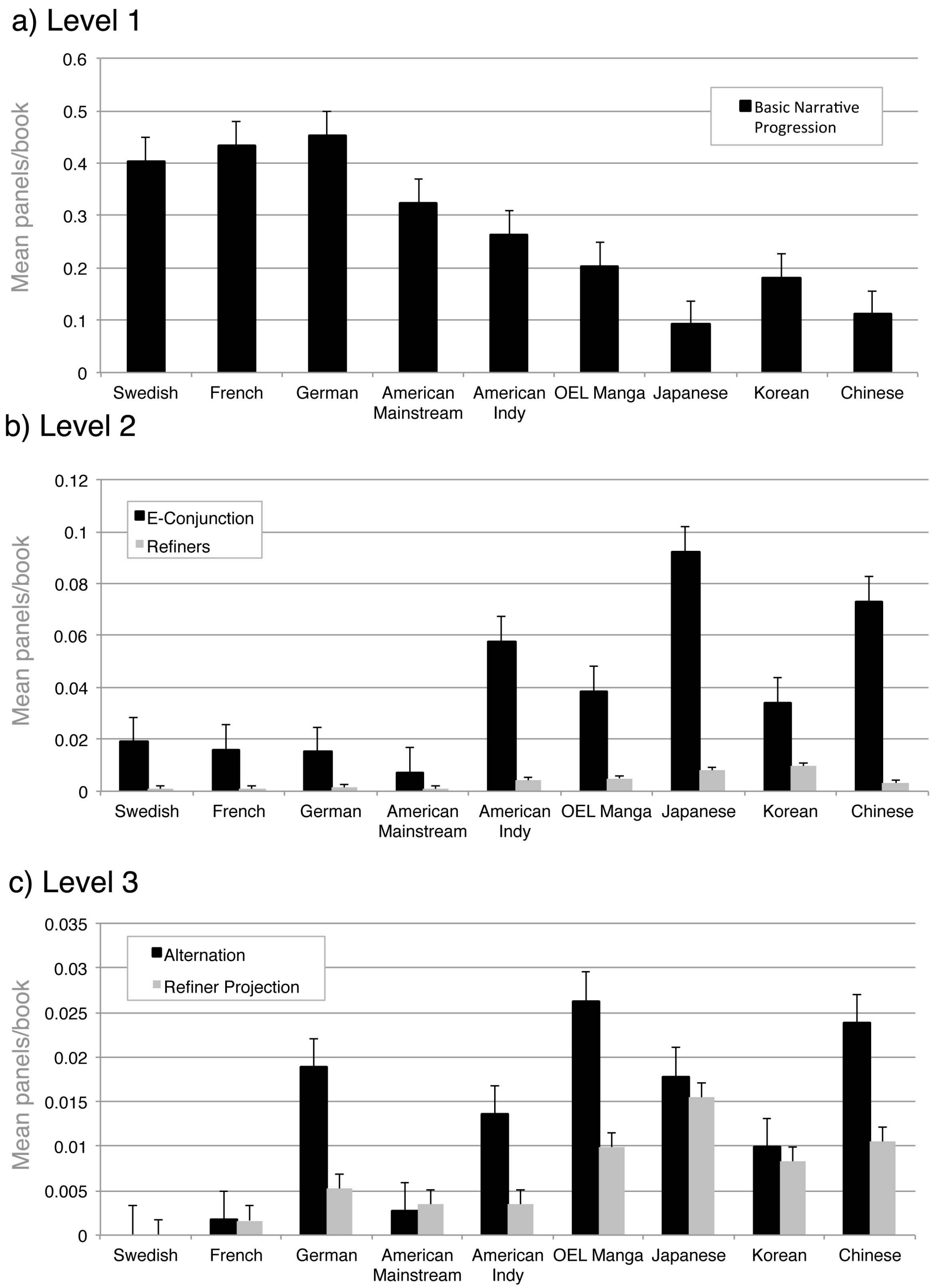

Figure 36. Mean panels per book between groups for each narrative pattern across the three "levels" of complexity. Note the difference in scales between each level.

These patterns of Euro-American narrative styles distinctly differed from Japanese manga and works imitative or inspired/influenced by them (Korean, OEL manga, American indy). Again, this aligns with McCloud's (1993) observation that semantic panel relations differed between Japanese manga and European/American comics. In particular, McCloud found that manga used greater numbers of changes in characters, consistent with our findings of more E-Conjunction in manga. In fact, in the Japanese manga from our sample, E-Conjunction appears nearly at the same proportion as the BNP. These observations also confirm the interpretations drawn from corpus studies of the framing of panels (Cohn 2011, Cohn, Taylor-Weiner, and Grossman 2012) that increased proportions of mono (single character) panels implied more E-Conjunction (Cohn 2013a).

In general, works imitative of, or inspired by, Japanese manga used the BNP less often, and used complex narrative patterns more often. These works often share surface features in their "drawing style" of graphic structure. However, these data suggest that they share structures extending beyond surface forms to consistent narrative grammars. Whether acquisition of this Japanese Visual Language grammar by non-Japanese authors (Korean manhwa, OEL manga) is explicitly or implicitly learned is an important follow up to exploring this consistency.

It is worth noting the trends shown by OEL manga, which were ostensibly created using the Japanese Visual Language (JVL) by English-speakers outside of Japan. These works demonstrated patterns closer to books using their intended visual language (Japanese manga and Korean manhwa) than to books from their culture of origin (America). The similarities of these works' narrative patterns to those in other manga-inspired books (Korean) indicate that narrative grammars are not contingent on cultural divisions (i.e., Asian vs. Euro-American), but rather are a facet of the "visual languages" themselves. That is, despite their varying countries of origin, Japanese manga, Korean manhwa, and OEL manga all appear to use the JVL grammar, or at least variants ("dialects") of it.

In addition, American indy comics use patterns trending in the direction of Japanese manga, away from American mainstream comics. This perhaps validates their status as a categorically different type of visual narrative in the United States. Their similarities towards books drawn in JVL is also noteworthy, as the contemporary indy movement has emerged in the wake of the mass importation of manga into the American comic market in the 1990s (Mazur and Danner 2014, Goldberg 2010), and many authors may have been influenced by manga more than mainstream American comics.

Altogether, these preliminary results suggest patterned ways that different visual languages use narrative structures. Just like the diversity of grammatical structures found in the syntax of the world's spoken languages, the overall trends between comics of the world likely reflect characteristics of particular narrative grammars for these diverse visual languages. It is worth reinforcing that such findings go against the idea that narratives are structured simply through uniform semantic juxtapositions, whereby no patterned information would be encoded in long-term memory (Bateman and Wildfeuer 2014, Magliano and Zacks 2011). Such cultural trends—and especially similarities between like-systems (Japanese manga and manga-inspired works)—are only possible if creators of such works retain these narrative patterns in memory.

Beyond this variation though, our dataset allowed us to investigate broader trends between narrative patterns, collapsing across individual populations. Are there particular relationships between levels of narrative complexity? We carried out correlations between means for all narrative patterns using an alpha set to .05. R-values are depicted in Figure 37.

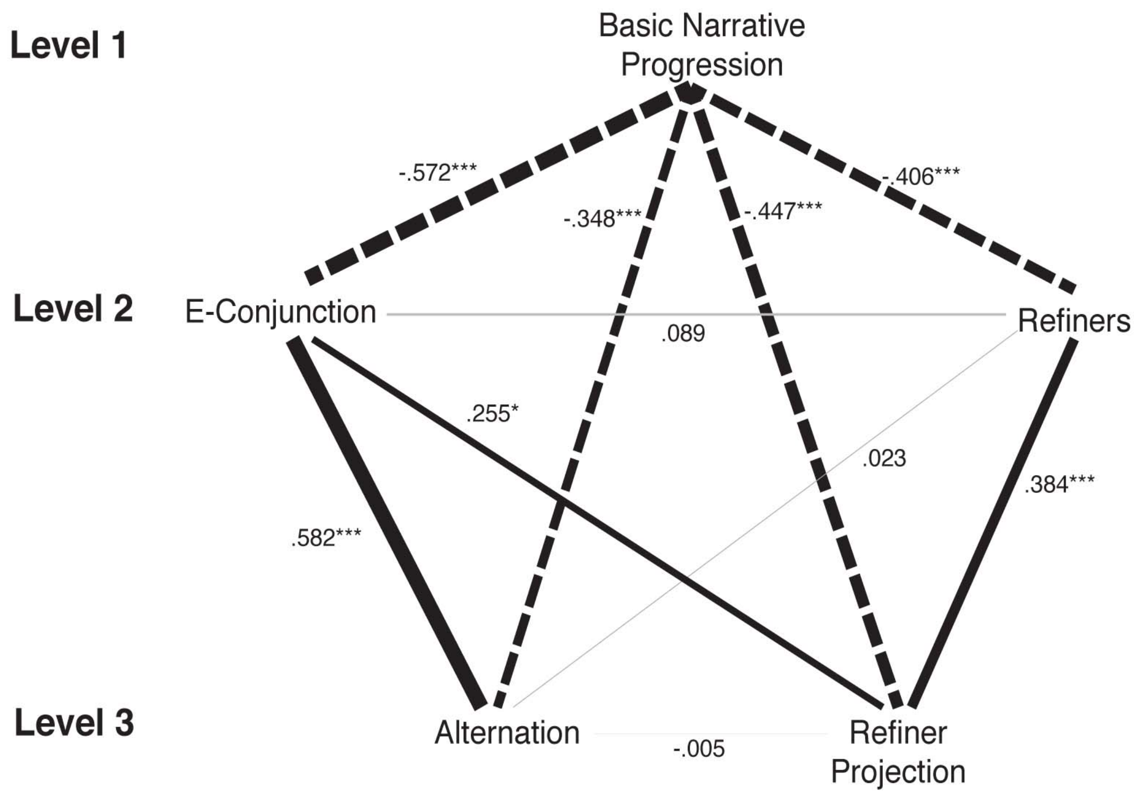

Figure 37. Correlation coefficients between each narrative pattern, collapsing across groups. Dashed line = negative correlation, Solid line = positive correlation, grey = not significant; line weight is equal to r-value. *p<.05, **p<.005, ***p<.001.

Negative correlations appeared between the Basic Narrative Progression and all Level 2 and Level 3 patterns (all ps < .001). This suggested that, across all systems, the BNP decreases with greater narrative complexity (E-Conjunction and Refiners). Within Level 2, E-Conjunction had no correlation with Refiners (p > .05), implying independence in narrative modification. Each of these Level 2 patterns also showed positive correlations with the more complicated Level 3 patterns that use them. E-Conjunction positively correlated with both Alternation and Refiner Projection (all ps < .001). This made sense, because E-Conjunction is used in both of those complex patterns. Similarly, Refiners positively correlated with Refiner Projection (p < .001), despite the intervening panel between the head-modifier structure in this more complex pattern (i.e., Refiners do not directly appear within Refiner Projection). These results implied that more complex patterns (Level 3) share relationships with their component schemas (Level 2).

Nevertheless, we found no correlation between Alternation and Refiners (p = .168) nor between Alternation and Refiner Projection (p = .159). This lack of a relationship suggested that "structural complexity" in a narrative system does not manifest uniformly across all structures. This also aligns with the lack of a correlation between E-Conjunction and Refiners. That is, it appears that these narrative patterns can manifest independently of each other in the conventions of a narrative system, perhaps suggesting alternative paths by which systems can develop complexity.

Finally, given the relationships between levels, it is worth noting that our notion of "narrative complexity" does not necessarily assign greater or lesser "value" to systems depending on their complexity. Rather, "complexity" here is viewed simply as a feature of typological variation, with proportion of patterns providing a descriptor for aspects of diversity between (visual) narrative systems. As in linguistic typology (Croft 2003), it may be the case that less "complex" systems of narrative grammar be "enriched" in other aspects of structure. For example, though Euro-American books relied on less complex narrative patterns, these works may compensate with additional within-panel complexity (similar to the tradeoff often shown between morphological and syntactic structure in verbal languages). This ongoing corpus work can hopefully investigate these other structures, ideally in combination with the type of data sampled here.

Psychological experimentation

The previous section illustrated that VNG can effectively characterize the patterns in various graphic systems. The question then remains: What evidence do we have that producers and readers of narrative draw on this type of narrative grammar in actual comprehension? As VNG is embedded within a linguistic and cognitive system, psychological instantiation remains an important test—perhaps the *main* test—of the validity of the theory. Below, I review the growing number of behavioral and neurocognitive studies that have directly tested the assumptions of VNG and its psychological validity.

First, evidence has supported that readers do update a mental model of a scene as they read a visual narrative sequence. Behavioral work using filmed visual narratives has consistently shown that readers are able to consciously attend to changes in dimensions of characters, spatial locations, and time (Magliano, Miller, and Zwaan 2001, Magliano and Zacks 2011). In addition, growing work has measured the electrophysiological neural response to panels while readers view visual narrative sequences. A brainwave response thought to index mental model updating—the *P600* (Kuperberg 2013)—has been observed across several studies. The P600 appears to be modulated by the changes in characters (Cohn and Kutas 2017), the generation of inference (Cohn and Kutas 2015, 2017), and incongruity to the structure of actions (Amoruso et al. 2013, Sitnikova, Holcomb, and Kuperberg 2008), such as images with omitted or incongruous motion lines (Cohn and Maher 2015). This process of mental model updating also appears to be ongoing and continuous throughout each panel of a visual narrative, not simply a response to incongruity (Cohn and Kutas 2015, Osaka et al. 2014). Such results support the idea that comprehenders do update a mental model of a scene at each image of a visual narrative, in line with theories of discourse (Zwaan and Radvansky 1998).

Nevertheless, if sequential image understanding relied solely on updating meaningful relations between images, we would expect that images would not necessarily play specific roles and behaviors in a sequence: all panels should function in the same ways relative to the ongoing comprehension process. We already saw evidence against this with the corpus data on narrative patterns. If visual narratives relied on only a uniform process of semantic updating, no stored patterns should appear in different systems of the world. Yet, the systematic variation across visual languages suggests that such patterns are indeed stored in memory of creators of visual narratives.

In addition, panels *do* play various roles in a sequence, which can be characterized by differing behaviors on a variety of tasks. For example, some "core" categories (Initials, Peaks) are chosen to be deleted from a sequence less often than more "peripheral" categories (Establishers, Releases), and these same "core" categories are more readily recognized when they are missing (Cohn 2014, Magliano et al. 2016). Also, some categories are more able to be moved around in a sequence than others (Cohn 2014), and different brain responses appear to grammatically appropriate, compared to inappropriate, categories in a sequence (Cohn 2012, Cohn and Kutas 2015). That is, some panels contain content that can flexibly play various narrative roles in a sequence, while other content is more fixed. Such findings contrast the assumption that *any discourse unit* can play a role in *any position* within a narrative (Sternberg 1982). Rather, the degree to which a panel plays different contextual roles depends on particular constraints on its semantic content in relation to a top-down narrative structure. These identifiable trends for various categories support that units play particular roles in the sequence, and thus go beyond just contrastive semantic relations between units.

In addition, a growing literature of psychological research shows that narrative grammar is processed separately from meaning. In one of the first studies of VNG, we compared coherent normal visual narrative sequences and totally scrambled sequences with those that had a well-formed narrative structure but no meaningful relations between images (Cohn et al. 2012). These "narrative-only" sequences had a well-formed narrative arc, but the content of the images had no relationships to each other—analogous to Chomksy's (1965) famous sentence *Colorless green ideas sleep furiously*. When participants pressed buttons to target panels in sequences, we found that their response times to panels in narrative-only sequences fell between the faster response times to panels in normal sequences and the slower times to those in scrambled sequences. Thus, the presence of only a narrative grammar gave an advantage to the processing of these sequences. However, subsequent analysis found a brain response typically associated with semantic processing—the *N400* (Kutas and Federmeier 2011). Here, the N400 did not differ between scrambled and narrative-only sequences. In other words, the brain response for semantic processing was insensitive to the presence of only narrative structure, perceiving the narrative-only sequence similarly to the scrambled sequence. Because the narrative grammar did not modulate the amplitude of the semantically-sensitive N400, despite differences in reaction times, it suggested that semantics and narrative grammar operate on different processing streams.

Subsequent research has found that manipulation of narrative grammar elicits different brain responses than manipulation of meaning (N400) in sequential images (Cohn 2012, Cohn et al. 2014, Cohn et al. 2012). For example, in one study, we inserted blank white "disruption" panels into visual sequences either within the narrative constituents or between the constituents. This design was inspired by classic studies of psycholinguistics (Fodor and Bever 1965) which found that disruptions (auditory clicks) played between syntactic constituents was easier to comprehend than those within constituents. Measuring brainwave responses, we observed *anterior negativities* to disruptions within the constituents in visual narratives compared to disruptions between groupings (Cohn et al. 2014). Similar anterior negativities have been associated with grammatical processing in both language and music (Kaan 2007, Patel 2003). Because the brain response was larger to disruptions within the narrative constituents than between them, it provides evidence that comprehenders make such groupings.

It is worth pointing out that these hierarchic constituents are not recognized on the basis of semantic discontinuity—situational changes in characters, location, or time—between units of visual discourse (Zacks and Magliano 2011). While semantic discontinuity often correlates with breaks between narrative constituents, grammatical categories are more predictive of narrative segmentation than semantic information like situational changes (Cohn and Bender 2017). Furthermore, these anterior negativities—taken to be indicative of grammatical processing (Kaan 2007, Patel 2003)—do not appear to be sensitive to manipulations of meaning, even when evoked in the context of grammatical patterning like conjunction (Cohn and Kutas 2017). Finally, in the aforementioned study of brain responses to constituent structure, we observed a larger anterior negativity to disruptions placed within the first constituent relative to those between constituents (Cohn et al. 2014). In both cases, viewers had not yet crossed the boundary to view the panel after the constituent break. Thus, they had no ability to observe semantic (dis)continuity, because they had not yet viewed the subsequent panel in order to make those semantic connections. Yet, their brain responses to these disruptions differed between within- and between-constituent disruptions, suggesting that they were able to make *predictions* about subsequent constituent structure on the basis of the cues within panels that signal narrative categories, not just the semantic relations between images.

Altogether, these findings support that visual narrative processing is more complex than just observing changes of meaning across individual units. While comprehenders must integrate information into a growing mental model of a scene, as supported by theories of semantic updating (Zacks and Magliano 2011), that information is mediated by a hierarchic narrative grammar that organizes semantic information in coherent, patterned ways. This narrative grammar appears to be largely unconscious to a comprehender; while experimental participants will consciously describe semantic aspects of sequences in post-experiment questionnaires, they rarely recognize aspects of this narrative grammar (Cohn and Bender 2017, Cohn et al. 2012). In addition, the brain responses elicited by these structures are consistent with those found to other domains, like language and music (Kaan 2007, Patel 2003). This suggests that visual narrative processing taps into general aspects of cognition that operate across domains (Cohn 2013a).

Finally, recent research has suggested that the processing of visual narratives is not uniform, as might be suggested by a view of semantic mental model updating. Rather, processing varies based on the patterns found in the structures of the particular "visual language" that a person reads—as indicated by the corpus analyses in the prior section. We recently examined participants' brainwaves to sequences that did or did not use E-Conjunction (Cohn and Kutas 2017). Across all participants, we found two neural responses: An anterior negativity suggesting combinatorial processing to the grammatical patterning, and a P600 suggesting a process of mental updating for the inference of combining different characters into a common spatial environment. These results were consistent with the idea that people use both mental model updating *and* a narrative grammar in the comprehension of sequential images.

However, a regression analysis showed further that these brain responses were modulated by participants' experience with reading manga while growing up. Frequent manga readers showed more combinatorial processing and less updating (larger anterior negativities, smaller P600s), while infrequent manga readers showed more updating and less combinatorial processing (larger P600s, smaller anterior negativities). As discussed above, Japanese manga use E-Conjunction more often than typical American and European comics. Such results imply that, because of their experience with this pattern, manga readers are thus able to rely more on combinatorial processing, while their relative inexperience leads non-manga readers to rely more on semantic processes of mental model updating. Such results are further intriguing given that these stimuli did not look like manga: they were manipulated American comic strips (*Peanuts*) that nevertheless used patterns similar to those in manga (E-Conjunction). Thus, these results suggest that not everyone processes visual narratives in the same way, but rather such processing is modulated by readers' familiarities with patterns found in particular visual languages of the world.

Conclusion

Overall, this paper has introduced a theory of narrative grammar grounded in the methods and tools of the linguistic and cognitive (neuro)sciences. By explicitly stating the representations and schemas involved in this structure, VNG reveals complexities of narrative structure beyond generalizations. This theoretical model can reveal differences in structure across cultures via corpus analyses, and can provide the basis for explicit and testable hypotheses for experimentation. Altogether, this combination of theoretical, corpus, and experimental research allows for important insights on the complexity of narrative and its cognitive foundations.

Finally, given that narrative structure appears across domains, VNG can offer insights into the structure of narrative in other modalities, such as film and discourse (Cohn 2013b). Just as broader meaning-making processes operate across domains (Magliano, Higgs, and Clinton This volume), narrative grammars must adapt to the affordances of different modalities (Cohn 2013a, 2016a), and such cross-domain application is already being examined (Cohn 2016a, Amini et al. 2015, Kim and Monroy-Hernandez 2015, Barnes 2017). Such formalization also allows us to characterize complexity in narratives that balance structures across modalities. Indeed, multimodality in contexts like comics or films involves not only the balancing semantics dominated by one modality or another, but also the presence or absence of narrative structures in either one or multiple modalities (Cohn 2016b). Such complexity is perhaps the pinnacle of research in narrative, and provides a worthy goal towards future research.

References

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.

Amoruso, Lucia, Carlos Gelormini, Francisco Aboitiz, Miguel Alvarez González, Facundo Manes, Juan Cardona, and Agustin Ibanez. 2013. "N400 ERPs for actions: Building meaning in context." *Frontiers in Human Neuroscience* 7. doi: 10.3389/fnhum.2013.00057.

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.

Bateman, John A., and Karl-Heinrich Schmidt. 2012. *Multimodal film analysis: How films mean*. New York: Routledge.

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.

Bordwell, David, and Kristin Thompson. 1997. *Film Art: An Introduction*. 5th Edition ed. New York, NY: McGraw-Hill.

Carroll, John M. 1980. *Toward a Structural Psychology of Cinema*. The Hague: Mouton

Chomsky, Noam. 1965. *Aspects of the Theory of Syntax*. Cambridge, MA: MIT Press.

Cohn, Neil. 2011. "A different kind of cultural frame: An analysis of panels in American comics and Japanese manga." *Image [&] Narrative* 12 (1):120-134.

Cohn, Neil. 2012. "Structure, meaning, and constituency in visual narrative comprehension." Doctoral Dissertation, Psychology, Tufts University.

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

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

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

Cohn, 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. 2016a. "From Visual Narrative Grammar to Filmic Narrative Grammar: The narrative structure of static and moving images." In *Film text analysis: New perspectives on the analysis of filmic meaning.*, edited by Janina Wildfeuer and John A. Bateman, 94-117. London: Routledge.

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

Cohn, Neil, and Patrick Bender. 2017. "Drawing the line between constituent structure and coherence relations in visual narratives." *Journal of Experimental Psychology: Learning, Memory, & Cognition* 43 (2):289-301. doi: http://dx.doi.org/10.1037/xlm0000290.

Cohn, 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, 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, Amaro Taylor-Weiner, and Suzanne Grossman. 2012. "Framing Attention in Japanese and American Comics: Cross-cultural Differences in Attentional Structure." *Frontiers in Psychology - Cultural Psychology* 3:1-12. doi: 10.3389/fpsyg.2012.00349.

Croft, William. 2003. *Typology and Universals*. 2nd ed. Cambridge, UK: Cambridge Univesity Press.

Culicover, Peter W., and Ray Jackendoff. 2005. *Simpler Syntax*. Oxford: Oxford University Press.

Fodor, Jerry, and Thomas G. Bever. 1965. "The psychological reality of linguistic segments." *Journal of Verbal Learning and Verbal Behavior* 4 (5):414-420.

Freytag, Gustav. 1894. *Technique of the Drama*. Chicago: S.C. Griggs & Company.

Goldberg, Adele. 1995. *Constructions: A Construction Grammar Approach to Argument Structure*. Chicago, IL: University of Chicago Press.

Goldberg, Wendy. 2010. "The Manga Phenomenon in America." In *Manga: An Anthology of Global and Cultural Perspectives*, edited by Toni Johnson-Woods, 281-296. New York: Continuum Books.

Jackendoff, Ray. 1990. *Semantic Structures*. Cambridge, MA: MIT Press.

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.

Kim, Joy, and Andres Monroy-Hernandez. 2015. "Storia: Summarizing Social Media Content based on Narrative Theory using Crowdsourcing." *arXiv preprint arXiv:1509.03026*.

Kuperberg, Gina R. 2013. "The pro-active comprehender: What event-related potentials tell us about the dynamics of reading comprehension." In *Unraveling the Behavioral, Neurobiological, and Genetic Components of Reading Comprehension*, edited by B. Miller, L. Cutting and P McCardle, 176-192. Baltimore: Paul Brookes Publishing.

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.

Magliano, Joseph, Karyn Higgs, and James A Clinton. This volume. "Sources of Complexity in Comprehension Across Modalities of Narrative Experience." In *Narrative Complexity and Media: Experiential and Cognitive Interfaces*, edited by Marina Grishakova and Maria Poulaki. Lincoln: University of Nebraska Press.

Magliano, Joseph P., Kristopher Kopp, Karyn Higgs, and David N. Rapp. 2016. "Filling in the Gaps: Memory Implications for Inferring Missing Content in Graphic Narratives." *Discourse Processes*:0-0. doi: 10.1080/0163853X.2015.1136870.

Magliano, 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., Jason Miller, and Rolf A. Zwaan. 2001. "Indexing space and time in film understanding." *Applied Cognitive Psychology* 15:533-545.

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.

Mandler, Jean M., and Nancy S. Johnson. 1977. "Remembrance of things parsed: Story structure and recall." *Cognitive Psychology* 9:111-151.

Mazur, Dan, and Alexander Danner. 2014. *Comics: A Global History, 1968 to the Present*. London: Thames & Hudson.

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

McCloud, Scott. 2006. *Making Comics*. New York, NY: Harper-Collins.

Metz, Christian. 1974. *Film Language: A Semiotics of the Cinema*. Translated by Michael Taylor. New York: Oxford University Press.

Osaka, Mariko, Ken Yaoi, Takehiro Minamoto, and Naoyuki Osaka. 2014. "Serial changes of humor comprehension for four-frame comic Manga: an fMRI study." *Scientific Reports* 4. doi: 10.1038/srep05828.

Patel, Aniruddh D. 2003. "Language, music, syntax and the brain." *Nature Neuroscience* 6 (7):674-681. doi: 10.1038/nn1082.

Sitnikova, Tatiana, Phillip J. Holcomb, and Gina R. Kuperberg. 2008. "Two neurocognitive mechanisms of semantic integration during the comprehension of visual real-world events." *Journal of Cognitive Neuroscience* 20 (11):1-21.

Sternberg, Meir. 1982. "Proteus in Quotation-Land: Mimesis and the forms of reported discourse." *Poetics Today* 3 (2):107-156.

van Dijk, Teun, and Walter Kintsch. 1983. *Strategies of Discourse Comprehension*. New York: Academic Press.

Zacks, Jeffrey M., and Joseph P. Magliano. 2011. "Film, narrative, and cognitive neuroscience." In *Art and the Senses*, edited by D.P. Melcher and F. Bacci, 435-454. New York: Oxford University Press.

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