# Generating Abstract Comics

Chris Martens and Rogelio E. Cardona-Rivera

North Carolina State University
{crmarten,recardon}@ncsu.edu

**Abstract.** We investigate a new approach to comic generation that explores the process of generating the contents of a panel given the contents of all previous panels. Our approach is based on leading discourse theories for comics by McCloud (panel transitions) and Cohn (narrative grammar), unified by cognitive theories of inference in visual language. We apply these theories to comics whose panel paremeters are abstract geometric shapes and their positions, contributing a computational realization of McCloud's and Cohn's comics theories, as well as a modular algorithm that affords further experimentation and evaluation of visual discourse theories.

**Keywords:** intelligent narrative technologies, comics, narrative generation

## 1 Introduction

Interactive digital storytelling has traditionally been concerned with text-based modes of discourse. Despite the fact that the word *fiction* makes no commitments to medium (book, film, play, et cetera), the term *interactive fiction* is used nearly synonymously with text-based digital storytelling. Meanwhile, comics, a relatively unexplored domain of computational narratology [**?**], present a wide range of expressive opportunities for interactive storytelling not afforded by text. Interactive comics invite many of the same questions addressed by interactive textual narrative research: what modes are available for a machine to tell visual stories in collaboration with a human player or author? How can algorithms introduce novel, generative content into a partially-constructed comic? Which decisions will the human and algorithmic components have available to them?

This work begins to address these questions by exploring generation of *purely visual*, *abstract* comics. "Purely visual" means that panels do not use words and text to communicate narrative, as in Figure **??**. "Abstract," as in abstract art, means that, aside from repetition of elements with a shape, color, and size, and spatial placement of said entities within panels, we do not intentionally ascribe literal meaning to the components of generated panels. Our purpose in choosing an abstract representation is partially due to ease of implementation, but partially also to separate issues of *structure* of comics (which can be perceived with abstract representations) from issues of contextual, reference-laden meaning that would be found especially in anthropomorphic figures.

Two leading theories on comics have examined constraints on sequences of panels and the relationships among them. McCloud [**?**] proposes a taxonomy of *transitions* between a panel and the one following it, enumerating six different ways readers might make sense of the connection between two consecutive panels "across the gutter" (gap between panels). Cohn [**?**], on the other hand, eschews transitions as a basis for comic structure and asserts that grammar-like syntax trees govern the formation of legible comics. We present a small-scale computational system [**?**] that operationalizes a hybrid of these two approaches. Our primary takeaway is that both *global and local* reasoning contribute important techniques to narrative generation: local reasoning is important for maintaining narrative coherence, and global reasoning is important for maintaining satisfying narrative structure. Both are thus important parts of creating comprehensible comics, and we present an outline of future work designed to explore the human interpretation of our generated artifacts.

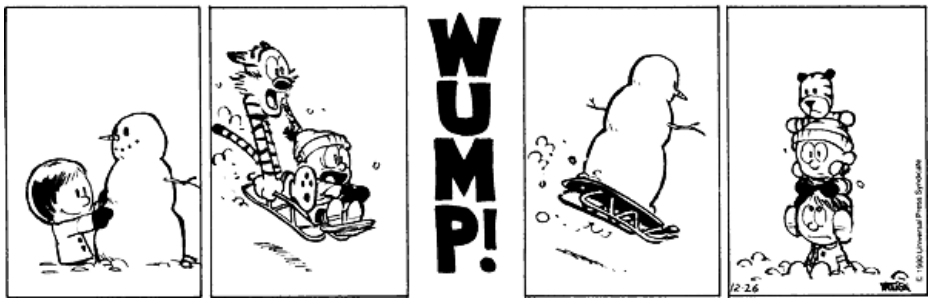


**Fig. 1.** This wordless Calvin and Hobbes comic strip (© Bill Watterson) exemplifies the kind of output we are targeting with our generator. The strip also illustrates how little plot structure informs this kind of short-form visual storytelling.

## 2 Comics Theory

The basis of McCloud's theory about making sense of panel sequences across the gutters, later validated experimentally, is that readers of comics optimize their consumption of relevant information [**?**], and work to construct inferences [**?**] about story content in these liminal spaces of discourse. Inferences for story content are constructed when they are needed for comprehension and enabled by what has been narrated thus far [**?**]. The dynamic between story authors and audiences parallels the dynamics of people engaged in cooperative conversation as outlined by the philosopher of language Grice [**?**]: the storyteller, as the active contributor to the ongoing communicative context, is expected to make her contributions to the discourse based on what is relevant to her narrative intent. McCloud's introduced six *panel transition types* for comics [**?**], enumerate the

different roles that the reader may infer from a well-written comic. These transition types are *moment-to-moment*, *action-to-action*, *subject-to-subject*, *aspect-to-aspect*, *scene-to-scene*, and non sequitur. While it is tempting to think we could simply operationalize these transitions in a generator, as Cohn [**?**] (Chapter 4) points out, so much of their meaning relies on contextual, real-world-situated understanding that it lends little help to computational authoring.

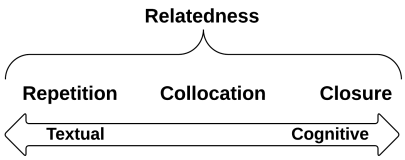


**Fig. 2.** The spectrum of *relatedness* as discussed by Saraceni [**?**]. Relatedness indicates how comic panels are connected or associated in the minds of readers, spanning from textual factors to cognitive factors. Along that spectrum, there are three distinguished categories of relatedness: *repetition*, *collocation*, and *closure*, which have demonstrably different effects on the construction of narrative mental models.

However, Saraceni [**?**] supports McCloud's hunch that readers create meaning from comics from perceived relationships between panels. Saraceni describes three notions of *relatedness* between comic elements, which are the building blocks from which readers may construct meaning inferences. Relatedness creates a comic's *cohesion* – the lexico-grammatical features that tie panels together – and *coherence* – the audience's perception of how individual panels contribute to her mental model of the unfolding events. Relatedness emerges from a spectrum of *textual*[1] factors to *cognitive* factors, illustrated in Figure **??**. Closer to the textual end of the spectrum is the *repetition* of visual elements across panels. Beyond repetition is *collocation*, which refers to an audience's expectation that related visual elements will appear given the ones that have been perceived. Closer to the cognitive end of the spectrum is the *closure* over comic elements, which refers to the way our minds complete narrative material given to us. Closure is terminologically borrowed from the field of visual cognition, but is intended as the mental process of inference that occurs as part of an audience's *search for meaning* [**?**]. The comic in Figure **??** depends on these three aspects of relatedness: first, repetition of the sled, snowman, and other figures maintains cohesion across panels. Second, the humor of the sled carrying off the snowman depends on our (non-grammatical) domain knowledge that the snowman is not a living character in the same sense as the other figures. Finally, the comic depends on closure for the audience to "fill in the gaps" to infer what must have happened during the third panel.

[1] *Textual* here does not mean the use of actual text, but rather is a shorthand for *surface code* [**?**].

## 3 Related Work

In terms of visual storytelling, Heider and Simmel [**?**] formulated an experiment based on abstract shapes in animation and evaluated whether audiences perceived consistent stories. Their work also depends on cognitive closure to achieve a narrative goal; however, the animations themselves were hand-authored rather than generated. Pioneering work on visual discourse by Jhala and Young [**?**] addresses visual storytelling in terms of camera control for cinema-style storytelling. We consider the representation and generation strategies emerging from this work [**?**] as strong candidates for future work in terms of scene specification; however, we look to comics as a simpler, more discretize domain that does not depend on a deeply simulated underlying narrative. Pérez y Pérez et al. [**?**] developed a visual illustrator to the MEXICA system [**?**], and verified the degree to which their 3-panel comic generator elicited in readers the same sense of story as a textual realization of the same MEXICA-generated plot. While this system also follows the pipeline model of narrative generation, we see their work as complementary: they developed an experiment methodology through which it is possible to empirically assess if their palette of designed visual elements denote story concepts as intended. Future work in comic generation will have to address this point going forward, and Pérez y Pérez et al. provide a step toward understanding the gap between story concepts and the pictorial symbols meant to encode them.

## 4 System Description

Our approach to generating visual narratives begins as a linear process that selects next comic panels based on the contents of previous panels, choosing randomly among indistinguishably-valid choices. The concepts we represent formally are *transitions*, *frames*, and *visual elements*, which we define below. There are two levels on which to make sense of these terms: the symbolic level, i.e. the intermediate, human-readable program data structures representing a comic, and the rendered level, designed to be consumed by human visual perception.

A **visual element (VE)** is a unique identifier from an infinite set, each of which is possible to map to a distinct visual representation. We do not explicitly tag visual elements with their roles in the narrative, such as characters, props, or scenery, making the symbolic representation agnostic to which of these narrative interpretations will apply. In the visual rendering, of course, our representation choices will influence readers' interpretation of VEs' narrative roles. A **frame** is a panel template; at the symbolic level, it includes an identifier or set of tags and a minimum number of required visual elements. The reason a frame specifies a minimum number of VEs is to allow for augmentation of the frame with pre-existing elements: for example, the *monologue* frame requires at least one visual element, indicating a single, central focal point, but other visual elements may be included as bystanding characters or scenery elements. At the rendering level, a frame includes instructions for where in the panel to place supplied VEs. A **panel** is a frame instantiated by specific visual elements. Finally, a **transition** is

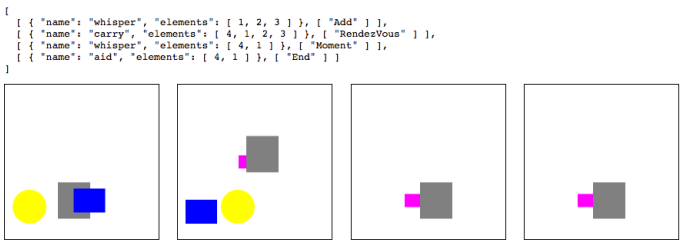


**Fig. 3.** Example of generator output. While the narrative here is ambiguous, we suggest the following readings: the repetition of the grey (largest) rectangle in every frame suggests it as a focal point, and the sudden appearance of the pink (smallest) rectangle suggests an interloper removing the grey rectangle from its initial context (established by the blue rectangle and yellow circle). Together with the names of the frames (reported in symbolic form above the comic), we can read the sequence as follows: the grey rectangle whispers to the blue rectangle, then is carried off by a pink rectangle, who whispers to the grey rectangle and then aids the grey rectangle.

a specification for how a panel should be formed as the next panel in a sequence. We took inspiration from McCloud transitions [**?**], developing a more syntactic notion defined purely in terms of frames and (abstract) visual elements, for which Saraceni's theory of relatedness [**?**] could be applied. For example, while McCloud could refer to an action-to-action transition as one where a character is depicted carrying out two distinct actions, we have no notion of *character* and *action* (these being semantic and contextual categories), so instead must refer to which visual elements appear, where they have appeared previously, and what their spatial relationships might be (potential frames). The rendering of a frame itself may position VEs in such a way that an audience would read certain actions or meaning into it; however, this kind of audience interpretation is not modeled to inform generation. Thus, we introduce six formal transition types: **Moment** transitions retain the same set of VEs as the previous panel, changing only the frame. **Add** transitions introduce a VE that didn't appear in the previous panel, but might have appeared earlier (or might be completely new). A new frame may be selected. **Subtract** transitions remove a VE from the previous panel and potentially choose a new frame. **Meanwhile** transitions select a new frame and show *only* VEs that did not appear in the previous panel, potentially generating new VEs. **Rendez-vous** transitions select a random subset of previously-appearing VEs and selects a new frame to accommodate them. We implemented our generator in OCaml and additionally implemented a front-end, a web-based renderer.[2] The renderer assigns each frame type to a set of coordinates given by percentage of the vertical and horizontal panel size, and then renders panels by placing visual elements at those coordinates. Visual elements are represented by randomly generated combinations of size, shape (circle or rectangle), and color. An example of the generator's output can be seen

[2] http://go.ncsu.edu/comicgen