Two-Fisted Comics Generation:

Comics as a Medium and as a Representation for Creative Meanings

Tony Veale,
School of Computer Science, University College Dublin, Ireland.
Tony.Veale@UCD.ie

Abstract

Every expressive medium allows us to ground meaning in different ways. Comics, or the so-called 9^{th} *art* (after film and TV), are sequential integrations of words and images that offer more possibilities than either words or images alone can offer. Like any art form, comics settle into clichéd norms – panels, balloons, tails – and give rise to genres that dominate at the expense of all others, such as the superhero genre. Yet comics can also use a vivid emotional expressivity to imbue physical actions with potent feelings, lending intuitive immediacy to the gamut of human concerns. This paper considers comics as both a medium for automated story-telling and as a meaning representation from which machines can shape specific meanings for particular audiences. We propose two XML standards for communicating via comics: one to define an underlying narrative, and another to define the comics derived from it. The latter uses a repository of visual assets to convey actions and emotions, placing posable characters against a backdrop of stock settings. We show how a combinatorial approach accommodates all of the outputs of an automated story-generator, and also explore how it adapts to the very human exchanges of an online debate, such as the fractious Twitter debate on vaccines. Comics appear to be a practical medium in which to make targeted interventions into a polarizing debate, to present opposing points of view to each side.

See You In The Funny Pages

Although frequently packaged in a disposable form, comics have been described as a *sequential art* by Eisner (1985) – for whom the Eisner prize in comics is named – and as *the ninth art* by Maurice De Bevere, the creator of *Lucky Luke*. At its simplest, a comic strip is a sequence of framed snap shots, called panels, separated by thin bands of whitespace, called gutters. Each panel is usually square or rectangular, typically framed in a black border, and sometimes labeled with a caption above or below the scene depicted within. A typical panel contains a mix of textual and visual elements, to depict a specific action in a certain setting, and to record any words that are spoken (or privately thought) in context. Those words are most often contained within text balloons, either rounded speech balloons or fluffy, cloud-like thought balloons, whose tails link them to the vocalizing characters.

These conventions have become the stuff of cliché, but as McCloud (1993) has shown, this normative grammar of comics allows for a great deal of divergence and creativity. Indeed, even text balloons can vary tremendously from one context to another, to shape meaning as well as just contain it (Forceville *et al*., 2010). Although the ease with which children adapt to the medium's mechanisms allows some to dismiss it as juvenile, this ease also reflects the depth of the cognitive foundations in which the medium is rooted (Cohn, 2013; Cohn & Magliano, 2020). For instance, one intuitively reads a comic strip in the order one reads a text, from top to bottom and left to right in the West, and from right to left and back to front in the East. No one needs to be taught how to read a comic strip. Readers simply adapt to the blended medium as a new form as visual reading.

This work explores the automated generation of comic strips as containers and communicators of meaning. While the marriage of visual and textual forms makes comics an ideal medium for computational creativity, our aim is to do more than produce comic-shaped outputs that readers may find attractive in their own right. Rather, our aim is to use the comic strip as a means of communication in which we pour meaning from one kind of container – such as a text – into another – a comic combining images *and* words – with no, or little, loss of meaning. Our goal is not an end-to-end production of comics in which interior levels of meaning go unexposed to scrutiny and manipulation by the producer, but a controlled, meaning-preserving translation from one explicit representation of meaning to another. To this end, we present a comics-generator that works with the outputs of an automated story-generator, translating each tale from a purely textual form to a vivid blend of words and images.

Although comics are entertaining in their own right, we explore a practical use of the medium here. Consider why they are called 'comics' or 'funny-books' in the first place: the name is a carry-over from the earliest newspaper strips in which short, whimsical diversions were illustrated (the first American "comic book", *Famous Funnies*, repackaged these newspaper funny pages as a standalone periodical). Even serious comicbooks – which some now call *Graphic Novels* – still carry traces of the comical and the unserious. The larger context of this work makes use of these vestiges to package polarizing and perhaps unwelcome meanings in more welcome and disarming forms. Those meanings arise in heated online debates, such as the Twitter debate around vaccines, in which disputants on each side show a tendency to dig in, tune out and listen only to those on the same side. Machines can help to break down these echo chambers by making targeted interventions into the debate, using comics to summarize and distill the main arguments on both sides. As a first step, we will examine how well a framework for producing comics from computer-generated tales can also produce comics from these arguments, preserving the gist of each argument and the gist of any given user's position.

With these goals in mind, the rest of the paper assumes the following structure. After exploring some related work and ideas in the next section, we present our combinatorial approach to comics production, which maps from an XML schema for machine stories to an XML schema for comics. We next consider applications of this mapping of XMLs, in dialogue-driven comics production and online intervention. For the latter, comics must be attuned to the dynamics of a debate as reflected in a representative dataset, so we model the debate via statistical analysis of a large Twitter corpus. Our analysis of the vaccine debate will show that a comics creator that is attuned to a sufficiently rich story creator is capable, through the liberal use of visual metaphor, to also accommodate the diverse arguments of a topical debate.

Related Work and Ideas

Comics are a sequential art that requires a narrative impetus. A generator can produce this impetus for itself, by creating its own stories, or it can acquire its narratives from another source, such as an existing story-generator, or from another medium, such as film (by e.g. reusing film scripts), theatre, online discussions (e.g., chatrooms, Twitter), or games. For example, a comic narrative might visualize the sequence of moves in a chess game as a sequence of real-world actions (Gervás, 2014), or mirror the sequence of moves in a video game. In the latter, game screenshots might also be used to provide the visual contents of the comic's panels.

The *Comic Chat* system of Kurlander *et al*. (1996) takes its narrative impetus from chatroom interactions, and turns those textual discussions into comic strips, filling one panel per conversational turn. Each interacting user is assigned a stock comic figure, such as a lantern-jawed jock, an exotic princess, or an anthropomorphic animal, where each figure has a small number of facial expressions and neutral poses. Those expressions, used to convey the basic emotions, are determined via a sentiment analysis of a user's contribution to a turn, which floats overhead in a text balloon. Because this narrative impetus tracks the human inputs, *Comic Chat* is free to focus on the technical craft of the medium, and it shows a firm grasp of the grammar of comics. It uses long-shots to start a conversation, and close-ups for subsequent turns. All figures, balloons and tails are intuitively ordered within a panel to ensure ease of reading from left to right, and a small number of backgrounds is used consistently to maintain continuity from one panel to the next.

The *Comics2D* system of Alves *et al*. (2007) builds its comics from the storyworld representations of a generator of dramatic fiction, such as that of Cavazza *et al*. (2003). If a fiction generator does more than generate narative texts, and also provides a visual representation of its story-world, a comics generator can tap into this visual model too, to fill its panels with snapshots of the story. Comics2D uses its own XML representation of a comic, via a schema it calls CSDL (or *Comic Strip Description Language*). The schema defines nodes for each of the principal elements of a comic, from panels and scenes to backgrounds and characters, and also explicitly tags what happens in the transitions between panels. *Comics2D* is a modular system that allows users to plug-in alternate renderers, and it should support any story-generator than works with CSDL, although the relationship between renderer and generator is typically a complex one.

A comic strip is a sequence of snapshots held together by the driving logic of a story, but this logic often lies hidden in the gutters between panels. Data-rich machine learning approaches can teach a machine to infer this logic, so that it can predict for itself what should come next in the story. To this end, Iyyer *et al.* (2017) have created their COMICS dataset from 1.2M comic panels, for which the text within is automatically transcribed. They estimate that most panel transitions are either *action-to-action* (~34%) or *subject-to-subject* (~32%), while ~17% extend a conversation, ~14% shift from one scene to another, and less than 1% illustrate the moment-to-moment dynamics of a single action. Iyyer *et al*. train a hierarchical network of LSTMs to derive a context model for a given sequence of panels, and use this model to score candidates for the words and images in the subsequent panels. Although the model still underperforms humans, it showcases the value of a multi-panel context and a multimodal integration of visual *and* textual features.

Such a model might, in principle, also generate the next panel, and not just prefer one or another panel continuation. Melistas *et al*. (2021) use *two* neural architectures to create comics in the style of Japanese manga: a language model, *GPT-2*, to produce the text of each panel, and a generative adversarial network, *StyleGAN2*, to synthesize the images. As with Iyyer *et al*., they create a dataset of manga comics for which textual transcriptions are automatically extracted. Text areas are then inpainted to yield text-free training data for the image synthesizer, whose training is further boosted with images of Japanese anime. Low-resolution outputs are then subsequently refined with a trained upscaling network. The approach is suggestive and highly experimental, as the generative models for text and image operate independently of each other, to produce sequences of images that have no pre-conceived relation to the text balloons that adorn them.

Agrawal *et al*. (2016) show it is feasible to make comics from a jumble of captioned photos, by using a mix of word and image features to infer the most natural narrative order. Those photos can, in turn, serve as a basis for generating a comics-style rendering for each photo/panel. Yet, as useful as these rich datasets are for machine-learning approaches, they lack a key dimension: a sense of the underlying story that gives the images and text their narrative momentum.

A Combinatorial Approach

Melistas *et al*. use a neural blackbox to generate the textual content of a comic strip in a single pass, so its output is a surface form that lacks an overt deep structure. A generator that produces XML-tagged surface forms can be subjected to tests of well-formedness and schema-specific validity, so that ill-formed or invalid outputs can simply be resampled, but raw surface forms can offer no such affordances. Multi-pass approaches that first generate a deep 'fabula' structure before producing the surface rendering of the story – as text for a narrator and dialogue for the characters – offer even more control to the creator of comic strips. This fabula can be mapped to the panel layout of the comic, while narration text can provide the panels' captions, and dialogue text can provide the contents of the panels' balloons. What may now seem like old-fashioned approaches to story-generation are thus well-placed to support step-by-step comics production.

Any story-generator that provides an explicit fabula built from a fixed inventory of action types, a surface rendering of the story, and dialogue for its characters, is well suited to automatic comic generation. While there are many story-generators that potentially fit this bill, from that of Cavazza *et al*. (2003) to Montfort *et al*. (2013) and Gervás (2014), we make use of the *Scéalextric* system here (Veale, 2017). *Scéalextric* structures each story as a sequence of "beats." As defined by Gaiman (2021), a *beat* is the smallest unit of action in a story or comic, a discretely resolvable event that is worthy of textual or visual rendering. A beat focalizes a moment in time, and is the ideal unit of comics structure. Each *Scéalextric* beat comprises a single two-person action from its inventory of 800 distinct types, which encompass the realms of romance, crime, medicine, politics, business, religion and war. Each beat is rendered as a 3rd-person view of the event, which can serve to caption the corresponding panel, and provides spoken (or internal) dialogue for each of the two participants, which can fill the panel's balloons.

We define a scene as a sequence of beats that focalize the same characters. A scene may juggle several characters, but each beat will focus on just two at a time. As the characters move in and out of a setting, the scene changes. A dramatic change typically tracks a change in location, and a change in background for the corresponding comic panels. Such a change may warrant a visual flourish such as a splash panel or an establishing shot for the new location, while a lower key shift may warrant just a close-up on a single character.

To tease apart the varying concerns of story and comic, we define two XML formats, one for each. *ScéalXML* is the schema that captures the nested layers of a *Scéalextric* tale, embedding the textual substance of narration and dialogue within a beat structure that also defines fabula-level events. This schema can, in principle, be used to encode the stories of other generators, either *as is* or with some extensions, so that the same mapping to *ComiXML*, and then onwards to a rendered comic strip, can serve those other generators also. *ComiXML* is the schema that encodes a comics-level view of the same events. *ScéalXML* composes scenes from beats, while *ComiXML* composes analogous chapters from panels. Although designed to mirror *ScéalXML* in the first instance, *ComiXML* has obvious overlaps with the CSDL schema of Alves *et al.* (2007), and with the CBML (or *Comic Book Markup Language*) of McIntosh (2005) and Walsh (2012).

In addition to parallel scene/chapter and beat/panel nodes, *ScéalXML* & *ComiXML* also define character/figure nodes. In a story, a <*character*> element names a recurring entity that participates in one or more beats as either the *agent* or the *patient* of an action. Nodes of this type also specify long and short names for a character, as well as the pronouns for referencing them obliquely. *Scéalextric*'s cast of thousands comprises well-established personae from fact and fiction, such as Cleopatra, Darth Vader, Bill Gates and Maleficent. This adds colour, and a potential for humorous incongruity, to the textual outputs of *Scéalextric*, but it poses a challenge for any comics application that must render them visually.

This challenge is two-fold: we need to render characters in ways that are recognizable, or at least differentiable, and we need to render them in emotionally expressive poses that reflect our intuitions of how one performs a specific action. We also need to imbue the outputs with an inherently zany or whimsical charm; a comic may have a serious intent, but to disarm a skeptical audience it must also appear flippant. We meet this challenge with a bespoke set of visual assets called *funny-bones*. Unlike those of Kurlander *et al.* (1996), which attach emotive heads to neutral bodies, these assets integrate a strong emotion with a vivid pose, since a feeling is not separable from the actions that cause it. Each funny-bone has a large expressive head with short rubber limbs, and modifiable hair, skin and lips that allow each to depict a male or female *Scéalextric* persona. For instance, Cleopatra is pre-defined with long black hair, olive skin and red lips, Julius Caesar is given short white hair, pale skin and pink lips, and a bald or nearly bald character, such as Joe Biden, is simply given short, skin-toned hair. We do not shoot for accuracy, just a cartoonish sense of who these people are.

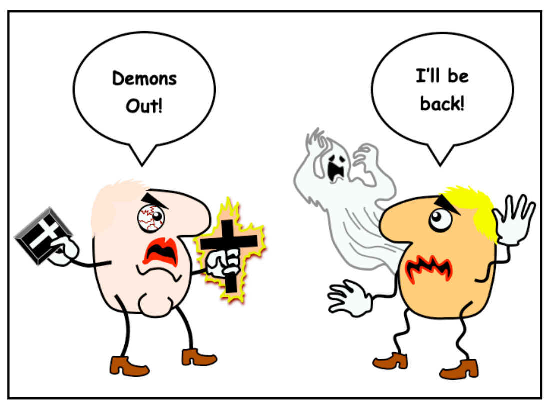

Fig. 1. A single panel with two funnybone figures, configured as *Joe Biden* and *Donald Trump*, in poses *exorcising* and *exorcised*.

Fig. 1. presents a panel for the single *Scéalextric* story beat, <*Joe_Biden exorcises Donald_Trump*>. Its two figures face each other against a blank setting, and are configured with appropriate hair, skin and lip values. In the XML encoding of this <*panel*> node, each is also specified with a location (*left* or *right* of panel), orientation (facing *left* or *right*), and a balloon node (*speech* or *thought*) with an apt text filling.

We consider the emotions expected of the agent and patient roles for each of the 800 *Scéalextric* actions, and how best to render them in mid-action. This leads us to produce 230 funny-bone assets – e.g., an angry figure attacking, a scared figure running away, a strident figure with a puffed chest – and assign one or more to each role of all 800 actions. These 230 assets serve, for the most part, as visual metaphors that concretize and exaggerate the action. So, destructive figures brandish giant hammers or axes; doctors wield scalpels and syringes, and mad scientists cackle as rats scurry underfoot. Zealous figures thump bibles or hurl fireballs, and sick ones slump in wheelchairs as sad ones stand under storm clouds. The assets favour immediacy over nuance, drama over tact. The same is true of the 100 backdrop images that we create to anchor an actor to the domain of their action. Backdrops include hospitals (interior and exterior), dungeons and labs, churches and offices, outdoor scenes (parks, farms, streets), edifices (police stations, court houses, banks, government buildings, jails) as well as battlefields, cafés, gyms and bars. These are mapped both to *Scéalextric* actions and to specific figure assets. The first mapping allows the system to pick a backdrop for a pre-defined action, and the second allows it to infer a suitable choice for any arbitrary pairing of figures. For instance, a stalker lurking in the bushes (*stalking* pose) must be placed against an outdoor scene, not an indoor one.

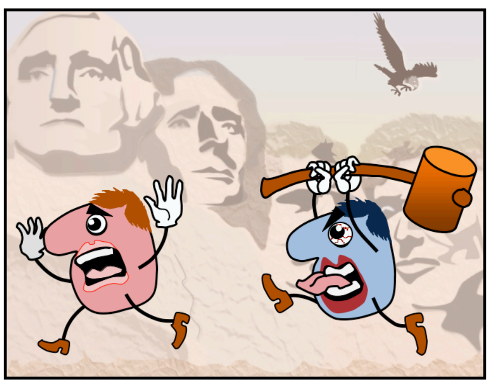

Fig. 2. A panel with "*Mount Rushmore*" as its backdrop. The two figures are colour-coded to visually represent *red* and *blue* states.

The <*setting*> element is used to add a background asset to a panel in ComiXML. As illustrated in Fig. 2, which shows one panel of a comic for the hashtag *#DemsAreDestroying America*, backdrops add another degree of metaphoricity to a panel's meaning. Here the background is a character in its own right, with the *Mt. Rushmore* setting used to depict the USA; the two figures are further colour-coded to represent *blue states* (i.e. "Dems") and *red states* (i.e. "not Dems"). A backdrop can be used literally, to denote a real setting such as a bar or kitchen, or metaphorically, as when a *battlefield* scene evokes bitter enmity, or a *police state* (barbed wire and searchlights at night) is used to denote oppressiveness.

Applications

These assets are sufficient to give every *Scéalextric* story a comic-strip form. The general application, named *Excelsior*, can sample *Scéalextric*'s ouevre at random, or retrieve tales that involve a given character, such as *Oprah* or *Bill Gates*. Examples of *Excelsior*'s comics for two *Scéalextric* stories are accessible online.[1] Its comics can be rendered as HTML documents that are suited to direct publication on the web, or as animated GIFs that may be shared within a tweet.

A co-creative variant of *Excelsior* engages in a dialogue with the user to refine the comic that is jointly produced. A metaphor-oriented chatbot, named *Figaro*, is repurposed to manage this dialogue using its rich lexicon and a system of pattern-matching rules. The narrative impetus for the joint work is at first provided by the user, in the form of a simple statement, typically a metaphor, that establishes the action, such as "my life is a joke" or "Donald is in the dog house." As in most chatbots, this input is then mapped via a series of stimulus:response rules into a corresponding output text. However, *Excelsior* does not generate raw surface outputs but XML forms that integrate a story-level understanding of the input with a ComiXML rendering of that interpretation.

It does this by tapping into the *Scéalextric* causal graph, which links every story action to every possible next action via *so*, *but* and *then* arcs. The Figaro lexicon connects the words of the user's input to relevant vertices in this graph, so that e.g., "life" maps onto *interact_with* (agent), "joke" maps to *is_entertained_by* (patient) and *laugh_at* (patient), and "in the dog house" maps to *criticize* (patient), *chastise* (patient) and *banish* (patient). *Excelsior* arrives at its story-based interpretation of a user input by seeking out a path in the causal graph between the vertices provided in the input. This pathway, a sequence of connected *Scéalextric* actions, is first converted into ScéalXML and then into ComiXML. For instance, the input "My boss is a tyrant" is mapped to a path linking *command*(agt, pnt) to *despise*(pnt, agt), which is summarized by the chatbot as: "A BOSS COMMANDS, BUT TAUNTS, SO EVENTUALLY IS DESPISED BY A HARSE CRITIC." To produce a three-panel strip from its narrative, the system chooses *Scéalextric* characters to portray its key roles, such as *Boudicca* & *Spartacus* or *Sideshow Bob* & *Lisa Simpson*. Notice how *Excelsior* infers that the user despises his boss, because the boss goes from giving orders to issuing taunts.

The initiative now passes back to the user, who can either accept the system's inferences with an affirming "yes" or "OK", or reject them with a "no" or a "try again." The user can also request a near alternative, by replying "almost" or "not quite", or require that the main roles be swapped by responding "the other way around." The user can elaborate an acceptable response by replying "so?" or "then what?", or introduce a dramatic kink to the tale with a "but." Each such reply prompts the system to add another action to the developing story, and another panel to the growing comic. New additions are in turn subject to acceptance or rejection by the user, who guides the system through the story space from the initial input to the final narrative and comic strip.

This co-creative, dialogue-driven variant of *Excelsior* has yet to be fully evaluated. However, its lexicon plays a key role in the web-scale application that we will evaluate next.

---

[1] https://tinyurl.com/ykdn33r3 and https://tinyurl.com/4b5ekc2y

Interventions: Comics With A Purpose

We can distinguish between comics that have a meaningful narrative and those whose narrative serves a larger meaning. The former tell a story that may divert and entertain us, but the latter carry a message that their creators need us to hear. Think of the difference between the newspaper comics that avoid politics and those, like Gary Trudeau's *Doonesbury*, that weave a political stance into the subtext of their stories. Comics can be used as a candy-coloured wrapper for views that some readers deem unpalatable, thereby increasing the diversity of the audience for those views. Or, as argued by Johnson *et al*. (2020) in the context of the rancorous online debate about vaccination, by increasing the heterogeneity of arguments in a community we can stop it from growing into a self-reinforcing echo-chamber. So, as a first step to using comics as our medium of intervention into early-stage echo chambers, let's explore this debate as it unfolds on Twitter.

Characterising the data and the debate

Twitter's streaming API was used to collate a dataset of all tweets that use a relevant hashtag from a list of approx. 60 tags, from *#GetVaccinated* to *#NoVaccinePassports*. During Q4 of 2021, as national vaccination drives ran at full tilt, a corpus of 1.6M tweets from nearly 400k users was gathered.

To characterize the debate, we take a rather simple view of a user's "stance" on vaccination, and assume that one is either *pro*- or *anti*-vaccines. The debate is more subtle than this dichotomy allows, and encompasses those who chafe at vaccine mandates, vaccine passports, masks, lockdowns, or at any curtailment of their pre-Covid lives. Nonetheless, there is still a sufficient consistency of attitudes to vaccines to make this pro/anti split a useful one. To assign a stance to each user, we build a graph of all retweets in the dataset, to indicate who retweets whom, and how often they do so. In this graph we identify the 100 most retweeted users, the so-called *influencers* or *evangelists*, and manually assign a pro (+1.0) or anti (-1.0) numeric stance to each. For every other user in the graph, we now estimate their stance as the weighted average of the stances of those that they retweet, weighted by the number of times they retweet them. After 50 iterations, the initial evangelist stances percolate down to every reachable user in the graph, to mark their position in the debate as a number between -1.0 and +1.0. In total, 149,162 users (38%) are assigned a positive *pro* stance and 214,066 users (55%) are assigned a negative *anti* stance. The remaining 7% are not in this retweet graph, or are not connected to one of our initial evangelists. Of those that are assigned a stance, the pro/anti split in the debate is 41/59%.

The dataset contains 39,366 distinct hashtags, many of which show a clear pro- or anti- bias, from *#VaccinesWork* to *#DoNotComply*. To evaluate our automatic assignment of stances, we look at the most frequently-used tags in the data and identify 100 clearly *pro* and 100 clearly *anti* tags. Looking at those who use these tags, we find clear support for our approach to stance allocation. The probability that a user who uses more pro-tags than anti-tags is assigned a pro-stance is .994, while the probability that one who uses more anti- than pro-tags is assigned an anti-stance is .999.

Hashtags condense arguments into compact, meme-like forms, such as *#VaccinesSaveLives* and *#FauciLied*, so to take the pulse of a debate we must get a handle on its tags. We first assign a stance to every hashtag, by estimating the stance of each as the weighted mean of the stances of those who use it, weighted by the number of times they use it. As a result, 39% of tags are assigned a positive *pro* stance, and 61% of tags are assigned a negative *anti* stance. These tags are viewed as discrete wholes, yet most tags are multiword forms with a headline-like structure. To unzip each hashtag into its headline content, we apply the *camel case* heuristic (as interior words tend to be capitalized) and a large lexicon (specifically, the *Figaro* lexicon of the previous section) to produce a sequence of words from each composite tag. 66% of the tag set, or 25,999 tags, can be segmented in this way.

Because Figaro's lexicon supports metaphorical analysis, it categorizes its entries along image-schematic dimensions such as *accepting* (up, love, praise, pick, etc.) and *rejecting* (down, fire, kill, dump, etc.). It also marks *negation* words (no, never, don't, etc.), and those related to *responsibilities* (e.g., rule, law, tax) and *rights* (e.g., freedom, truth, choice). We extend this lexicon to also mark geographic locations, such as Australia, the USA, China, Europe and their cities. We also mark references to the left or right of the political spectrum (e.g. Dems, Biden on *left*, GOP, Trump on *right*). This allows us to characterize the argument carried in a tag along these dimensions (e.g., responsibilities in Europe, or an acceptance of rights in the USA). If we now view users as aggregates of the arguments they use, i.e. as aggregates as the hashtags they use, we can apply this characterization to users too. For instance, we can say whether a given user is more accepting than rejecting (i.e., uses more accepting arguments than rejecting ones), or more focused on rights over responsibilities (i.e. uses more tags focused on rights), or more likely to reference the political left than the right.

Unsurprisingly, the hashtags of the anti-vaccination side show a clear preference for overt negation. The probability that a negated tag is assigned an anti-stance is .762, while a negated tag that highlights responsibilities has a probability of .981 of being labeled *anti*. Similarly, any tag that shows a rejecting attitude to responsibilities has a .942 probability of being labeled *anti* (these results are significant at the $p < .0001$ level). A clear political faultline is also evident at the hashtag level. Hashtags that mix rejection and a reference to the political left have a .957 probability of being labeled *anti*, while those that mix rejection and a reference to the political right have a .920 probability of being labeled *pro*.

These regularities also hold at the user level. A user that is more accepting than rejecting, and uses more accepting arguments than rejecting ones, has a .898 probability of being labeled *pro*, while one that is more rejecting than accepting has a .915 probability of being labeled *anti*. This simple criterion accounts for 75% of all stanced users. The battlelines are very clearly drawn in this debate, since any user that makes even a single rejecting reference to the idea of responsibility (177,319 do, or 45% of all users) has a probability of .966 of being labeled *anti*. Once again, each of these findings is significant at the $p < .0001$ level.

Evaluation: Data-Driven Comics

Excelsior has been adapted to work with the outputs of the *Scéalextric* story-generator, but can it be used to represent the cut and thrust of arguments in the vaccination domain? To test its applicability to online debates, a sample of 1,500 tweets is selected at random from our vaccine dataset. Each tweet has one or more hashtags with a quantifiable stance, and each contains one or more well-formed sentences that can be used as its narrative content. The sample has an even split of *pro* and *anti* sentiments (750 tweets of each stance). To test the expressive range of the Excelsior representation, we consider whether humans versed in this representation can effectively map these 1,500 tweets into a comics form. If so, we can have faith that the representation is expressive enough for generating comics about wide-ranging concerns. Moreover, the resulting mapping provides a parallel corpus for training an automatic translator of tweets into comics – this is a subject for future work and another paper – and as a baseline for evaluating comics generated from hashtags.

Annotators are first familiarized with Excelsior's assets and its markup conventions. A lightweight markup is used in lieu of full ComiXML, so that annotators need not create detailed XML forms. Rather, the markup simply segments each text into separate panels, and identifies the figure that speaks (or thinks) each chunk of text. For each figure in the panel, a pose and an orientation is also defined, and for the panel itself, a backdrop asset may also be specified. This is sufficient for a machine to construct the full XML for itself.

All 1,500 tweets were translated into a comics form, and none were discarded or labeled as too difficult to translate. The annotators used 95% of the available assets to markup the sample, which suggests they have a broad applicability. For the sample as a whole, the 5 most frequently used pose assets are: *operating* (a figure in a surgical mask carries a syringe and knife); *experimenting* (the same figure, without the mask, cackles as a rat scurries underfoot); *defensive* (an anxious figure retreats with its arms outstretched); *running away* (a scared figure flees in terror); and *rude* (an angry figure "flips the finger"). The 5 backdrop assets most often used are: *hospital* (interior); *hospital* (exterior); *graveyard* (tombstones and grass); *government* (a view of congress); and *battlefield* (barbed wire, ruins and scorched earth).

Each tweet text to be annotated contains, on average, 14 words. When annotators segment these texts into a series of comics panels, the mean number of panels per tweet is 2.82 (sd.=.92). Most comics are thus two to four panels in length. The mean number of words per text segment – and thus, per panel – is 4.68 (sd.=2.85). Most text balloons will thus contain between two and eight words apiece. Specific assets are favoured for the depiction of the arguments from opposing stances. Vaccination is generally depicted using the *operating* pose for the "pro" tweets, and depicted using the more sinister *experimenting* pose for the "anti" tweets. The *graveyard* and *hospital* backdrops find equal favour in pro and anti tweets – the graveyard is a terminus for those who refuse vaccination, or who die from its side effects – but *government* and *battlefield* are preferred by those who campaign against (rather than for) vaccination mandates.

Hashtag Comics: Automated Generation

We estimate that 120 person-hours of effort were needed to translate this sample of 1,500 tweets into ComiXML. This investment of time and effort might be repaid by a machine learning approach to generation that uses this tagged dataset for supervised training, but it is still a significant outlay for each new debate topic to be modeled (e.g., climate change). Any automated approach to *argument-by-comics* will likely lack the finesse of a human-led one, but it should be easier to adapt to new debate topics. Whole tweets are too loosely structured to serve as the basis of this automated approach, but hashtags – which work best when used as hooks for the content they adorn – are ideal: they capture the gist of an argument in a pithy form that one hopes will "go viral."

To begin, we pass every hashtag in the dataset through a segmenter, to identify its individual words. As noted earlier, 25,999 tags (or 66% of the total) can be segmented using a lexicon and the norms of "camel casing." The latter allows a tag to be segmented even if some of its words are not in the lexicon, such as "#*CovidIsAHoax*. Figaro's lexicon will cover most of the domain-independent terms that are used, such as "is" and "hoax", but important terms like "Covid" must also be added, as must the names of central figures in the debate, such as "Fauci," "Pfizer" and "Biden." To plug the largest holes in the lexicon, we rank the unlisted terms by frequency in the tag set, and add entries for the top 200. This leads us to add entries for "Covid" and "Covid19", as well as for casual references to vaccines ("vax", "vaxxed") and for recurring characters in the debate (Tony Fauci, Joe Biden, Scott Morrison, Boris Johnson, etc.) and the various ways in which they are named in a tag. For these characters we define Excelsior specifications for their cartoon effigies (e.g., white hair for Fauci, blond for Boris, bald for Biden), and for "Covid" we specify the poses *stalking* and *preying*. Every comic hero needs a comic villain, and named figures are convenient bogeymen (or bogeywomen) for the debate. Any user that mentions *Fauci* in even a single tweet is very likely to hold anti-vax views (probability = .914), while for *Bill Gates* that probability rises to .947; for *Jacinda Ardern*, the prime minister of New Zealand, it rises to .996.

A hashtag can be automatically translated into ComiXML if: every word in the segmented hashtag has a lexical entry; the hashtag provides enough material for at least one panel; its lexical entries specify poses for two figures in each one; and, for certain terms, suggests a backdrop for the panel too. The hashtag may contain lexical items that do not translate into any visual element, such as function words ("a", "the", etc.), as these can contribute to the content of text balloons, but it may not contain a term that falls outside the lexicon. These restrictions lead to 9,802 hashtags (or one third of all tags that can be segmented) being mapped into ComiXML. In turn, 36% of these comics are produced for hashtags that are associated with a pro-stance, and 64% are generated for tags that suggest an anti-stance. For this dataset, there is a political dimension to how tags convey stances – *anti* tags lean right and *pro* tags lean left – so the translator uses this finding to colour the figures in its panels. For a comic that is created for a *pro* tag, the protagonist – the figure who utters the words of the tag – is blue, and the antagonist is red. For an *anti*-leaning tag, the comic's protagonist is red and the antagonist is blue. In the comic of Fig. 3, for the *anti* tag *#DemocratsAreDestroyingAmerica,* "Democrats" is defined by the lexicon as *one who votes for Biden*, so the red protagonist is instead paired with a cartoon Joe Biden:

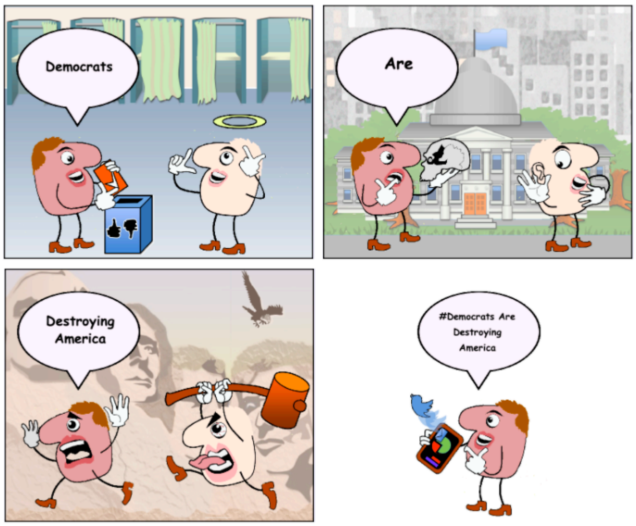

Fig. 3. Comic for the *anti* tag *#DemocratsAreDestroyingAmerica*.

As a comic unfurls from left-to-right, it reveals successive words in the hashtag. Each panel hinges on a single hashtag term that suggests a pairing of figure poses. "Democrats," for instance, suggests *voting for* and *saintly*, where the latter is the pose struck by the cartoon Biden. The setting for this panel is suggested by the *voting for* pose, which is strongly linked to backdrop *polling station* and only weakly tied to the backdrop *government*. This weak association suggests a scene change in the second panel, which stages the copula term "are" in a dramatic, Hamlet-like fashion. In the third panel, the action term "destroying" suggests a pairing of the poses *destructive* and *running away*, while "America" is rendered in the same panel as the backdrop *Mt. Rushmore*. The narrative impetus of the comic, the original hashtag, is ultimately summarized with a final borderless, blank panel.

When a hashtag takes a dramatic turn, its comic does too. By forcing the narrative into a two-fisted tale of protagonist vs. antagonist, the red protagonist becomes both Democrat and anti-Democrat, a figure that starts the tale by voting for Biden and ends it by fleeing from him in terror. We see just as dramatic an arc in the comic of Fig. 4, which is derived from the pro-leaning tag *#CovidIsNotGone*. "Covid" ranks high in the list of lexicon gaps for the vaccine dataset, and we plug the gap with an entry that specifies a colour (*green*) and a pose (*preying*). This makes "Covid" a state of being, a posture that can change from one panel to the next. In the first panel, the antagonist is portrayed as "Covid", preying on a scared protagonist. In the next, which depicts negation with a wag of the finger, the antagonist resumes her "pro" colouring. The emergent effect suggests that Covid can lurk within a character, waiting for a chance to express itself.

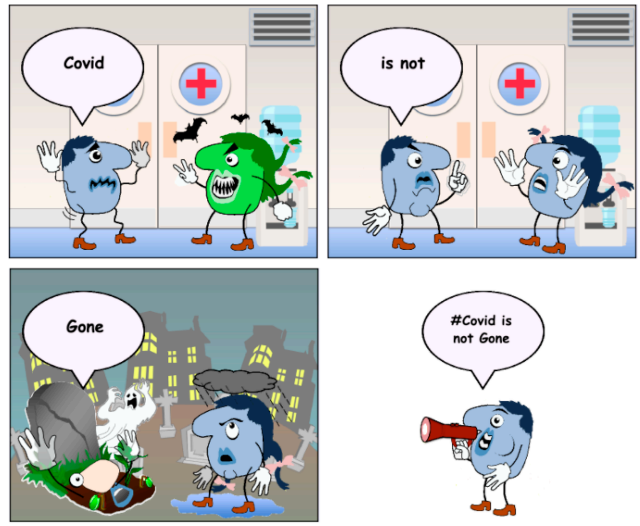

Fig. 4. A comic generated for the *pro* hashtag *#CovidIsNotGone*.

Discussion

How do the comics that are generated automatically from hashtags compare with those created manually from entire tweets? On average, a hashtag-based comic contains 2.83 panels (sd. = .687), including the final summary panel, and its average speech balloon has 1.46 words (sd. = .55). The typical hashtag comic is thus punchier than a typical tweet-based comic, yet the tag comics hit the same high notes. As was the case in the human-translated sample, this larger set uses 95% of Excelsior's assets, and likewise numbers the backdrops *hospital interior* and *government*, and the poses *operating* and *defensive*, among its top 10 most-used assets. Other highly ranked assets include the poses *preying*, *sick* and *scared* (to depict coronavirus and its victims), and the backdrops *graveyard* (for pro- and anti- deaths) and *police state* (a dark variant of *government*, to suggest oppression).

Most of Excelsior's poses depict highly emotional states, bound up in the character actions that evoke those states. A significant proportion of Excelsior's backdrops also evoke – or stand in for – emotional states, such as the *graveyard*, *police state*, *dungeon*, and *battlefield* backgrounds, which can be used literally – in a *Scéalextric* tale, for instance – or metaphorically, in debate spaces such as that for vaccines. Our sample of 1,500 pro- and anti-vaccine tweets serves as a representative slice of the Twitter debate, and since these are mapped onto Excelsior's emotional images by hand, we can be confident that these choices capture the tenor of the online debate. But how well do our hashtag comics capture the dominant themes and emotions of the debate? To assess how closely our automated comics hit the same notes, we measure the correlation between the usage frequencies of each asset for both manual and automatic comic generation.

Comparing the usage frequencies of all Excelsior assets across both sets of comics, we find that *Pearson's* $r = .824$. This reflects a strong correlation between the choices that humans make when visualizing online arguments, and those that the machine makes when visualizing them for itself.

Concluding Remarks

Comics are an appealing delivery mechanism and framing device for our stories, wherever they happen to come from. Data-driven comics are a specific form of "creativity with a purpose" (as opposed to "art for art's sake") that uses the expressive and representational affordances of the medium to convey a specific message and advance a particular goal. So, on the back of some large-scale data analysis and some small-scale plugging of lexical gaps, a comics generator can be adapted to tell the stories of a diverse body of users, and in doing so affect the ways in which they interact. That, at least, is our goal here: to tell the stories, and express the concerns, of different users to a wider audience than they might otherwise reach, and thus increase the heterogeneity of viewpoints within their opinion "silos." We have shown how individual arguments, as expressed in individual tags, can be translated into comics without loss of emotionality. A machine can understand those arguments, like those who make them, at a general level e.g., as rejecting or accepting of particular people, policies or ideas. But, as a next step, we must do more than echo the arguments in comics form, and target them at communities that are least open to them. These practical interventions will pose the best test of the formats, resources and tools that we have presented here.

Comics can be added as a medium of expression to many kinds of computational creativity system. Story-generators are just the most obvious, and we anticipate that *ScéalXML* can be adapted – with minor extensions – to suit the needs of systems other than our own choice, *Scéalextric*. In cases where this necessitates the creation of new assets, such as pre-Columbian imagery for the *Mexica* stories of Pérez y Pérez (2007), these additions will benefit both systems. We have shown here how comics can serve as a means of *data visualization* i.e., as a means of emotively visualizing the drama inherent in aspects of a large data set. A comic can be the main creative output of a system, or a useful means of framing the system's explanations of its own creativity. If used for the latter, we expect to find a great many outlets for comics-based creativity, in a wide range of applications that go beyond what we typically think of as story-centred.

Author Contributions

The reported research is wholly the work of the sole author.

Acknowledgements

The author acknowledges the intangible yet invaluable outcomes of his many discussions with colleagues and friends.

References

Agrawal, H., Chandrasekaran, A., Batra, D., Parikh, D., & M. Bansal. (2016). Sort story: Sorting jumbled images and captions into stories. *Proc. of Empirical Methods in NLP*, 925–931.

Alves, T., McMichael, A., Simões, A., Vala, M., Paiva, A. & Aylett, R. (2007). Comics2D: Describing and creating comics from story-based applications with autonomous characters. *In Proceedings of CASA, the 20th Annual Conference on Computer Animation and Social Agents*. Hasselt, Belgium.

Cavazza, M., Charles, F., & Mead, S. J. (2003). Interactive Storytelling: From AI Experiment to New Media. *Proceedings of the 2nd international conference on Entertainment computing*, 1-8.

Cohn, N.. (2013). *The Visual Language of Comics: Introduction to the Structure and Cognition of Sequential Images.* London: Bloomsbury.

Cohn, N. & Magliano, J.P. (2020). Visual Narrative Research: An Emerging Field in Cognitive Science. *Topics in Cognitive Science,* 12:197–223.

Eisner, W. (1985). Comics & Sequential Art. Tamarac, Florida: Poorhouse Press.

Forceville, C., Veale, T., & Feyaerts, K. (2010). Balloonics: The visuals of balloons in comics. *The rise and reason of comics and graphic literature: Critical essays on the form*. J. Goggin & D. Hassler-Forest (Eds.). Jefferson NC: McFarland, pp 56–73.

Gaiman, N. (2021). How To Create A Comic Book. *Masterclass course.* https://www.masterclass.com/

Gervás, P. (2014). Composing narrative discourse for stories of many characters: A case study over a chess game. *Literary and Linguistic Computing*, 29(4):511–531.

Iyyer, M., Manjunatha, V., Guha, A., Vyas, Y., Boyd-Graber, J., Daumé, H., III & Davis, L. (2017). The amazing mysteries of the gutter: Drawing inferences between panels in comic book narratives. *Proc. of Computer Vision & Pattern Recognition*, 6478–87.

Johnson, N.F., Velásquez, N., Restrepo, N.J., Leahy, R., Gabriel, R., El Oud, S., Zheng, M., Manrique, P., Wuchty, S. & Lupu, Y. (2020). The online competition between pro- and anti-vaccination views. *Nature*, 582:230–233.

Kurlander, D., Skelly, T. & Salesin, D. (1996). *Comic Chat*. In proceedings of SIGGRAPH'96, the 23rd annual conference on Computer graphics and interactive techniques, pp 225–236.

McCloud, S. (1993). *Understanding Comics: The Invisible Art.* New York: Harper Collins.

McIntosh, J. (2005). ComicsML: A proposed simple markup language for online comics. *https://jmac.org/projects/comics_ml/*

Melistas, T., Siglidis, Y., Kalogiannis, F. & Manouach, I. (2021). A Deep Learning Pipeline for the Synthesis of Graphic Novels. In *Proc. of ICCC'21, the 12th International Conf. on Computational Creativity*, Mexico City.

Montfort, N., Pérez y Pérez, R., Harrell, F. & Campana, A. (2013). Slant: A blackboard system to generate plot, figuration, and narrative discourse aspects of stories. In *Proc. of the 4th International Conf. on Computational Creativity*. Sidney, Australia, June 12-14.

Pérez y Pérez, R. (2007). Employing emotions to drive plot generation in a computer-based storyteller. *Cognitive Systems Research* 8(2):89-109.

Veale, T. (2017). Déjà Vu All Over Again: On the Creative Value of Familiar Elements in the Telling of Original Tales. In Proc. of *ICCC 2017, the 8th Int. Conf. on Comput. Creativity, Atlanta*.

Walsh, J. (2012). Comic book markup language: An introduction and rationale. *Digital humanities quarterly* 6(1).