Reading without words: Eye movements in the comprehension of comic strips

Tom Foulsham[1]*, Dean Wybrow[1] and Neil Cohn[2]

[1] Department of Psychology, University of Essex, UK

[2] Department of Cognitive Science, University of California at San Diego, USA

*Corresponding author:

Department of Psychology, Wivenhoe Park, Colchester, Essex, UK, CO4 3SQ.

Tel. (+44) 1206 874159. Email: foulsham@essex.ac.uk

Abstract

The study of attention in pictures is mostly limited to individual images. When we "read" a visual narrative (e.g., a comic strip), the pictures have a coherent sequence, but it is not known how this affects attention. In two experiments, we eyetracked participants in order to investigate how disrupting the visual sequence of a comic strip would affect attention. Both when panels were presented one at a time (Experiment 1) and when a sequence was presented all together (Experiment 2), pictures were understood more quickly and with fewer fixations when in their original order. When order was randomised, the same pictures required more attention and additional "regressions". Fixation distributions also differed when the narrative was intact, showing that context affects where we look. This reveals the role of top-down structures when we attend to pictorial information, as well as providing a springboard for applied research into attention within image sequences.

Introduction

In many contexts, information is presented to people in a graphical or diagrammatic form. Although some researchers have applied experimental methods to how we understand advertisements (e.g., Rayner et al., 2001), diagrams (e.g., Hegarty & Just, 1993) and instructions with pictures and text (e.g., Holsanova et al., 2009), these media are most often limited to single displays. In contrast, the sequential images found in visual narratives like comics use a deliberate and guided sequence that provides order to the component parts of a larger visual display. Such deliberate ordering is reminiscent of "reading" words in language, and indeed, when people engage with comics, they often describe it as "reading," whether or not it contains text along with the images. Setting aside the inclusion of writing, in order to understand wordless visual narratives, "readers" must select sequential units in a particular order (i.e., choosing which panel to look at), comprehend the meaning of individual components within those units (i.e., characters, objects etc.), and integrate these elements into a coherent discourse. Such a process of comprehending sequential images has often been compared to the comprehension of words in sentences (Cohn, 2013b, 2013c; McCloud, 1993).

Much has been discovered about attention during text reading and the perception of individual images, but few studies have examined how people view the static sequential images found in visual narratives like comics. In particular, there are no studies which have used careful experimental manipulation of the structure of these visual narratives in order to establish how participants view meaningful sequences of images. Our experiments examine how participants view visual narratives in the absence

of text. The aim is to apply what we know about attention in text and images to simple visual narratives, but also to utilise these narratives as a way of experimentally manipulating context in the processing of pictorial information. We will begin by considering previous work on attention in scenes and text reading, as well as the comprehension of visual narratives.

Eye movements and attention in text and pictures

When we view a complex stimulus, we move our eyes frequently in order to align our high-acuity central vision with particular items of importance. This is true both when we are looking at a picture, and when we are reading text. In text, readers of English make a series of fixations, in a mostly left-to-right sequence along a line. After several decades of research, much has been learned about the factors which affect where we fixate in reading, as well as how long we fixate there (see Rayner, 2009, for a review). For example, the duration of individual fixations is influenced by visual factors (such as font), as well as by aspects of the fixated word (e.g., its frequency and its predictability given the sentence context). These effects demonstrate that eye movements reflect on-going cognitive processes involved in comprehending text. When comprehension is more difficult, fixation durations are prolonged and participants are more likely to make a saccade to a previous word (a regression).

While viewing scenes, we also make a series of saccades and fixations. However, perhaps because images are less easily broken down into constituent parts, it has proven more difficult to elucidate the factors involved in where and when we fixate (see

Foulsham, 2015, for a recent review). In general, people fixate locations which are informative for what they are doing (e.g., Yarbus, 1967). For example, when searching for something in an image, participants look at items which are similar in appearance to the target and at locations where, given the context, that target is likely to appear (Torralba et al., 2006). As with reading, the duration of fixations in picture viewing is believed to largely reflect the processing of attended information at the fovea. Therefore, degraded or semantically unpredictable regions are fixated for longer (e.g., Underwood & Foulsham, 2006).

In many contexts (e.g., a newspaper cartoon or a textbook illustration), a picture is combined with text and the observer must interpret both to fully understand the stimulus. Carroll et al., (1992) asked participants to rate the humour of various single-image cartoons. Eyetracking results showed a brief inspection of the image, followed by the reading of the caption and finally a more detailed re-inspection of the picture. In Hegarty and Just (1993), meanwhile, participants alternated their gaze between a mechanical diagram and the caption describing it. In these studies, the information in the text and the image complemented each other, and observers tended to use both, mostly using the text and then referring to the image to check details. Rayner et al. (2001) also reported fixations on text and images in print advertisements, finding a higher frequency of the former. Applying similar principles, Holsanova et al., (2009) found that text and images in an information graphic were read more efficiently and more deeply when they were arranged in an integrated, serial fashion (with captions aligned to their relevant pictorial content and a clear global layout).

The present study investigates comic-strip style visual narratives in the absence of text. The research described thus far demonstrates that picture viewing is strongly affected by the task and the cognitive processes of the observer (and not just the content of the image, i.e., by "top-down" processes). Despite this, a popular approach for predicting fixations on pictures, the saliency map approach, relies principally on context-free image features (Itti & Koch, 2000; Parkhurst et al., 2002). How does context affect attention? Visual narratives, which consist of individual pictures bound into a semantically-meaningful sequence, provide an ideal way to answer this question. Unlike studies which use "free-viewing" or memorization tasks to examine picture understanding (e.g., Parkhurst et al., 2002; Underwood & Foulsham, 2006), visual narrative comprehension requires integration of information into a mental model (as do text-picture displays). Thus the study of visual narrative comprehension can provide theoretical insights (by assessing what sorts of context affect attention) as well as being readily applied to sequentially organized graphics such as instruction manuals.

Although a few studies have previously applied eyetracking to visual narratives of comics, these studies have been ad hoc, and none have manipulated the order of the narrative to test predictions about visual attention (Omori, Ishii & Kurata, 2004; Nakazawa, 2002, 2005). In all these studies, the stimuli contain both text and images, but there is little description or control of their presentation, and no detailed analysis of eye movements which could be compared to work on reading and scenes. Interestingly, Kunze et al., (2013) recently showed that eye movements from various reading materials, including manga, textbooks and reading, could be discriminated using machine learning.

This suggests that eye movements in visual narratives are distinct, which may have important implications for the design of comics and information graphics.

The comprehension of visual narratives

While several early theories compared the structure of comics and language in more metaphoric terms (for a review, see Cohn, 2012), recent research has explored the analogy between visual narratives and language with increasing seriousness using tools from contemporary linguistics and experimental psychology (Cohn, 2013b). Theoretical work has argued that coherent sequences are processed using a "narrative grammar" to order the meaning of sequential images. This grammar uses analogous functional principles as syntax uses to order the word units in sentences, such that sequential images take on categorical roles which are organised into constituent structures, only it does so at a discourse level of meaning (Cohn, Jackendoff, Holcomb, & Kuperberg, 2014). That is, though individual panel-units typically contain more information than individual word-units—and thus are closer in meaning to whole sentences—similar "grammatical" principles operate to organize images and words into well-formed sequences (Cohn, 2013c; Cohn, et al., 2012). This parallelism has been born out by experimentation measuring event-related brain potentials (ERPs), which has suggested that the same neural effects appear from the manipulation of this narrative grammar in sequential images as from manipulations of syntax in sentences (Cohn, et al., 2014; Cohn, et al., 2012). The processing of sequential images may therefore be similar to the processing of sequential words, even at a neurocognitive level, despite the differences in levels of

semantics. The existence of a narrative grammar suggests that the rich context in a sequence of images will have systematic effects on fixations within each image (akin to those seen on words within a sentence).

Aside from the content of each component image, the reading order of comic page layouts is likely to be similar to text. When forced to choose, readers typically prefer to order page layouts using a left-to-right and top-to-bottom strategy that maintains the "Z-path" of reading inherited from writing systems, even when those comic pages have no content in the panels (Cohn, 2013a; Cohn & Campbell, 2015). These findings go against prominent theories about comics that readers are fairly unconstrained in their engagement with page layouts, and "the eye's movements on the surface of the page are relatively erratic and do not respect any precise protocol"—ideas which were purportedly influenced by unpublished French eyetracking experiments in the 1980s (Groensteen, 2007, p. 47).

The present study

In the present study, participants viewed sequential images in either an "original" order, which had a consistent and humourous narrative, or in a randomized order, with no coherent structure. Given the paucity of eyetracking data in comics, we used this reliable manipulation (Cohn & Wittenberg, 2015; Cohn, et al., 2012) in order to confirm previous findings, while extending the methodology to focus on aspects of attention. In accordance with previous studies of visual narrative (Cohn & Wittenberg, 2015; Cohn, et al., 2012; Gernsbacher, Varner, & Faust, 1990), we expected that participants would find it more

difficult to understand randomly ordered panels, and we checked their understanding with a self-report measure. How would this difficulty manifest in the overt visual attention deployed to panels and visual details? There were a number of predictions. First, if panels in a sequence are interpreted according to the narrative of the comic strip, as are words in a structured sentence or sentences in a discourse, then we expected to replicate our previous findings that randomly ordered panels should take longer to view and understand (Cohn & Wittenberg, 2015; Cohn, et al., 2012).

Second, increased processing difficulty for panels in a random order might have an impact on the way in which strips are attended, in the same way that lexical and syntactic difficulty affect eye movements in reading. For example, randomly ordered panels which are more difficult to process might be looked at with more fixations and/or fixations of a longer duration. Confirming this prediction is important, because it addresses whether attention to an image is driven mostly by the content of the image—in which case panels should be looked at consistently regardless of their order—or whether even individual fixations reflect the meaning of the panel within the context of the strip. A significant body of recent research has tried to predict fixation locations on the basis of simple image features (Itti & Koch, 2000; Foulsham, 2015). Describing if and how context affects eye movements will lay the groundwork for improving such models by representing top-down information.

Third, we predicted that the narrative sequence, the presentation order and the selection of particular panels would all interact. For example, previous work has consistently shown that, like when reading text, self-paced viewing times to panels starting a sequence are longer than others in the sequence, facilitating the understanding

of a new situation (Cohn, 2014b; Cohn & Wittenberg, 2015; Cohn & Paczynski, 2013),

even in randomly ordered sequences (Cohn & Wittenberg, 2015). Moreover, in sentence

reading, processing difficulty often results in regressions which shift attention back to a

previously examined word. If the same is true in the processing of visual narratives, we

would expect participants to sometimes depart from the sequential order in which

viewing is designed, especially if the narrative is hard to follow (as in the random panels).

This prediction is tested in Experiment 2, which presented all six panels of a sequence at

one time, allowing participants the freedom to move between the panels and look back at

earlier elements. Very little empirical data currently exists regarding how participants

attend to panels in this situation. As described above, some theorists have claimed that

inspection patterns in comics are erratic (e.g., Groensteen, 2007). When presented with

multiple panels, participants might not inspect the images in a strict order. They might

skip ahead, flit regularly between earlier and later images, or build up an overall idea of

the narrative before inspecting more closely. Our second experiment will provide

measures of such patterns, determine how they are affected by context, and lay the

foundation for subsequent research in this medium.

Experiment 1

In our first experiment, we asked participants to view visual narratives one image at a

time at their own pace and then rate the coherency of the sequences. This "self-paced

viewing" paradigm has been used as a reliable measurement of comprehension in other

studies of visual narrative (Cohn, 2014b; Cohn & Wittenberg, 2015; Cohn & Paczynski,

2013), inlcuding the contrast between coherent and scrambled sequences (Cohn &

Wittenberg, 2015). These precedents therefore provide a useful place to start applying eyetracking to narrative comprehension.

Like word reading, sequential images order panels in a sequence, constrained by a grammar—in this case a narrative structure (Cohn, 2013b). However, unlike the words in sentences, the image-units in in visual narratives are two-dimensional and complex, normally conveying more information than a single word. In terms of information-level, images are closer to sentences than individual words, though sequences of both may be governed by similar combinatorial systems. Thus, while even long words often receive just one or two fixations, we expect each scene depicted in a panel to receive more attention. We therefore tested our first two predictions by examining the frequency and duration of fixations within panels, while maintaining the control afforded by the self-paced, panel-by-panel method.

Method

Participants

Sixteen student volunteers (12 female) from the University of Essex took part in return for payment. The mean age of participants was 19 years. All participants had normal or corrected-to-normal visual acuity. In a short questionnaire after the experiment, all but two of the participants said that they recognised characters from the *Peanuts* comic strips used, but that they had not read strips from this series recently. Participants' expertise with reading comics was assessed using the Visual Language Fluency Index (VLFI), which asked participants to rate the frequency both currently and while growing up that they read a variety of visual narratives (comic books, comic strips, graphic novels,

Japanese manga), the frequency that they draw comics, and their self-rated "expertise" in reading and drawing comics (see Cohn, et al., 2012 for more details). These ratings were then combined using a formula that computed a "VLFI score" that has been shown to correlate with both behavioural and neurophysiological responses (Cohn, et al., 2012). Using this metric, an idealized average would be near 12, anything below 8 would be low, and anything near or above 22 would be high. Participants in this study had a mean VLFI score of 9.9, a low average (range, 4.8 – 21.0; SD = 4.8).

Stimuli and Design

The stimuli consisted of panels culled from six volumes of the *Complete Peanuts* by Charles Schulz (1952-1968), which were edited and recombined into novel sequences, and stimuli from this broader corpus have been used in several studies (Cohn, et al., 2014; Cohn & Paczynski, 2013; Cohn, et al., 2012). Novel sequences were generated to create sequences of six panels in length (as opposed to the standard four-panel daily strip length), and to reduce participants' familiarity with any specific sequences. These panels were grayscale images, normalized to a consistent size, featured a few main characters (such as Charlie Brown and Snoopy, recognisable to many audiences) and did not feature any text. For the current experiments, we used 96 visual sequences of 6-panels in length each of which showed a visual narrative evolving in a humourous or entertaining way (see Figure 1 for an example).

Sequential order was manipulated as a within-subjects factor. Each participant saw half of the strips with panels in the original sequential order, which was narratively coherent. The other half were viewed in a randomized order, shuffled for each participant.

Two sets of stimuli were created such that, across participants, each particular set of 6

panels appeared in both the original and random conditions. Each panel was presented in

the centre of the screen and, at the viewing distance in the experiment, subtended

approximately 16 by 13 degrees of visual angle.

---Figure 1 about here---

Apparatus

Participants viewed panels on a 19-inch monitor from a fixed distance of 60cm. Eye

movements during viewing were recorded using the EyeLink 1000 system (SR Research),

which calculated monocular eye position every millisecond based on the pupil and

corneal reflection. Using the default EyeLink parser, saccades were detected based on

velocity and acceleration thresholds of $30°/s$ and $8000°/s^2$. Head position was restricted

with a chin-rest. Validation indicated that the average error in the position measurements

was 0.4°. The experiment was controlled by SR Research's Experiment Builder software,

and participants used a keyboard to respond during each trial.

Procedure

The experiment began with a short practice example that familiarized participants with

the mode of presentation, followed by a 9-point calibration with the eyetracker.

During each experimental trial (see Figure 1), participants viewed a 6-panel comic strip, one panel at a time, in a self-paced fashion. Before each strip, a central drift-check marker ensured that participants were fixating the centre of the screen, which they confirmed by pressing a key on the keyboard. The first panel then appeared, and participants were instructed to press the spacebar when they were ready to view the next panel. Following the presentation of the final panel in the sequence, a response screen appeared asking participants to give a coherency rating for "how difficult the comic strip had been to understand" on a 7-point, Likert-type scale (1=very easy to understand, 7=very hard to understand). The next trial began after this response was made.

Viewing continued in a self-paced fashion, with original and random conditions interleaved. Participants were instructed to "follow what happens" in the comic, and could take breaks where necessary. The whole experiment took approximately 30 minutes to complete.

Results

Data from one participant were excluded due to missing eye position data for a significant proportion of trials. We analysed viewing time, self-reported difficulty and eye movement measures by comparing participant means in the two conditions with paired-samples t-tests and repeated measures ANOVA. In addition, we focus on estimating standardized effect sizes for the key paired differences. Specifically, we report *d* as the mean difference normalised by the standard deviation of the difference scores (see Morris & Deshon, 2002), and we use bootstrapping to calculate 95% confidence intervals (CIs) around this effect size (Cumming & Finch, 2001).

Response measures

Panel viewing durations were calculated as the time from the onset of each panel to the participant's terminating keyboard response, and we assumed that this response would reflect viewing difficulty. In addition, participants gave explicit self-reports of their ease of understanding each strip.

Participants took about 10s to read each strip, equivalent to approximately 1.7s per panel. Mean viewing duration per panel was significantly shorter in the original condition (1696ms, SD = 435ms) than in the random condition (1774ms, SD = 423ms; $t(14) = 2.6$, $p = .02$, $d = 0.68$, 95% CIs = [0.36, 1.37]). This was consistent with the mean coherency rating for the sequences. Participants rated original orders (2.4, SD = 0.64) as significantly easier to understand than randomly ordered strips (3.6, SD = 1.04), $t(14) = 6.3$, $p < .001$, $d = 1.63$, 95% CIs = [1.22, 2.47]. Across strips, the correlation between

self-reported viewing difficulty and mean viewing time was positive and statistically

significant, $r = .28, p < .01$.

Eye movements

The longer time spent viewing each panel in the random condition could potentially be

associated with more fixations (perhaps indicating that more regions needed to be

inspected in order to understand the panel) or fixations with a longer duration (as would

be the case if particular elements needed more extensive processing). Table 1 shows that

the original and random conditions were different in both respects.

---Table 1 about here---

A paired t-test demonstrated that a significantly greater number of fixations were

made during the viewing of randomly ordered panels than during the viewing of

originally ordered panels, $t(14) = 2.2, p = .041, d = 0.58$, 95% CIs = [0.23, 1.07]. These

fixations were also longer, on average, although the mean difference of 7ms did not reach

the conventional level of statistical significance, $t(14) = 2.1, p = .056, d = 0.54$, 95% CIs

= [0.09, 1.12]. Changes in the processing of visual information are also sometimes

associated with larger or smaller saccades. For example, degraded displays which are

harder to see may be viewed with smaller saccades because information from a smaller

region can be processed on each fixation (e.g., Foulsham et al., 2011). There was no

effect of panel viewing order on the average size of saccades, $t(14) < 1$, d = 0.06, 95%

CIs = [-0.39, 0.62].

Presentation order

The results so far demonstrate that comic panels are viewed differently depending on

their order. It is important to note that, because the same panels were present in both

conditions, these differences cannot be due to the visual content of the individual panels

and must instead be because of their context within a visual narrative. To gain a further

understanding of this context we examined viewing duration as a function of original and

random presentation order. The mean self-paced viewing time is shown in Figure 2, as a

function of presentation order.

---Figure 2 about here---

A 2 x 6 repeated measures ANOVA with factors of condition and presentation

order confirmed that viewing times were longer for randomly presented panels than for

those in the original order, $F(1,14) = 6.95$, $p = .02$, $\eta_p^2 = .33$. There was also a significant

effect of presentation order, $F(5,70) = 40.02$, $p < .001$, $\eta_p^2 = .74$. Bonferroni-corrected

pairwise comparisons showed that the first panel was viewed significantly longer than all

other panels, but no other levels were different from each other. Although the effect of

presentation order was very similar across conditions, there was also an interaction,

$F(5,70) = 2.76, p = .03, \eta_p^2 = .16$. The simple main effect of condition was reliable for the first, second and fourth panels, but not for any other panel in the sequence.

The first panels in a sequence might be inspected for longer for a number of reasons, including differences in the content or complexity of these images. It is noteworthy, therefore, that even when the same panels were shown in a random presentation order, the first panel took longer to view. Moreover, when data from the random condition were re-organised according to their original sequence order (dashed line in Figure 2) the line was essentially flat. Therefore, it was presentation order, and not the way in which the panels were originally ordered, which had a systematic effect on self-paced viewing times.

Discussion

This experiment tested the effect of a coherently-ordered visual narrative on how people inspect the panels within comic strips. There were a number of clear effects of the presentation order, which confirmed the prediction that individual panels are "read", not as isolated pictures, but relative to the evolving narrative in a sequence.

As in prior studies, panels shown in the wrong order were rated as harder to understand, and they were looked at for longer, presumably because they required extra processing to understand what was happening (Cohn & Wittenberg, 2015; Cohn, et al., 2012). We might expect such processing to result in more fixations because more regions had to be fixated or re-fixated. This was indeed the case, and there was also a small and marginally significant tendency for randomly ordered panels to yield a longer average fixation duration. When viewing scenes, individual fixation durations are associated with

semantic properties of fixated objects. For example, objects that are semantically incongruous with their background are looked at for longer (Henderson et al., 1999; Underwood & Foulsham, 2006). Task also affects average fixation durations, with fixations when searching for something in a picture being generally shorter than when trying to remember the details or just freely viewing an image (e.g., Mills et al., 2011).

There are numerous examples of top-down effects on fixations in picture viewing (e.g., Henderson et al., 1999; Underwood & Foulsham, 2006; Yarbus, 1967). However, most research into image viewing uses single isolated pictures and some models assume that images are inspected independently. There are very few previous demonstrations that the context of an image within a sequence affects fixations. Our findings show that narrative structure, like other sorts of top-down information, reduce processing time and change the way that panels are inspected. For example, often the same or similar characters repeat across panels and so it may take only a brief fixation to confirm this. Other, novel regions may be prioritized for attention instead. In contrast, when panels appear out of sequence, more elements will be novel and need to be inspected for longer. We might therefore expect that viewing in the random condition might result in fixations on different regions. We will defer discussion of the location of fixations within a panel until we have described our second experiment.

There was also an effect of presentation order, with longer viewing times on the starting panel of a sequence. These findings replicate previous work showing that viewing times are longer to the first images of a visual narrative sequence than those in subsequent panels (Cohn, 2014b; Cohn & Wittenberg, 2015; Cohn & Paczynski, 2013), and reaction times are slower to target panels earlier in a narrative sequence (Cohn, et al.,

2012). This initial acquisition of information is described as a process of "laying a foundation" (Gernsbacher, 1990) for orienting to the content of the narrative sequence, be it verbal discourse (Glanzer, Fischer, & Dorfman, 1984; Haberlandt, 1984), or visual narratives (Cohn et al., 2012; Gernsbacher, 1985). Indeed, there does not seem to be anything special about the specific panels which are originally positioned first in a sequence, because these same panels were not fixated longer on average when in a random sequential position. Instead, longer processing times are likely caused by characters and background being more novel at that point—regardless of subsequent panel ordering—and thus require a process of "laying a foundation" for the comprehension of the rest of the sequence.

Of course, self-paced viewing paradigms of visual narratives are artificial when compared with the standard methods of presentation. In a more naturalistic context, comprehenders engage with panels arranged in a spatial layout on a page (Cohn, 2013a; Cohn & Campbell, 2015), where they are able to select panels in a different order to that prescribed by the artist. It has been suggested that participants may not have a consistent "protocol" for page reading (Groensteen, 2007), and we might also expect some people to look back at panels or skip some altogether. Experiment 2 investigated these possibilities.

Experiment 2

In Experiment 1, we found that participants had more fixations and longer viewing times for images in randomly ordered sequences than those in coherent visual narratives. In Experiment 2, we displayed the whole strip at once, therefore allowing us to observe any

possible between-panel shifts, as well as enabling us to check whether effects of panel order were also seen when participants could see all panels simultaneously.

Method

Participants

Fourteen participants (10 female, mean age 21 years) took part in exchange for payment. Participants had a mean VLFI score of 10.8, a low average (range, 2.6 – 26.0, SD = 7.2).

Stimuli, Apparatus and Design

The same comic strips were used as in Experiment 1, but here all panels were presented at the same time, in two rows of three. This is a typical arrangement for comic strips in the context of newspapers, in addition to being a convenient arrangement given the size of the display. Exactly the same eyetracker and recording set-up was used as in Experiment 1. As previously, originally- and randomly-ordered strips were interleaved throughout the experiment, with 48 trials in each condition. Random orders were generated uniquely for each participant and strip, and across participants each particular strip appeared in both conditions.

Procedure

The participants' instructions were exactly the same as in Experiment 1, except that participants were told that the strip started in the top left of the page and that they should press a key when they had finished reading the entire strip. A short practice session, with separate strips, familiarized the participants with the display. In particular, the "Z-path"

(Cohn, 2013a; Cohn & Campbell, 2015) arrangement of panels from left-to-right and top-to-bottom was identified. Trials began with a drift correct marker in the centre of the screen, as in Experiment 1. After each trial, they made a judgement about how difficult the strip was to understand.

Results

We began by looking at participants' manual responses, before looking at the distribution of fixations between panels.

Response measures

As previously, the sequential order of the panels affected the amount of time participants took to view the sequences, $t(13) = 4.47$, $p < .001$, $d = 1.2$, 95% CIs on $d = [1.0, 2.3]$. Participants took a mean of 9.5 seconds before terminating an originally ordered trial (SD = 2.6s), considerably shorter than in random trials (M = 11.9s, SD = 4.4s). When these viewing times are divided between the six panels in each display, they are very similar to the per-panel viewing times in Experiment 1 (i.e., 1.6 - 2.0s per panel). Participants also rated randomly ordered strips as more difficult to understand (M = 2.9, SD = 0.49) than the originally ordered sequences (M = 2.2, SD = 0.55); $t(13) = 7.2$, $p < .001$, $d = 1.9$, 95% CIs = [1.5, 3.2]).

Eye movements

Participants made a large number of eye movements while viewing a single strip (see Figure 3 for an example). Because viewing was constrained to begin in the centre of the

screen, the first one to two saccades almost always shifted fixation to the top left panel,

where viewing began. We will investigate the panel viewing sequence in more detail later

in the results, but for subsequent analyses we excluded the first fixation and considered

only those fixations following the first fixation on the upper left panel. Fixations were

then categorised according to the panel on which they occurred. Fewer than 1% of all

fixations were made to areas outside of the 6 panels.

---Figure 3 about here---

Table 2 shows key eye movement measures averaged across all panels in the

original and random conditions. As in Experiment 1, significantly more fixations were

made on each panel, on average, when strips were shown in a randomly jumbled order,

than when they were shown in the original order, $t(13) = 4.4$, $p<.001$, $d = 1.18$, 95% CIs

$= [1.0, 1.9]$. However, there was no reliable effect of condition on the mean fixation

duration or the mean saccade amplitude, both $t(13) < 1$.

It is also interesting to compare these measures with those from Experiment 1,

where panels were shown one at a time. Participants in Experiment 2 made only slightly

more fixations per panel than those in Experiment 1 (compare Tables 1 and 2;

independent samples t-test, $t(27) = 1.1$, p $= .29$). More striking, viewing in this

experiment was characterised by fixations with a shorter average duration and much

larger saccades. These differences between experiments were statistically reliable

(fixation duration: $t(27) = 3.20$, $p < .005$, d (using the pooled standard deviation) = 1.2,

95% CIs = [0.5, 2.2]; saccade amplitude: $t(27) = 18.47$, $p<.0001$, $d = 6.9$, 95% CIs = [5.9,

9.2]). It is likely that these changes reflect the size and arrangement of the display, which

in Experiment 2 allowed viewers to revisit previously viewed panels and make large

saccades between parts of the sequence.

---Table 2 about here---

In order to examine the way in which panels were inspected in more detail, we

computed two additional measures of panel processing time: the first gaze duration and

the total gaze duration. The first gaze duration (which could also be called the first-pass

inspection time) was defined as the sum duration of all fixations made on a panel the first

time that panel was fixated, and before moving to another region. The total gaze duration

(or total inspection time), meanwhile, includes the sum duration of all subsequent

fixations and therefore captures reinspections on later looks. As expected given the

greater number of fixations during random trials, the total gaze duration was reliably

affected by condition, $t(13) = 4.3$, $p < .001$, $d = 1.1$, 95% CIs=[0.9, 2.3]. Importantly, the

same difference was found with the first gaze duration, which was significantly longer in

random trials than when panels were shown in their original order, $t(13) = 4.1$, $p = .001$, d

= 1.1, 95% CIs=[0.7, 1.7]. Thus, the longer time spent on randomly-ordered panels was

not due only to coming back to them later, but manifested on the first gaze.

Did the random condition also elicit more "regressions", which move gaze backwards from a later panel to an earlier panel? About 40% of all valid saccades in the experiment began in one panel and ended in another, meaning that they acted to shift gaze between panels. Of these saccades, a larger proportion went "backwards" to a previously inspected panel in the random condition than in the original condition, $t(13) = 6.6$, $p < .001$, $d = 1.7$, 95% CIs=[1.0, 3.4].

Panel order

The results above show that, although participants often followed the sequence of the panels, they also made fairly common movements backwards to previously inspected panels. Figure 4 shows a representation of the frequency of shifts between different panels, across all participants and conditions. The reading sequence is clear enough: shifts from panel 1 were most likely to move to panel 2 and so on following the canonical "Z-path" of reading comic panel layouts (Cohn, 2013a; Cohn & Campbell, 2015). However, regressions were also common, and tended to move to the immediately preceding panel.

---Figure 4 about here---

To examine the viewing of each panel in a sequence, we computed first and total gaze durations for each condition. Unlike in Experiment 1, participants were not required

to view every panel. However, skips were very rare, and almost all panels were fixated at least once (a mean of 98% of all panels were fixated in both original and random conditions). The gaze duration results for each condition are shown in Figure 5 as a function of panel index.

---Figure 5 about here---

These data were analysed using repeated-measures ANOVA with factors of panel index and condition, separately for the total and first gaze duration. As we have seen consistently across measures, randomly ordered panels had a longer total gaze duration than those shown in the original order, $F(1, 13) = 18.0$, $p = .001$, $\eta_p^2 = .58$. There was also a main effect of panel order on total gaze duration ($F(5, 65) = 12.3$, $p < .001$, $\eta_p^2 = .49$) and a significant interaction demonstrating that the effect of order was distinct in the two conditions, $F(5, 65) = 12.1$, $p < .001$, $\eta_p^2 = .48$. There was a significant effect of order in both conditions (simple main effects, both $F(5,65) > 7$, $p < .01$). However, it was only in the random condition that the first panel received significantly more attention than subsequent panels (different from all except panel 2, Bonferonni-adjusted pairwise comparisons, ps<.01). Interestingly, the total time spent on random panels was longer than that on originally ordered panels, in all cases except for the 6th and final panel where the two conditions were not significantly different.

26

The results for first gaze duration showed a different pattern. The main effect of condition has already been described and was again statistically reliable, $F(1, 13) = 15.0$, $p = .002$, $\eta_p^2 = .54$. There was also a main effect of panel index ($F(5, 65) = 27.9$, $p < .001$, $\eta_p^2 = .68$), qualified by a significant interaction, $F(5, 65) = 5.0$, $p = .001$, $\eta_p^2 = .28$. Unlike the total gaze duration, these effects were driven by a prolonged first gaze duration on the final panel in the sequence. There were significant simple main effects of order in both original and random conditions ($Fs(5,65) > 16.9$, $p < .01$). In the original condition, the first gaze was significantly longer on the 6th panel than on all others and significantly shorter on the 5th panel (all $ps < .005$). The results were similar in the random condition (6th panel significantly different from all except panel 1, $ps < .01$). In general, the first gaze duration did not vary as much between conditions, and random and original trials were only reliably different in this respect on the 1st and 3rd panel.

Fixation locations in Experiments 1 and 2

The aim of this study was to investigate general differences in viewing visual narratives given the context or order of the panels. Our analysis so far, therefore, has not considered where people looked within a panel. Investigating fixation locations in this task is difficult given the wide range of content within the visual narratives, and predictions about what exactly participants will look at will be tested in future work. Despite these caveats, it is interesting to ask whether the overall pattern of fixations within a panel is consistent between experiments (where we manipulated the context of the strip by showing either isolated panels or a full strip). Such analysis can help address the widely held presumption that layouts in comics are tied to content (Groensteen, 2007; Bateman

& Wildfeuer, 2014), such that changing the layout also changes meaning (Postema, 2013: 29). Moreover, data from the current experiments can be used to investigate whether fixations in this task are mostly driven by the image (in which case they will be similar in both original and random orders, as predicted by a feature-based saliency model) or whether where people look also depends on the context of the visual narrative. To address these questions, we conducted a novel, data-driven analysis comparing fixation distributions in each panel.

Figure 6 shows an example of all the fixations made by participants viewing one of the panels, separately for each condition. In order to compare the fixated locations between experiments and conditions, we transformed these points into a continuous fixation distribution (e.g., Wooding, 2002), resulting in a fixation map for each panel. Maps were generated in MATLAB by adding a two-dimensional Gaussian kernel (with width approximately 2 degrees) at the location of each fixation, and then normalising the results so that each map had a constant scale from 0 to 1. Any two maps can then be compared using a two-dimensional correlation coefficient, where a strong positive correlation would indicate that the regions inspected in one map tended to be inspected in another. We used this approach in a number of comparisons.

---Figure 6 about here---

Overall fixation distributions

28

First, we compared the overall fixation distributions on a panel in Experiment 1 with the corresponding pattern from all fixations and re-fixations on the same panel in Experiment 2. Locations in Experiment 2 were re-scaled to reflect the change in panel size between experiments. Even in different panels, we would expect some correlation merely because of regularities in the position of panel elements, or because of systematic tendencies to look at certain parts of a panel. For example, it has previously been reported that there is a strong bias for people to attend to the centre of images, both within a single picture and within images in a multi-picture display (Stainer, Scott-Brown and Tatler, 2013). To take into account such biases, we calculated an additional control comparison by comparing each panel's fixation map in Experiment 1 to that from a randomly selected, different panel in Experiment 2. The results are summarised in Table 3.

In both conditions, there was a greater correlation between people looking at the same panel in different experiments than between people looking at different panels. This was true in almost all panels (97%), and the mean, between-experiment correlation was far greater than the mean control correlation (more than 2 standard deviations away in both conditions). These results confirm that our approach can capture commonalities in where people look. The control comparison, meanwhile, is still greater than zero, which is to be expected given consistencies in the position of characters and the attentional biases noted above. That both the control and the between-experiment comparison are greater in the random trials may reflect a more systematic pattern of attention in this condition.

---Table 3 about here---

Fixation distributions by condition

Next, we compared the fixation distributions for the same panels presented in the original or random order, separately for each experiment. Clearly, because the image content of the panel was the same in each condition we would expect a reasonably strong correlation. However, if the places fixated in the random panel order are different from those fixated in the original panel order, then the correlation between conditions will be lower. Specifically, the correlation between two groups of viewers looking at the same panel in *different* conditions will be lower than the correlation between two groups of viewers looking at this panel in the *same* condition. To test this, we used a "split half" technique. For each panel, we randomly sampled half of the fixations from the original condition and half of those from the random condition. Then, using the fixation map method, we compared the locations of these fixations between the two conditions, to produce a between-condition correlation. The fixations were then compared to the remaining half from the same condition, to produce estimates of the within-condition correlation. The within-condition correlation reflects how similar different fixations are within a set condition, and thus provides an upper-bound on the overall fixation behaviour we might be able to explain.

---Table 4 about here---

Table 4 shows that, on average, fixation maps from different conditions were less similar than those from the same conditions. In Experiment 1, the within-condition correlation was greater than the between-condition comparison in 73% of original strips and 78% of random strips. In Experiment 2, the within-condition correlation was greater in 63% of original strips and 82% of random strips. To compare the mean correlations, we transformed the coefficients using Fisher's z transformation and computed a paired t-test and effect sizes across panels (see Table 4). In each case, effects were medium-to-large and statistically significant. This confirms that participants looked at different regions when the same panels were presented out of sequence. It is notable that all the correlations in this analysis are weaker in Experiment 2, which may reflect larger inter-observer differences in the multi-panel displays.

The previous analysis compared each panel presented in the original condition (where it was always in a fixed sequential position) with the same panel in the random condition (where it could appear in any of the six sequential positions, randomized across participants). It is therefore possible that the difference in fixations patterns is confounded by sequential order. In the data from Experiment 1, viewing times were greater on the first panel (see Figure 2). However, when the analysis was repeated without the first panels, within-condition correlations remained significantly greater than between-condition correlations (original: $t(480)=11.1$, $p<.001$, $d=0.50$, 95% CIs=[0.42, 0.59]; random: $t(480)=17.1$, $p<.001$, $d=0.78$, 95% CIs=[0.70, 0.88]). Indeed, within-condition correlations from the original condition were consistently greater than between-

condition correlations across all panel indexes, and across both experiments. Therefore there was no evidence that sequential order, in and of itself, could explain the difference in fixation patterns.

General Discussion

In this paper we have described two experiments investigating where people look in comic strips. We applied methods and theory from text and image understanding to the reading of visual narratives. On the one hand, our aim was quite modest: to describe how people view sequential images, in a similar way to the detailed information that has been gathered from participants reading text and viewing single images. This is important both for those interested in visual narratives and information graphics, but also for those who use images as a testing-ground for natural visual attention. On the other hand, we had a number of predictions based around disrupting the overall narrative (by changing the order of panels), while keeping the content within a panel the same. Testing these predictions provides information about the degree to which attention is guided based on semantic context, as well as how the processes of attention support the visual language of comics. Our results provide clear evidence of contextual influences on attention, through the novel medium of comics, and they also lay the groundwork for future studies investigating the processing of visual narrative.

Altogether, we found a number of differences between the original and the random conditions, which indicated widespread effects of a consistent narrative on cognitive processing and the moment-to-moment allocation of attention. Randomly-ordered panels were read more slowly, and perceived as being more difficult to

understand. This was true in both experiments, and replicates previous findings that randomly ordered images are harder to comprehend and remember than those in coherent sequences (Cohn, et al., 2012; Cohn & Wittenberg, 2015; Gernsbacher, et al., 1990). Although this result was expected, it shows that the process involved in understanding a panel involves both recognition of the visible components, and some sort of integration with a coherent narrative structure (e.g., Cohn, 2013c, Cohn et al., 2012).

Of course, research comparing meaningful sentences with unordered text has shown effects of context on word reading times. For example, Morton (1964a, b) showed that reading speed increased as randomly-selected words became closer in approximation to correct English, and that the duration required to recognise a word in a sentence depends on its predictability from the context. Sentence context is a major determinant of how fast people read individual words (Pelli & Tillman, 2007). With this in mind, it is interesting to note that, in Experiment 1, panel viewing times in original and random conditions were different, even on the first panel in a sequence, which replicates findings from our early studies (Cohn & Wittenberg, 2015). As no prior context is yet available on this panel, this must be due to some panels being particularly hard to understand when appearing as the first in a randomized sequence. For example, the end result of an action (a ball falling from the sky) may require extra processing if one has not already seen the preparatory context (Charlie Brown hitting a baseball). This has been shown experimentally with longer viewing times appearing at sequence-initial panels that contained more "climactic" information than those that merely set up characters for subsequent actions (Cohn, 2014b). Consistent with this, in Experiment 1 starting panels were looked at for longer than later panels across both conditions, which we attribute to

the extra time required to "laying a foundation" (Gernsbacher, 1990) of a new situation and characters without prior context (i.e., to "set up" a new sequence).

In Experiment 1, participants viewed the sequences one panel at a time in a self-paced fashion. The pictures involved in a visual narrative can be complex, and so each panel required multiple fixations to understand. For this reason, it was important to consider differences in overt attention, even when the sequence of viewing was proscribed. The context of the sequence principally affected the number of fixations. Randomly ordered panels required more fixations, indicating increased sampling of the visual scene. A slightly increased average fixation duration appeared in random trials, although the difference was only marginally significant. There was no effect of context on average saccade amplitude in Experiment 1. This measure may therefore reflect the visual content of the images, and suggests that there was no overall change in scanning (e.g., a shift from global to local attention like that found when participants view single, scrambled images; Foulsham, Alan & Kingstone, 2011).

In Experiment 2, participants were shown visual narratives in full layouts, which meant that we could observe both the time spent initially looking at a panel and the time spent going back to previous elements. There are very few other systematic descriptions of how people navigate such narratives. Participants showed a dominant scanning pattern where they looked at the panels in order, following the left-to-right, top-to-bottom Z-path that is the default order for navigating comic page layouts (Cohn, 2013a; Cohn & Campbell, 2015). These findings go against the idea that comic pages are comprehended "holistically" (e.g., Barber, 2002) and counter the idea that readers move across layouts in "relatively erratic" ways that "do not respect any precise protocol" (Groensteen, 2007,

p. 47). Panels were rarely skipped. However, regressions were also common (accounting for about 40% of inter-panel shifts), which was different from previous eye-tracking studies of comics, where regressions were reported only for certain "non-grid" aspects of comic layout or content (Omori, et al., 2004). Our observed rate of regressions indicated that they were more common here than in text reading (where regressions occur 10-15% of the time; Rayner, 2009), but that is perhaps to be expected given that the comic images are larger and more complex than words. Regressions almost always moved to the previous panel, although some "longer-range" regressions occurred, often after the participant had reached the end of the strip.

Investigating the way that people look back in order to re-analyse the sequence of images in different conditions will be a useful measure for future research. For example, previous eyetracking studies found regressions for vertically stacked panels to the left of a long horizontal panel, because it breaks the Z-path; yet, these regressions were ameliorated by post hoc manipulations of a sequence into a grid pattern (Omori, et al., 2004). In contrast, behavioural research on participants' preferred reading orders of page layouts showed that a vertical reading of these stacked panels is much preferred to a horizontal reading (Cohn, 2013a; Cohn & Campbell, 2015). Thus, dedicated manipulation of page layouts analysed using eyetracking can help better understand these types of disparities between previous findings. In addition, analysing how attention shifts between panels can tell us about the process of integrating narrative information across a sequence. For example, making a regression to check on the meaning of a final humourous panel requires an overall understanding of the story, as well as some kind of representation of where previously seen elements were encountered. Such regressions

may be expected in sequences that feature long-distance connections between panels, as in centre-embedded clauses (Cohn, 2013c), and in sequences that require "reanalysis" of prior information (Cohn, et al., 2014). Similar processes of integration have been shown with picture-text displays by Hegarty and Just (1993) and Holsanova et al., (2009).

Most of the effects of context were replicated in Experiment 2. Participants spent longer, in total, and made more fixations on panels presented in a random order. However, fixations were much shorter when participants were scanning the whole page, which probably reflects the constraints of the panel-by-panel procedure in Experiment 1 (which may have encouraged participants to fixate for longer) as well as the occurrence of some short fixations while transitioning between panels in Experiment 2. For these reasons, and in order to look at the reinspections afforded by looking back to an earlier panel, it is useful to look at the number of regressions and more local measures of processing time.

Participants made more regressions when viewing randomly-ordered panels than when viewing coherent sequences. While this is comparable to what happens with words appearing in unpredictable or randomized contexts (Morton, 1964a), regressions between panels might occur for a range of reasons (including general narrative confusion as well as violation of grammatical rules). An increased frequency of regressions may go some way to explaining why the difference in average fixation duration shown in Experiment 1 was not found here. In Experiment 1, participants knew that they would not be able to go back to a panel, and so they spent more time processing each panel on the first pass. On the other hand, when we divided processing time into the first gaze duration and the total inspection time, effects of context remained. Panels in the random condition were initially looked at for longer, and they were also reinspected for longer. The same effects

36

were seen broadly across all six presented panels. However, the final panel in the strip received a particularly long first gaze duration, in both conditions. This could be an indication of a wrap-up effect analogous to that found in reading (Rayner, Kambe & Duffy, 2000). Importantly, this was seen in both original and random conditions and so it cannot be explained by the visual complexity of panels which terminated an original sequence.

Discussions of where people look in pictures have recently been dominated by the question of whether fixation allocation is determined by features in the image ("bottom-up", e.g., Parkhurst et al., 2002) or the viewer's task and cognition ("top-down"; e.g., Underwood & Foulsham, 2006; Henderson, Malcolm & Schandl, 2009). Our experiments provided another opportunity to consider this question. Across experiments and conditions, we presented exactly the same panels but in different sequences. This was a strong test of image-driven fixations: if the allocation of attention is mostly determined by the picture then fixation patterns on a given panel should be similar regardless of when this panel appears in a narrative. However, using a novel data-driven approach we detected differences in where people looked in the original and random condition. Thus, even in fairly simple, black and white illustrations, the items being attended changed, top-down, according to what had been seen before and what might come next. This was true across both experiments. Moreover, our data-driven approach was able to detect similarity in viewing patterns between the two experiments (i.e., in very different presentation formats). This runs counter to the notion held by various comic theorists that changes in layout produce large changes in reading and comprehension (e.g., Postema, 2013). It would also be interesting to investigate how presentation order and position on

the page might influence fixation positions, regardless of panel content. For example, a page-based central bias might change the distribution of fixations on central versus edge panels. Figure 4 suggests that some panel positions might be less likely to be regressed into, due to the standard Z-path. Such issues are ripe for further study.

One of our aims was to provide a set of measures which can be used to investigate the processing of visual narratives in future experiments, and as such we have focused on estimating effect sizes. Our results revealed widespread differences between the original and random orders, and these were seen across multiple measures and two experiments. Standardised effect size estimates, accounting for repeated measures, ranged from medium to large. In other cases, however, we can be less confident, and future studies, with more statistical power, would be necessary to confirm which measures are most informative. In particular, it will be interesting to look at fixation duration on panel elements and how this is affected by the narrative. This is one example of how, using more controlled manipulations, researchers could link prediction in comics with processes in reading (where there are detailed descriptions of lexical and syntactic variables affecting fixation durations on individual words).

Research into how people view images has proceeded with studies of disjointed, individual scenes shown one at a time to observers. These images may act as a proxy for the real world, but one of the many ways in which they are limited is that, unlike the visual environment, they are not part of a consistent (and often predictable) sequence. In the present experiments we found that the sequence context had robust effects on how attention was allocated. In this way, viewing coherently-ordered comic panels is perhaps more similar to watching a movie, where attention within a frame is more regular than

expected from static scenes (Smith et al., 2012). The attentional synchrony found in movies is partly due to a conventional editing style which takes advantage of our expectations about where things will be in the future (Smith, 2012). What drives the difference between attention in our original and random conditions in the present experiment? We have emphasised that comics provide a context and that this is one of many ways in which top-down expectations override feature-driven attention (see Foulsham, 2015). These effects could be explained by basic mechanisms such as the priming of matching features and locations between panels (thus reducing processing time) and a preference for new objects. However, behaviours such as regressions between panels and the change in processing time over a sequence suggest that comic viewing also contains specific processes akin to reading. Future investigation of these processes could reveal much about the narrative grammar believed to underlie the visual language of comics (Cohn, 2013c).

Fully understanding attentional allocation in comic-style visual narratives will require theories not only of the content of the images, but also of the conventional organisation of panels on a page, the narrative structures involved in interpreting them, and the connections between these structures. Our experiments show how the narrative structure of a comic or information graphic affects fixations on a picture. Just as context plays a role when we read words, it must also affect our attention when we "read" comics.

Acknowledgements

References

Barber, J. (2002). *The Phenomenon of Multiple Dialectics in Comics Layout*. (Masters
Thesis), London College of Printing, London.

Bateman, J. A., & Wildfeuer, J. (2014). A multimodal discourse theory of visual narrative.
Journal of Pragmatics, 74, 180-208. doi:10.1016/j.pragma.2014.10.001

Cohn, N. (2012). Comics, Linguistics, and Visual Language: The Past and Future of a
Field. In F. Bramlett (Ed.), *Linguistics and the Study of Comics* (pp. 92-118). New
York: Palgrave MacMillan.

Cohn, N. (2013a). Navigating comics: An empirical and theoretical approach to strategies
of reading comic page layouts. *Frontiers in Psychology—Cognitive Science, 4*, 1-
15.

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

Cohn, N. (2013c). Visual narrative structure. *Cognitive Science, 37*(3), 413-452.

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

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

Cohn, N., & Wittenberg E. (2015). Action starring narratives and events: Structure and
inference in visual narrative comprehension. *Journal of Cognitive Psychology*
27(7), 812-828. doi: 10.1080/20445911.2015.1051535.

Cohn, N., Jackendoff, R., Holcomb, P. J., & Kuperberg, G. R. (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, N., & Campbell, H. (2015). Navigating comics II: Constraints on the reading order of page layouts. *Applied Cognitive Psychology*, *29, 193-199*. doi: 10.1002/acp.3086.

Cohn, N., & Paczynski, M. (2013). Prediction, events, and the advantage of Agents: The processing of semantic roles in visual narrative. *Cognitive Psychology, 67*(3), 73-97.

Cohn, N., Paczynski, M., Jackendoff, R., Holcomb, P. J., & Kuperberg, G. R. (2012). (Pea)nuts and bolts of visual narrative: Structure and meaning in sequential image comprehension. *Cognitive Psychology, 65*(1), 1-38.

Cumming, G., & Finch, S. (2001). A primer on the understanding, use, and calculation of confidence intervals that are based on central and noncentral distributions. *Educational and Psychological Measurement*, *61*(4), 532-574.

Foulsham, T., Alan, R., & Kingstone, A. (2011). Scrambled eyes? Disrupting scene structure impedes focal processing and increases bottom-up guidance. *Attention, Perception, & Psychophysics*, *73*(7), 2008-2025.

Foulsham, T., Teszka, R., & Kingstone, A. (2011). Saccade control in natural images is shaped by the information visible at fixation: evidence from asymmetric gaze-contingent windows. *Attention, Perception, & Psychophysics*, *73*(1), 266-283.

Foulsham (2015). Scene Perception. In J. Fawcett, E.F. Risko and A. Kingstone (Eds.), *The Handbook of Attention*. Cambridge, MA: MIT Press.

Gernsbacher, M. A. (1985). Surface information loss in comprehension. *Cognitive Psychology*, 17, 324-363.

Gernsbacher, M. A. (1990). *Language Comprehension as Structure Building*. Hillsdale, NJ: Lawrence Earlbaum.

Gernsbacher, M. A., Varner, K. R., & Faust, M. (1990). Investigating differences in general comprehension skill. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 16*, 430-445.

Glanzer, M., Fischer, B., & Dorfman, D. (1984). Short-term storage in reading. *Journal of Verbal Learning and Verbal Behavior*, 17, 467-486.

Groensteen, T. (2007). *The System of Comics* (B. Beaty & N. Nguyen, Trans.). Jackson: University of Mississippi Press.

Haberlandt, K. (1984). Components of sentence and word reading times. In D. E. Kieras & M. A. Just (Eds.), *New methods in reading comprehension research* (pp. 219-251). Hillsdale, NJ: Erlbaum.

Henderson, J. M., Weeks Jr, P. A., & Hollingworth, A. (1999). The effects of semantic consistency on eye movements during complex scene viewing. *Journal of Experimental Psychology: Human Perception and Performance*, 25(1), 210.

Henderson, J. M., Malcolm, G. L., & Schandl, C. (2009). Searching in the dark: Cognitive relevance drives attention in real-world scenes. *Psychonomic Bulletin & Review*, 16(5), 850-856.

Holsanova, J., Holmberg, N., & Holmqvist, K. (2009). Reading information graphics: The role of spatial contiguity and dual attentional guidance. *Applied Cognitive Psychology*, *23*(9), 1215-1226.

Itti, L., & Koch, C. (2000). A saliency-based search mechanism for overt and covert shifts of visual attention. *Vision Research*, *40*(10), 1489-1506.

Kennedy, A., & Pynte, J. (2005). Parafoveal-on-foveal effects in normal reading. *Vision Research*, *45*(2), 153-168.

Kunze, K., Utsumi, Y., Shiga, Y., Kise, K., & Bulling, A. (2013). I know what you are reading: recognition of document types using mobile eye tracking. In *Proceedings of the 17th annual international symposium on wearable computers* (pp. 113-116). ACM.

Kutas, M., & Federmeier, K. D. (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.

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

Mills, M., Hollingworth, A., Van der Stigchel, S., Hoffman, L., & Dodd, M. D. (2011). Examining the influence of task set on eye movements and fixations. *Journal of Vision*, *11*(8), 17.

Mital, P. K., Smith, T. J., Hill, R. L., & Henderson, J. M. (2011). Clustering of gaze during dynamic scene viewing is predicted by motion. *Cognitive Computation*, *3*(1), 5-24.

Morris, S. B., & DeShon, R. P. (2002). Combining effect size estimates in meta-analysis with repeated measures and independent-groups designs. *Psychological Methods*, *7*(1), 105.

Morton, J. (1964a). The effects of context upon speed of reading, eye movements and eye-voice span. *Quarterly Journal of Experimental Psychology*, *16*(4), 340-354.

Morton, J. (1964b). The effects of context on the visual duration threshold for words. *British Journal of Psychology*, *55*(2), 165-180.

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

Nakazawa, J. (2005). Development of manga (comic book) literacy in children. In D. W. Shwalb, J. Nakazawa & B. J. Shwalb (Eds.), *Applied Developmental Psychology: Theory, Practice, and Research from Japan* (pp. 23-42). Greenwich, CT: Information Age Publishing.

Omori, T., Ishii, T., & Kurata, K. (2004). *Eye catchers in comics: Controlling eye movements in reading pictorial and textual media*. Paper presented at the 28th International Congress of Psychology. Retrieved from http://www.cirm.keio.ac.jp/media/contents/2004ohmori.pdf

Parkhurst, D., Law, K., & Niebur, E. (2002). Modeling the role of salience in the allocation of overt visual attention. *Vision research*, *42*(1), 107-123.

Pelli, D. G., & Tillman, K. A. (2007). Parts, wholes, and context in reading: A triple dissociation. *PLOS One*, *2*(8), e680.

Postema, B. (2013). *Narrative Structure in Comics: Making Sense of Fragments*. Rochester, NY: RIT Press.

Rayner, K., Kambe, G., & Duffy, S. A. (2000). The effect of clause wrap-up on eye
movements during reading. *The Quarterly Journal of Experimental Psychology:
Section A*, *53*(4), 1061-1080.

Rayner, K. (2009). Eye movements and attention in reading, scene perception, and visual
search. *The Quarterly Journal of Experimental Psychology*, *62*(8), 1457-1506.

Smith, T. J., Levin, D., & Cutting, J. E. (2012). A Window on Reality Perceiving Edited
Moving Images. *Current Directions in Psychological Science*, *21*(2), 107-113.

Smith, T. J. (2012). The attentional theory of cinematic continuity. *Projections*, *6*(1), 1-
27.

Stainer, M. J., Scott-Brown, K. C., & Tatler, B. W. (2013). Behavioral biases when
viewing multiplexed scenes: scene structure and frames of reference for
inspection. *Frontiers in Psychology*, *4*.

Torralba, A., Oliva, A., Castelhano, M. S., & Henderson, J. M. (2006). Contextual
guidance of eye movements and attention in real-world scenes: the role of global
features in object search. *Psychological Review*, *113*(4), 766.

Underwood, G., & Foulsham, T. (2006). Visual saliency and semantic incongruency
influence eye movements when inspecting pictures. *The Quarterly Journal of
Experimental Psychology*, *59*(11), 1931-1949.

Wooding, D. S. (2002). Eye movements of large populations: II. Deriving regions of
interest, coverage, and similarity using fixation maps. *Behavior Research
Methods, Instruments, & Computers*, *34*(4), 518-528.

Yarbus, A. L. (1967). *Eye movements and vision*. New York: Plenum Press.

	Original		Random	
	M	SD	M	SD
Number of fixations per panel	5.42	1.59	5.63	1.53
Mean fixation duration (ms)	281.2	43.4	287.4	51.5
Mean saccade amplitude (°)	3.40	0.42	3.41	0.38

Table 1. General eye movement measures taken in Experiment 1.

	Original		Random	
	M	SD	M	SD
Number of fixations per panel	5.57	1.64	6.98	2.67
Mean fixation duration (ms)	237.5	35.9	234.4	34.6
Mean saccade amplitude (°)	6.38	0.49	6.40	0.48
First gaze duration on panel (ms)	693	134	759	153
Total gaze duration on panel (ms)	1274	392	1601	638
Proportion of regressions	0.36	0.05	0.39	0.06

Table 2. Eye movement indices of attention from Experiment 2.

	Original		Random	
	M	SD	M	SD
Between Experiments 1 and 2	0.45	0.13	0.50	0.12
Between-panel control	0.16	0.07	0.18	0.07

Table 3. Correlation coefficients quantifying the similarity between fixation patterns in Experiments 1 and 2. Values show the mean and standard deviation across all 576 panels.

		Experiment 1			Experiment 2		
		M	SD	*d* [95% CIs]	M	SD	*d* [95% CIs]
Between-condition comparison	-	0.44	0.12	-	0.30	0.09	-
Within-condition comparisons	Original	0.50	0.14	0.61** [0.53, 0.69]	0.35	0.11	0.41** [0.33, 0.48]
	Random	0.52	0.12	0.73** [0.65, 0.81]	0.40	0.11	0.87** [0.80, 0.96]

Table 4. Mean and standard deviation correlation coefficients quantifying the similarity between fixation patterns. Values are averaged across all panels in each experiment. Also shown is the effect size (with 95% CIs) for the difference between within-condition similarity and the between-condition baseline. This difference was statistically significant in all cases (**$p<.001$, paired t-tests on z-transformed data, with all ts(574)>9).
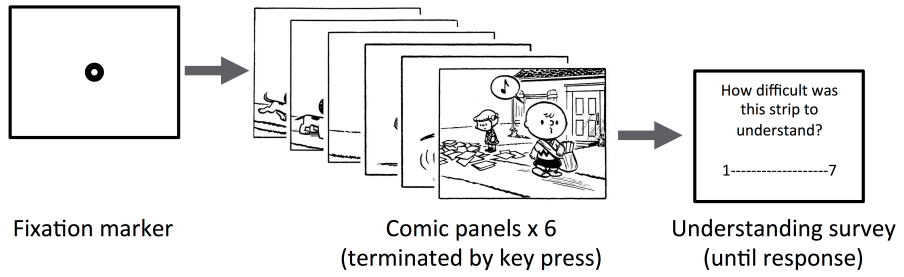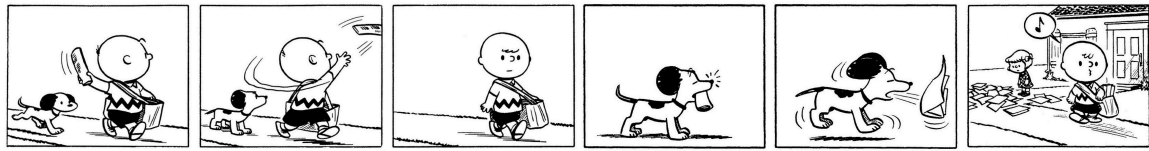
Figure Captions

Figure 1. An example comic strip (top) and the procedure for one trial in Experiment 1. After responding about the ease of understanding the panel, the next trial began.
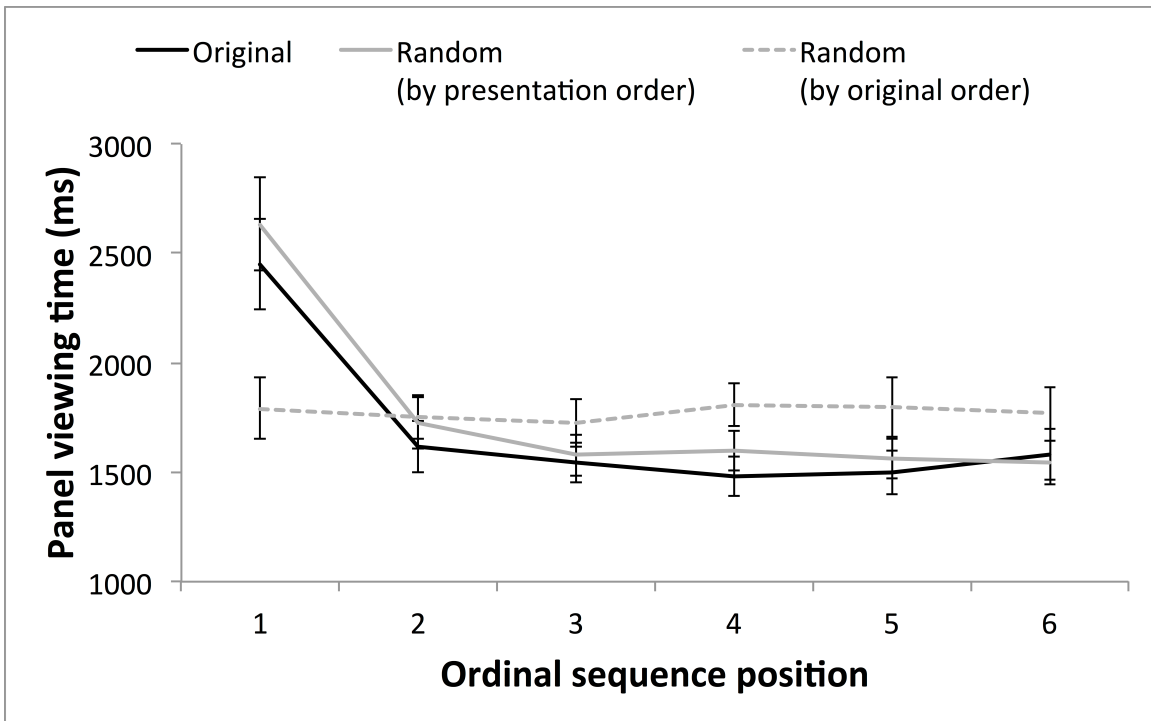
Figure 2. Panel viewing times as a function of order within the visual sequence. Data from the random condition is shown averaged across each presented order, and separately as a function of the original order for that strip. Data points show the mean across participants (with standard error bars).
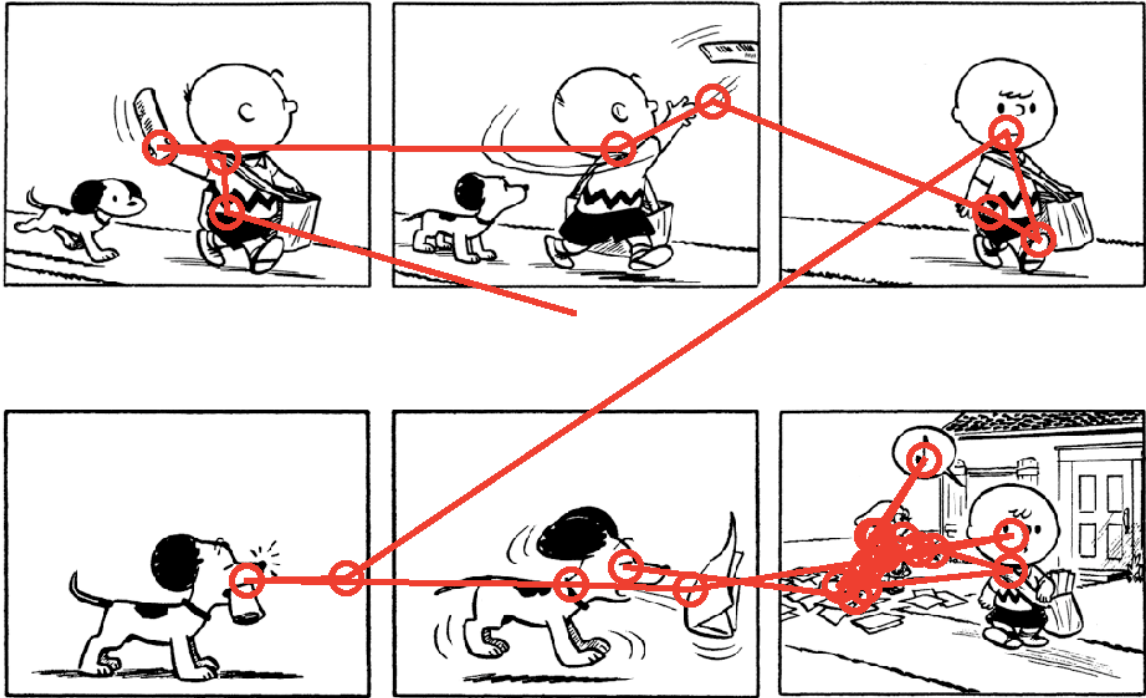
Figure 3. An example eye movement scanpath from one participant viewing a strip in Experiment 2. Viewing started in the centre (below the second panel) and subsequent saccades and fixations are shown as lines and circles, respectively.

Figure 4. The relative frequency of transitions from each panel to a different panel (left, with rows summing to 100%), and a schematic representing the most common transitions (right). Arrows are scaled proportionally to the frequency of this transition across all participants and conditions. The least frequent transitions (<10% of all shifts from a region) are omitted. Regressions that move back to a previously fixated panel are shaded in grey. Forward transitions to the next panel were most common (bold outline in table).

Figure 5. Panel gaze durations as a function of the presented order on the page, for strips in the original condition (top) and those in the random condition (bottom).

Figure 6. The fixations made on a single panel in the original (top) and random (bottom) conditions. Fixations locations from all observers are shown as dots and combined into fixation distributions (right). The correlation between the two distributions depicted is 0.77.

Figure 1. An example comic strip (top) and the procedure for one trial in Experiment 1.

After responding about the ease of understanding the panel, the next trial began.

Figure 2. Panel viewing times as a function of order within the visual sequence. Data from the random condition is shown averaged across each presented order, and separately as a function of the original order for that strip. Data points show the mean across participants (with standard error bars).

Figure 3. An example eye movement scanpath from one participant viewing a strip in Experiment 2. Viewing started in the centre (below the second panel) and subsequent saccades and fixations are shown as lines and circles, respectively.

Figure 4. The relative frequency of transitions from each panel to a different panel (left, with rows summing to 100%), and a schematic representing the most common transitions (right). Arrows are scaled proportionally to the frequency of this transition across all participants and conditions. The least frequent transitions (<10% of all shifts from a region) are omitted. Regressions that move back to a previously fixated panel are shaded in grey. Forward transitions to the next panel were most common (bold outline in table).

Figure 5. Panel gaze durations as a function of the presented order on the page, for strips in the original condition (top) and those in the random condition (bottom).
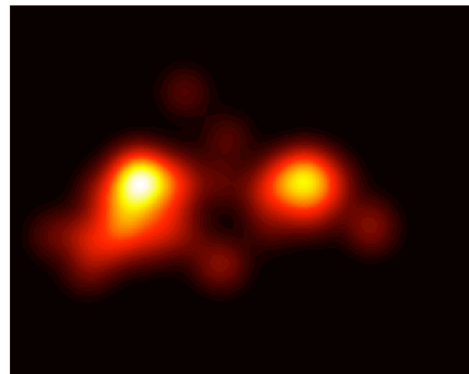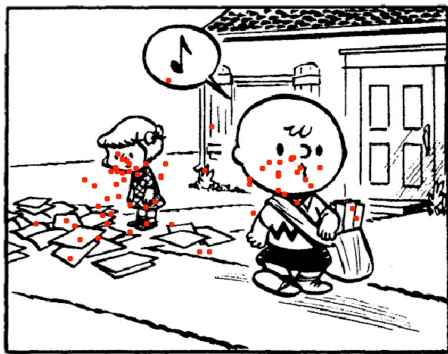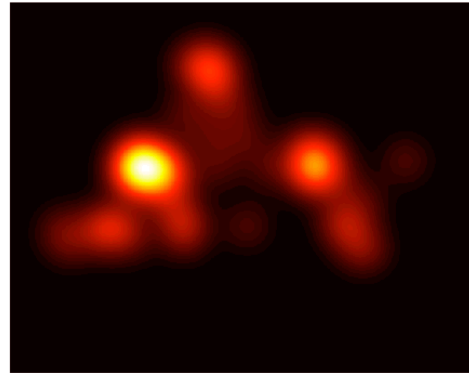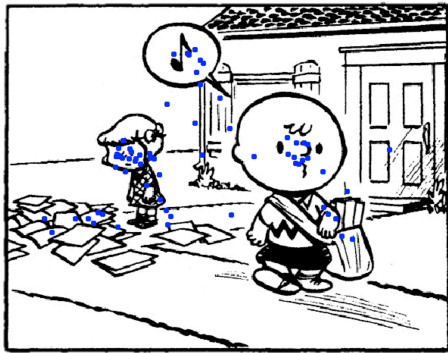
Figure 6. The fixations made on a single panel in the original (top) and random (bottom) conditions. Fixations locations from all observers are shown as dots and combined into fixation distributions (right). The correlation between the two distributions depicted is 0.77.