SEEING THE SEQUENCE: A LITERATURE REVIEW ON COMIC BOOK READING

BACHELOR THESIS ARTIFICIAL INTELLIGENCE - 7.5 ECTS

Saif Abdoelrazak 5655366
Faculty of Humanities
Artificial Intelligence
Utrecht University
Domplein 29, 3512 JE Utrecht
s.abdoelrazak2@students.uu.nl

Supervised by
Martijn van Ackooij MSc
m.vanackooij@uu.nl
And
Prof. dr. Stefan van der Stigchel
s.vanderstigchel@uu.nl

June 28, 2019

ABSTRACT

Comic book reading is a multi-modal experience that requires visual and textual processing. While these two modalities on its own have been subject to abundant research and experiments, the two modalities combined have not. In this literature review we seek to find in more detail how humans navigate through comic book pages and how the content of a page influences the reader's eye movements. Certain layouts of comic book pages influence the reader's way of navigating the page. Faces and text are able to draw the attention of the reader since these components of a comic book page are highly salient. Using the findings of the reviewed literature, a preliminary framework for a comic book eye movement predictor and a comic book generator is created. This reveals a necessity to create a more precise model of comic book reading, for which a research proposal is presented.

1 Introduction

Attention is a behavioral and cognitive process that is heavily researched within the field of neuroscience. To build highly intelligent machines, a deeper understanding of the human brain is useful. These neuroscientific topics therefore have a high relevance towards developments in artificial intelligence. Visual attention is one aspect of attention that is commonly researched. When visual attention is being researched this usually incorporates reading or picture/scene viewing. Different reading models have been created, as well as models to predict eye movements when viewing art or when reading. These two actions, scene viewing and reading, can be seen as two different modalities. Combining two modalities proves to be a difficult task for AI [1]. Therefore, it is beneficial to conduct more research in multimodal media to make further developments in artificial intelligence. The comic book is a medium that blends these two modalities into a single format. Comic books have been around since the 1800s, but the comic books we know and love today originate from the 'Golden Age Era' of Comic Books from the 1930s. Ever since, comic books have been a platform for creators to tell stories that require multi-modal thinking – meaning that comic books are a multi-modal medium that engages multiple literacies in the reader [2]. Comic book pages are separated into several panels, that are in turn separated by the gutters. Within these gutters, the reader must fill in the blanks between the comic book panels and make a connection between them. According to Jacobs in *More Than Words* [2]: ' images of people, objects, animals, and settings, word balloons, lettering, sound effects, and gutters all come together to form page layouts that work to create meaning in distinctive ways and in multiple realms of meaning making'. This is reconfirmed by Iyyer, who states that '[comic books] are not just visual: creators push their stories forward through text – speech balloons, thought clouds, and narrative boxes [. . .]' [3]. This unique multi-modality of comic books is what makes them an interesting concept to analyse from a neuroscientific perspective, which in turn makes research in this field a valuable contribution to the progress of artificial intelligence.

Analyzing how people read comic books might give us an indication of how we can make AI understand comic books better. As of now, AI is able to understand text [4] and AI is also able to correctly recognise objects and items

within images [5]. However, it has difficulties understanding comic books, as mentioned in Iyyer et al. [3]. AI has severe difficulties drawing inferences between the panels of comic books, therefore making it difficult to understand what is actually happening in these comics. Likewise, even though there are deep learning algorithms to create 'art' from regular photos or images (such as *prisma-ai* [6]) and stories (such as OpenAI's stories generator [4]), the ability for AI to create comic books is not available as of now. This might be because of the same difficulties mentioned before that are found in Iyyer et al.'s research. The multi-modal combination of text and images is already difficult for AI to understand, therefore it might be even more difficult to develop any sensible creations when combining the two modalities. Because of these difficulties, it makes sense to first get a grasp of how humans understand comic books in more detail before we can start implementing these concepts in AI to create or understand comic books.

The aim of this paper is to achieve a more coherent knowledge of how comic book pages are viewed, as well as to propose an idea for analyzing comic book pages in the future using this gathered knowledge. To substantiate this proposal, three research papers will be heavily scrutinized which will act as the groundwork for this. The first two relate to a global way of analyzing viewing behaviour over a comic book page, with a focus on panel-to-panel eye movements. The second delves into the more local way of analysing eye movements over comic book pages, namely by looking at how the eyes fixate at texts and graphics. Lastly, this paper will offer a brief preliminary blueprint of sorts for future comic book-reading AI models.

2 Previous Research

2.1 Background Information

To further understand how humans read comic book pages, it is beneficial to gather some background information on how we look at text and images, since these elements make up a comic book page. Although the way we look at text and images might not completely reflect how our eyes move over comic book pages, it can give us an indication for findings we can expect in comic book viewing.

When studying eye movements, certain terminology is important to note. *Saccades* and *fixations* are important characteristics of the movements of our eyes. Saccades are sudden jumps that our eyes make when attending a scene or reading text. Fixations are short stops that our eyes make in between these saccades. When reading, these fixations usually last about 100 to 500 ms. In reading experiments, it has been found that practically no information is extracted during the saccades , meaning the information is instead extracted during the fixations [7]. This information is however processed during the saccades, according to Holmqvist et al. [8]. Saccades are not only made from left to right (when attending to a text that is read from left to right); return sweeps are made to transport the reader's eyes to the beginning of the next line. Furthermore, readers every now and then need to move backward in text, so we make *regressions* with our saccades. *Scanning* is a way of reading where saccades are much longer and go in every direction. Scanning is used to find interesting entry points at which deeper reading can be continued.

When attending to scenes or images, eye movements are different from the movements that are made when reading. There are factors that influence our eye movements, such as top-down and bottom-up factors. Top-down factors refer to how our brain makes use of information that has already been brought to our attention, i.e. our own knowledge is an influence on how we look at things. Bottom-up factors refer to the sensory information that an object contains i.e. when an object is shown, our eyes detect the features, our brain pieces these features together which creates our perception of this object. Top-down factors have more impact on fixations compared to bottom-up factors, according to Onat et al. [9]. What this means is that the reader's current knowledge influences the way they interpret what they are seeing. People also tend to fixate more on faces when these are present in a scene or image [10]. Furthermore, an experiment in print advertisements has shown that readers prioritise text over images when text is integrated into an image and is needed for comprehension [11].

Combining these observations, it can be expected that comic book readers to prioritize text balloons and panels over other elements while reading a comic book page. In fact, in an experiment conducted by Caroll et al., it has been found that the processing of pictures and their respective text captions are two isolated events and that readers often only attend the picture until the caption has been read [12]. Thus, when combining the two modalities of text and images together, in a comic book we can expect readers to first fully read the text within a panel before they start looking at the art. Afterwards, faces and familiar elements will be fixated on with priority over other elements that are shown on a page. By exploring how people view text and images, we have covered a great part of comic book viewing, but comic books are more than only text and images. Comic books are divided into panels, which determine the flow of reading on a page. Additionally, comic book artists use different styles and techniques to creatively attract the attention of the reader and manipulate the flow in which a comic book page is read. In the next section, three different research papers will be looked at. Neil Cohn, a pioneer in comic book viewing research, has written two papers about comic book

layouts and how people navigate a comic book page through these layouts. Subsequently, in a preliminary research report by Rigaud et al., a research method for measuring eye movements over a comic book page is presented.

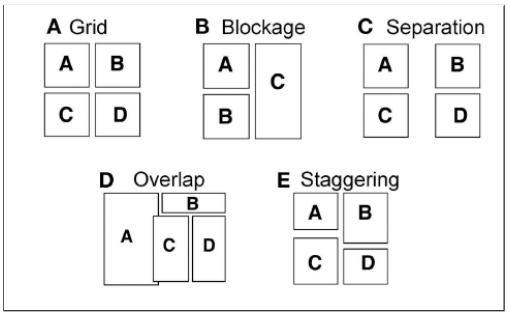

Figure 1: Manipulations of comic page layouts. Taken from [13].

2.2 Neil Cohn's Navigating Comics

In *Navigating Comics: An Empirical and Theoretical Approach To Strategies of Reading Comic Page Layouts*, Neil Cohn addresses the question as to how people know how to navigate comic page layouts, or as the author puts it: *how readers create a deliberate sequence out of the unconstrained spatial array of analog visual information*. There are several factors that contribute to how a reader traverses over a comic page, such as the colour of panels, composition within a panel, or character positioning and eye gaze. When reading a comic book, readers usually read from left-to-right and downward, also known as the 'Z-Path' [13]. This left-to-right reading orientation is assumed to come from the path that followed in a culture's written language. For instance, in Japanese manga, the comic pages are read from right-to-left.

Cohn proposes five different manipulations of comic page layouts, as seen in **Figure 1**. The manipulations are used to show the different types of layouts that challenge the usual Z-Path that is used on a conventional grid layout, line in **Figure 1A**. **Figure 1C** and **Figure 1D** show a *separation* and *overlapping* manipulation respectively. These manipulations vary the proximity of the panels. The Gestalt law of proximity states that "objects or shapes that are close to one another appear to form groups" . Therefore, these manipulations beg the question as to whether such groupings based on proximity would be preferred if the reader flouts the Z-Path. **Figure 1E** shows another challenge to the Z-Path, namely *staggering*, because the horizontal gutter terminates earlier because of the B-panel that extends ever so slightly downwards. A more extreme version of this occurs when an entire panel blocks the horizontal gutter in its entirety. This *blockage* can be seen in **Figure 1B** , in which the C-panel fully extends downwards. In this instance, following the Z-Path causes panel C to be read before panel B (as opposed to the blockage path, which is read in the ABC order). Therefore, any subsequent panel requires backtracking in the opposite direction of the Z-Path. The question that arises then is whether readers prefer to follow the blockage path or the Z-Path in a situation like this. To investigate these particular manipulations, an experiment was conducted in which participants were asked to 'read' 12 comic pages with empty panels, meaning that all the pages were devoid of any graphics (as can be seen in **Figure 3**. These stimuli pages were created specifically for this experiment, except for the last stimuli page, which is taken from a comic book page drawn by the legendary artist Jim Steranko (original seen in **Figure 2**). **Figure 3A** features a 2x3 panel grid that is predicted to be ordered in the Z-Path, since it offers the most basic and conventional type of comic page layouts. Because of this, **Figure 3A** serves as the control stimulus to be compared with the others reading strategies. As can be seen in **Figure 3**, each stimulus page except the first features one or more manipulations. The manipulations that occurred and were studied were the following:

Figure 2: Comic book page created by Jim Steranko. A version of this page with empty panels is used as a stimulus page.

1. **Blockage** A "blocking panel" blocks that path of two or more rows of panels.
2. **Separation** A large gap separates two panels. In the case of **Figure 3C** and **F**, this sponsored a vertical way of reading.
3. **Overlap** When panels overlap each other. In **Figure 3K**, the overlap across three panels could either reinforce a blockage path (the reader is guided from the bottom left panel diagonally upward) or a Z-path (the reader is guided horizontally, then down to the diagonal left).
4. **Staggering Figures 3C,E,H** show staggering in such a way that a continuation of the gutter moved vertically against the Z-Path, instead of horizontally.
5. **Insets Figure 3E** shows an inset panel. Inset panels are smaller panels within a bigger one.

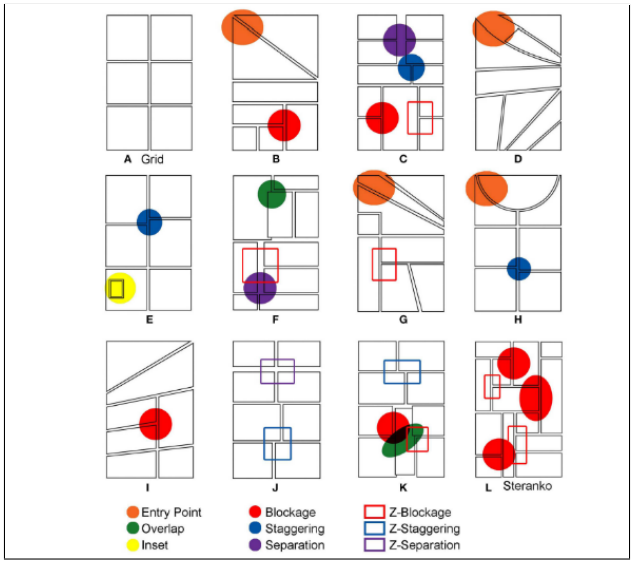

Figure 3: All Stimuli pages. Taken from [13].

6. **Entry-points** This manipulation looks into where readers enter the page that have no 'entry point'. **Figures 3B,D,G,H** show no clearly defined entry panel in the upper left corner, leaving the starting panel ambiguous.

In addition to the aforementioned panel manipulations, 'fillers' were created. These fillers showed aspects of layouts that looked similar to the challenging panel organizations, yet they did not violate the Z-Path. These fillers were introduced to give a sense of complexity to the pages, without challenging the Z-Path. These fillers can be seen in **Figure 3** under the Z-Blockage, Z-Staggering, and Z-Separation manipulations.

For each manipulation, the question was how often the Z-Path was taken, i.e. which manipulation was the most challenging to the Z-Path? In **Figure 4** the frequency of using the Z-path under the different manipulations can be seen. This confirms that the conventional grid layout offers the most straightforward reading order and can therefore justifies the usage as control manipulation. The most problematic manipulation for the Z-Path is the blockage manipulation. Omori et al. found that readers often skip over vertical panels, such as panel B in **Figure 1B**, in blockage situations [14]. They also found that when these panels were modified to appear as a horizontal path, the amount of people who skipped these panels decreased. This shows that participants prefer a horizontal Z-path in these blockage situations. It was found that the usage of the blockage path was correlated to the frequency of comic book reading of the participants. Frequent readers were more likely to follow the blockage path, whereas novice readers preferred the Z-Path. Novice readers are unfamiliar with blockage scenarios, which makes them fall back on the conventional and familiar Z-Path that is derived from text reading.

The separation manipulation is interesting because this layout might affect the reader's reading order, due to their sensitivity to the law of proximity, a Gestalt principle. It is assumed that when participants are sensitive to it, they will flout the Z-Path and order the panels that are closest together for reading. Cohn's findings however showed that the separation of panels did impact the preferences for navigational order within the participants, but not much. The participants chose the Z-Path three times more often than they chose the 'Gestalt Path', i.e. grouping the panels together. It does however play a role in navigating a page, as can be seen in in **Figure 4**.

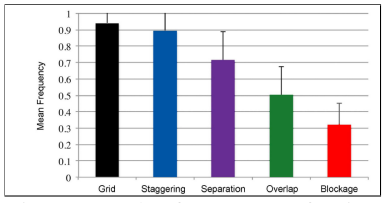

Figure 4: The frequency of using the Z-Path under each manipulation. Taken from [13].

Overlapping panels also influenced the navigation order, however the preference for flouting the Z-Path differed between the two stimuli in which overlap was presented. In **Figure 3F** and **K**, overlap was presented. **F** showed that the readers had a preference for the Z-Path, while the readers had a preference for taking the blockage path when presented with page **K**. Cohn concluded that the results imply that overlap does not provide sufficient influence to dramatically alter the preference for the Z-Path, meaning that the overlap over other panels in itself is not influential on the reader, rather the layout of the panels are.

Staggering was the manipulation that impacted the participants in the least way. This is interesting because staggering is similar to blockage. As Cohn puts it: *"they differ only in the degree to which the stagger meets the Blocking Panel"*.

That is to say, when the staggered panel would be extended, it would appear more like a blockage manipulation. This would later become the premise for the follow-up paper that will be discussed further down this section.

Lastly, the entry point of a comic book page is the same as regular text, in the sense that like regular text, comic books are also read from the top left corner. However, when the entry point is ambiguous, like in **Figure 3B, D, G, H**, it might be difficult for the reader to know where to start. Unsurprisingly, in the experiment it was found that when a panel was clearly defined in the upper left corner, this panel would be the starting point for the majority of the readers. However, when no panel was clearly defined in that position, the preference for choosing the leftmost panel dropped dramatically. The results show that more participants prefer a left-to-right and bottom-to-top reading order. People who read more manga (which is read from right to left) showed a lower frequency of using the left-to-right reading order. This is in congruence with the theory that readers of writing systems with left-to-right paths attend to different parts of a page, compared with those of right-to-left systems [15].

Using the results of each aforementioned manipulation, a set of rules on how to read comic books was created, which is a combination of a strategy called Assemblage and a set of rules called the external compositional structure preference rules, or ECSPR. External Compositional Structure of a comic book page refers to *the structure governing the organization of the physical layout of comicpages* [16]. These rules are constraints that navigate the reader through a comic page. There are four general preferences that guide Assemblage. These are:

1. Grouped areas > non-grouped areas
2. Smooth paths > broken paths
3. Do not jump over units
4. Do not leave gaps

The last Assemblage preference requires further clarification. Take for example a blockage situation. When a horizontal z-path would be taken, a 'gap' will be left in the broader shape of the panels' additive space. In other words, a part of a set of panels would be neglected, which wouldn't be when the blockage path is taken. Therefore it makes sense to follow a blockage path, and thus not leaving a gap, rather than reading a blockage situation using the Z-Path.

Combined with the principles of Assemblage, Cohn also created a set of constraints called the ECS (External Compositional Structure) Preference Rules, or ECSPR. These rules show the operations that occur in the reader when the reader is at one panel and is looking to move to the next. Firstly, there are two entry constraints, because the reader must find a starting panel before they can move on. The two constraints are:

1. ECSPR E1: Go to the top left corner
2. ECSPR E2: If no top left panel, go to either the (1) highest and/or (2) leftmost panel

Now that the entry panel is established, there are six navigational contraints that guide the reader from panel to panel.

1. ECSPR 1: Follow the outer border (Assemblage constraint 1), which is the border closest to the border of the actual page
2. ECSPR 2: Follow the inner border (Assemblage constraint 2), which are the borders that are positioned towards the inside of the actual page
3. ECSPR 3: Move to the right (Z-Path constraint 1) when ECSPR 1 or 2 can be followed
4. ECSPR 4: Move straight down (Z-Path constraint 2). Given ECSPR 3, if a rightward movement is unavailable.
5. ECSPR 5: If nothing is to the right, go to the far left and down (Z-path constraint 3)
6. ECSPR 6: Go to the panel that has not been read yet

These navigational rules could provide an initial start into the describing of the principles of comic page navigation, however they are not precise enough yet to be followed blindly. To improve these rules, a future experiment could be carried out to examine the probabilistic weights by the ECSPR to see how one rule is chosen over another by readers.

In the follow-up paper *Navigating Comics II: Constraints on the Reading Order of Comic Page Layouts*, a closer look was taken at blockage, separation , and overlap layouts [17]. In this paper, an experiment was conducted to see at what point readers switch from using the Z-Path to deviating from it. To test this, a set of page layouts were created. In these pages, the manipulations that were tested varied in their appearance, to test what specific layouts were more or less Z-path coherent. These layouts can be seen in **Figure 5**. Again, a grid layout was used as a control to test against the other experimental manipulations.

In the earlier paper, it was found that blockage was a manipulation that influenced participants the most to deviate from the Z-Path, while staggering only marginally influenced these participants. What was asked was how far down a rightward gutter needs to be placed for participants to move downward vertically, instead of horizontally (and thus following the Z-Path). The results suggested that participants are more likely to deviate from the Z-Path when a left-hand panel's bottom gutter is contiguous with its adjacent panel to the right. This is consistent with the Assemblage constraint where contiguous groupings of panels are built in order to create a smooth path of reading. These results also suggest that readers seek to build groupings of panels using cues from the contiguity of the gutters between panels. That is to say that the gutters rather than the panels themselves guide the reader from panel to panel.

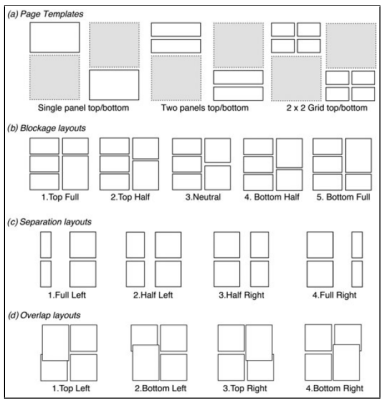

Figure 5: Page layouts using different stimuli. Row (a) depicts template pages where the gray area was filled with the layouts in (b), (c), and (d). Row (b) shows different blockage manipulations. Row (c) shows different separation manipulations. Row (d) shows different overlap manipulations. Taken from [17].

For the separation manipulation, Cohn wondered whether the increasing of the separation between the panels towards the right or left side of the page would drive participants to depart from the Z-Path. This was indeed the case, and the results showed that increasing the separation would also drive participants to depart more often from the Z-Path. The laterality of the gutter, that is to say the side at which the panels are squished to, did not matter. These findings are also consistent with the Assemblage constraint of attempting to build contiguous units of grouped areas. Readers are more inclined to follow Gestalt grouping cues rather than the Z-Path, which would require them to jump over a gap - which is another Assemblage constraint.

Overlapping also led to departures from the Z-Path, the most influential page being **Figure 5(d)**: bottom left layout. It is believed that the deviation by the overlap in this particular layout was greater because readers interact first with panels on the left than those on the right. It seems that readers create a contiguous vertical grouping from the leftward panels when they overlap. Overlap created by a bottom panel causes the top panel to not have a bottom border. Because of this, there is no contiguity with an adjacent gutter and the 'lost' bottom border. Because of this, there is a greater need to group panels when a bottom panel overlaps the top than when the top overlaps the bottom.

These results show that the aspects of spatial arrangements in comic book pages lead readers to deviate from the conventional Z-path layout, but the reasoning behind this departure is not random. Instead, it reinforces that readers seemingly follow a set of principles that drive them to navigate like this - namely, the Assemblage and ECSPR principles.

Cohn's research has predominantly been about the structure of a page, not taking into account the insides of the panels. The next research paper that will be viewed delves more into the content of panels and how we can extract information from them.

2.3 Rigaud's Semi-Automatic Text and Graphics Extraction

Manga comics are comics originating from Japan that conform to a traditional Japanese art style. The art in manga comics is usually in black and white and the comics are read from right to left. In essence, a manga comic is read like a mirrored western comic. The general way to read the pages in manga is to follow the "Reverse-Z" shape. This means that the pages are read from right to left, top to bottom. In 2016, Rigaud et al. conducted research in the field of Japanese manga comics. In the process, they presented a semi-automatic extraction method for text and graphics regions, based on eye gaze fixation and saccade analysis. They used eye-tracking on different readers on a set of manga images, which gave more insight into reading behavior regarding comic book content [18]. This research could provide us more insight for further research into eye movements over comic book pages. In Rigaud et al, a new approach for comic image analysis using eye gaze data is proposed. According to Biedert et al., the most important parts of a text document can be found by their proposed reading-skimming classifier. This classifier distinguishes actual reading from skimming [19]. Similarly, this could theoretically also be done for comic book pages by analyzing the eye gaze data for comic books.

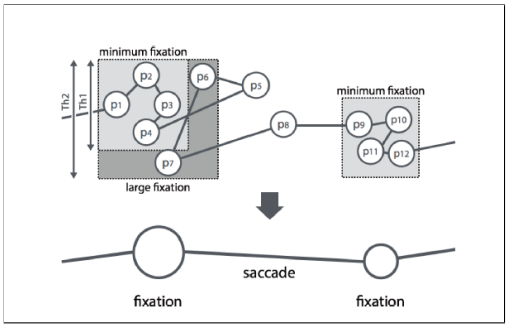

Figure 6: Two fixations detected using Rigaud's detection process. p_5 and p_8 are outside the rectangle and therefore written off as noise. Taken from [18].

To distinguish between saccades and fixations, fixations are detected when four consecutive points of attention on a location are in a rectangle of Th_1 pixels, like for example p_1, p_2, p_3, and p_4 in **Figure 6**, with Th referring to a rectangular area of pixels. The rectangle expands whenever a gaze point falls within the rectangle of Th_2 pixels. When a gaze point is outside of this rectangle, these points are written off as noise. However, when three consecutive points are out of the rectangle, the next minimum fixation is detected, starting from the first point outside the rectangle. The middle of each fixation rectangle and the corresponding timestamp are stored as a new fixation coordinate.

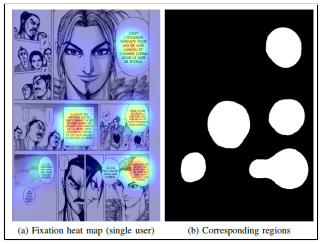
(a) Fixation heat map (single user) (b) Corresponding regions

Figure 7: The fixation heat map and the corresponding text regions. Taken from [18].

On the comic book page, the text balloons and text captions are classified as text regions. These regions are extracted by combining the position and duration of fixation points. This is done by using a Gaussian distribution centered on each fixation. From this, a heat map view can be derived. From the heat map, the text regions can be extracted using a connected component analysis on a binary segmentation of the distributions. These extracted regions only roughly correspond to the speech balloon regions, as can be seen in **Figure 7**.

A saccade map is composed from the eye tracking data. From these maps, we can clearly see which parts of the graphics region are focused on and are therefore shown to be salient. The most watched graphic regions are extracted by subtracting the text regions and background regions from the saccade path. This saccade map is essentially a trajectory map which shows us at which regions the readers fixate, while the heat map more clearly shows the duration of the fixations. The two proposed methods for text and graphics regions were then compared to two other methods. These methods were compared by calculating the recall, precision, and F-Score. The recall was calculated using the following formula: $R = \frac{TP}{TP+FN}$ and the precision with: $P = \frac{TP}{TP+FP}$, with TP = True Positive, FP = False Positive, and FN = False Negative. The authors of the paper did not specify how the F-score was calculated. The common way to calculate this is through the following formula, which can be assumed was used in the paper: $F1 = \frac{2 \cdot P \cdot R}{P+R}$.

The text region method proposed in the paper was compared with a speech balloon extractor method, created by Rigaud et al. in 2015 [20]. They found that even though the method proposed in this paper had a lower level of recall, the precision of the proposed method scored better than the speech balloon extractor method from [20]. This was because it is based on the user's eye movement, which is highly reliable.

The graphics region method proposed in the paper was compared with a method able to extract manga character faces, created by Iwata et al [21]. It was found that the recall of the proposed method was better, though the precision of the proposed method was lower, due to saccade paths being continuous and therefore overlapping other regions than the facial regions. For both extraction methods, the F-Measure or F-Score was calculated as well, determining the accuracy of the proposed methods in this paper. Both methods performed better than their comparisons, suggesting that the extraction methods based on eye gaze movements are superior. These methods can therefore be used as blueprints for future research in the field of eye movements over comic book pages.

3 Discussion and Future Research

Cohn and Rigaud et al. have confirmed the initial expectations mentioned in the introduction. Readers indeed do prioritize text and faces over other elements on the page, which could be seen in the heat maps and saccade maps that were provided in Rigaud et al's paper. Cohn showed that the panels on a comic book page do in fact influence the way of reading and gave us more insight into when people follow or flout the Z-Path. Cohn provided a set of rules for reading generic comic book layouts. These rules attempt to model human comic book reading and could therefore be relevant for AI comic book comprehension and creation. As referred to earlier in the introduction, AI has trouble understanding comic books and an AI comic book generator has yet to be created. Using the knowledge that the articles in this literature review have provided us, we could build a framework for AI comic book interpretation and creation.

As Iyyer et al. have shown, AI has trouble understanding comic books because of their inability to draw inferences between comic book panels. Therefore, it might be more valuable to first try to create a model that predicts where humans look on a comic book page. When this prediction model is successful, it can highlight the important components on a comic book page. These important components of a page can then later be used by AI to get a better understanding of the gist of a comic book story. Using Cohn's Assemblage and ECSPR, these principles and rules can be used as a guideline for AI to navigate over a comic book page in a global way. These rules can quite literally be used in this eye movement predictor. Rigaud has shown that their model is able to extract panels and backgrounds, as well as extracting text balloons. Since these guidelines model human comic book reading, they are relevant for an AI comic book eye movement predictor. A similar experiment has been done for scenes, in *Learning to Predict Where Humans Look* [22]. In this paper, Judd et al, created an eye movement predictor based on a set of 1003 images. This model was able to

make predictions using eye tracking data from these 1003 images. After training the model, the model was able to extract the salient components from scenes that were given to it.

A similar prediction model could possibly be made for comic books, which could be even more refined because of the set of rules that readers seem to follow most of the time. The problem however with these rules is that they only model eye movements from a global level. Cohn has not accounted for the sequence in which eyes move within comic book panels, while Rigaud et al. did. However, there is an oversight that Rigaud et al. did not include within their papers. The oversight is the use of colours within comic books. Rigaud's experiments were aimed towards manga comics and did therefore not contain coloured graphics, but colours in comics are widespread in Western comic books . In a study conducted by Massaro et al., colour was one of the factors that guided eye movement in naive art viewers [23]. This shows that colour could potentially be a factor that also helps a reader navigate a comic book, besides other factors such as faces and text. It makes sense, since panels that have similar colours usually group together, just like the Gestalt Principle. My expectation would be that challenging page layouts, such as the blockage scenario, would be followed in the correct order more frequently compared to when these page layouts would be empty. In order to acquire a more specific model for comic book layout viewing, a follow-up study to Cohn's *Navigating Comics II* is necessary, which will combine his comic book layouts with Rigaud's text and graphics extractor. For us to attain a more precise view of human comic book reading, three sets of comic book layouts will be studied. The first set of layouts will act as the control stimulus, in which the layouts are identical to the layouts used in *Navigating Comics II*, meaning that they are empty. This will be done to assess whether the results are similar to Cohn's. The second set of layouts will contain graphics and text, but no colour i.e. the pages will be in black and white. The third set will be identical to the second, but with colour. Eye movements will be tracked using the methods presented in Rigaud's paper, which would produce a fixation and a heat map based on the reader's eye movements. As mentioned before, a probabilistic weighted reading model could potentially be derived from this, which is useful for eye movement prediction. Readers are expected to follow the correct navigational path increasingly more with each set. The faces and text will guide the reader towards the correct panel, as would the colours, which could be seen from the fixation maps. Readers of the full colour pages would be less 'confused' when reading challenging layouts, that is to say that the right navigational path would be taken more frequently without regressions to previous panels. The main factor for this would be that salient structures on a panel would determine the flow on a page more clearly. The results of this experiment would represent human comic book layout viewing, and in essence, also human comic book reading, in a more realistic manner. Therefore it is certainly beneficial to conduct more research in comic book navigation in future studies.

These results would not only help with creating an eye movement predictor for comic books, but the development of a comic book generator would also benefit from this. A comic book generator is something that has yet to be created, because of the many complications that need to be overcome in order to create something like this. A comic book generator would need to be able to generate a story and art. Stories can already be generated through the use of OpenAI's story generator [4]. One of the many difficulties arises when the story needs to be matched with the correct art. Generating art based on text is a challenge in and of itself, since most existing models still produce unrealistic, incoherent, dream-like images. This is due to the fact that multi-modal AI creation is still very difficult, as was mentioned before. When technology advances, the predictor and generator can work together in harmony, since the data that a comic book eye viewing predictor produces can offer the comic book generator data for where certain components of an image should be placed. This is important to create a sensible comic book in which the flow between the panels is navigable.

4 Conclusion

In this literature review, the question of how human readers navigate over comics book pages has been answered. Different papers that model and analyze human eye movements across comic book pages have been carefully looked at. Comic books contain different high level and low level elements that attract the attention of our eyes, which makes us navigate through them in certain ways. Comic book layouts show how the structure of the panels can influence the reader's eye movements, and also how certain layouts present difficulties during the flow of reading. The spatial arrangement of these panels can therefore sometimes lead readers to deviate from the Z-Path. The reading behaviour of comic book readers can be attributed to a set of rules, namely the Assemblage and ECSPR principles. Furthermore, an efficient method for conducting experiments on comic book reading has been found, which also effectively can extract text and graphics from the pages. We have learned that comic book layouts can certainly influence our eye movement behaviour, as well as the use of faces, text, and evidently, colour. The topic of comic books and comic book reading has been clearly set in the landscape of artificial intelligence and its developments, stating its relevance for this area of science. For future work an interest has been taken in a more realistic model of human comic book reading. For this reason, a draft for future research has been set up, which builds upon the research articles that have been analysed in

this paper. This will in turn also prove its relevance towards artificial intelligence related developments concerning comic books.

References

[1] Baltrušaitis, T., Ahuja, C., & Morency, L. P. (2018). Multimodal machine learning: A survey and taxonomy. IEEE Transactions on Pattern Analysis and Machine Intelligence, 41(2), 423-443.

[2] Jacobs, D. (2007). More than words: Comics as a means of teaching multiple literacies. *English Journal*, 19-25.

[3] Iyyer, M., Manjunatha, V., Guha, A., Vyas, Y., Boyd-Graber, J., Daume, H., & Davis, L. S. (2017). The amazing mysteries of the gutter: Drawing inferences between panels in comic book narratives. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition* (pp. 7186-7195).

[4] Radford, A., Wu, J., Child, R., Luan, D., Amodei, D., & Sutskever, I. (2019). Language models are unsupervised multitask learners. *OpenAI Blog*, 1(8).

[5] Vision AI | Derive Image Insights via ML | Google Cloud. (n.d.). Retrieved June 9, 2019, from https://cloud.google.com/vision/

[6] Prisma Labs. (n.d.). Retrieved June 9, 2019, from https://prisma-ai.com/

[7] Wolverton, G., & Zola, D. (1983). The temporal charactersistics of visual information extraction during reading. In *K. Rayner, Perceptual and Language Processes* (pp. 41-51). New York: Academic Press.

[8] Holmqvist, K., Holsanova, J., Barthelson, M., & Lundqvist, D. (2003). Reading or scanning? A study of newspaper and net paper reading. In *The Mind's Eye* (pp. 657-670). North-Holland

[9] Onat, S., Açık, A., Schumann, F., & König, P. (2014). The contributions of image content and behavioral relevancy to overt attention. *PloS One*, 9(4), e93254.

[10] Tullis, T., Siegel, M., & Sun, E. (2009, April). Are people drawn to faces on webpages?. In *CHI'09 Extended Abstracts on Human Factors in Computing Systems* (pp. 4207-4212). ACM.

[11] Rayner, K., Rotello, C. M., Stewart, A. J., Keir, J., & Duffy, S. A. (2001). Integrating text and pictorial information: eye movements when looking at print advertisements. *Journal of experimental psychology: Applied*, 7(3), 219.

[12] Carroll, P. J., Young, J. R., & Guertin, M. S. (1992). Visual analysis of cartoons: A view from the far side. In *Eye movements and visual cognition* (pp. 444-461). Springer, New York, NY.

[13] Cohn, N. (2013). Navigating comics: an empirical and theoretical approach to strategies of reading comic page layouts. *Frontiers in psychology, 4,* 186.

[14] Omori, T., Ishii, T., & Kurata, K. (2004, August). Eye catchers in comics: Controlling eye movements in reading pictorial and textual media. In 28th international congress of psychology (pp. 8-13).

[15] Chan, T. T., & Bergen, B. (2005, July). Writing direction influences spatial cognition. In *Proceedings of the 27th annual conference of the cognitive science society* (pp. 412-417).

[16] Cohn, N. (2014). The architecture of visual narrative comprehension: the interaction of narrative structure and page layout in understanding comics. *Frontiers in Psychology*, 5, 680.

[17] Cohn, N., & Campbell, H. (2015). Navigating comics II: Constraints on the reading order of comic page layouts. *Applied Cognitive Psychology*, 29(2), 193-199.

[18] Rigaud, C., Le, N., Burie, J.-C., & Ogier, J.-M. (2016). Semi-automatic Text and Graphics Extraction of Manga Using Eye Tracking Information. *2016 12th IAPR Workshop on Document Analysis Systems*, 120-125.

[19] Biedert, R., Hees, J., Dengel, A., & Buscher, G. (2012). A robust realtime reading-skimming classifier. *Proceedings of the Symposium on Eye Tracking Research and Applications*, 120-130.

[20] Rigaud, C., Le, N., Burie, J.-C., & Ogier, J.-M. (2015). Text-independent speech balloon segmentation for comics and manga. *Proceedings of the 11th IAPR International Workshop on Graphics Recognition (GREC)*

[21] Iwata, M., Ito, A., & Kise, K. (2014, April). A study to achieve manga character retrieval method for manga images. *In 2014 11th IAPR International Workshop on Document Analysis Systems* (pp. 309-313). IEEE.

[22] Judd, T., Ehinger, K., Durand, F., & Torralba, A. (2009, September). Learning to predict where humans look. In *2009 IEEE 12th international conference on computer vision* (pp. 2106-2113). IEEE.

[23] Massaro, D., Savazzi, F., Di Dio, C., Freedberg, D., Gallese, V., Gilli, G., & Marchetti, A. (2012). When art moves the eyes: a behavioral and eye-tracking study. *PloS one*, 7(5), e37285.