FRAMESHIFT: SHIFT YOUR ATTENTION, SHIFT THE STORY

A Thesis

Submitted to the Faculty

in partial fulfillment of the requirements for the

degree of

Master of Science
in
Computer Science with a Concentration in Digital Arts

by

Tim Tregubov

in Conjunction with Rukmini Goswami

Department of Computer Science

DARTMOUTH COLLEGE

Hanover, New Hampshire

May 15, 2015

Examining Committee:

Chair _____
Lorie Loeb

Member _____
Michael Cohen

Member _____
Michael Casey

_____
F. Jon Kull, Ph.D.
Dean of Graduate Studies

ABSTRACT

Attention is a limited resource that intrinsically dictates our perceptions, memories, and behaviors. Further, visuospatial attention correlates highly with user engagement, heart rate, and arousal [El-Nasr et al., 2010]. Artists and interactive game designers strive to capture and direct attention, yet even in the most carefully crafted graphic narratives viewer eye paths – a proxy for attention – vary up to 20 percent [McCloud, 1994; Jain et al., 2012a]. Our aim is to use attentional measures to enrich graphic novel narratives. FrameShift uses eye tracking to measure reader attention and changes text and visual elements later on in the story accordingly. We have built an extensible framework for using attention to introduce perceptual changes in narratives. We use attention as an indirect method for interactions and introduce shiftable frame nodes that change readers' belief states over time.

Acknowledgements

To our ever-understanding committee members who helped us make it work.

To our families who supported us as guinea pigs in a new program.

To our friends who we fell in and out of touch with during crunch times.

To our fellow MS-DA students with whom we turned DALI into a home.

To Parag Mital without whom our references and analyses would be much shorter.

And to those we haven't acknowledged yet: if you're reading this at all, thank you!

And to Microsoft Research for giving us a Microsoft Surface to play with!

Contents

List of Tables
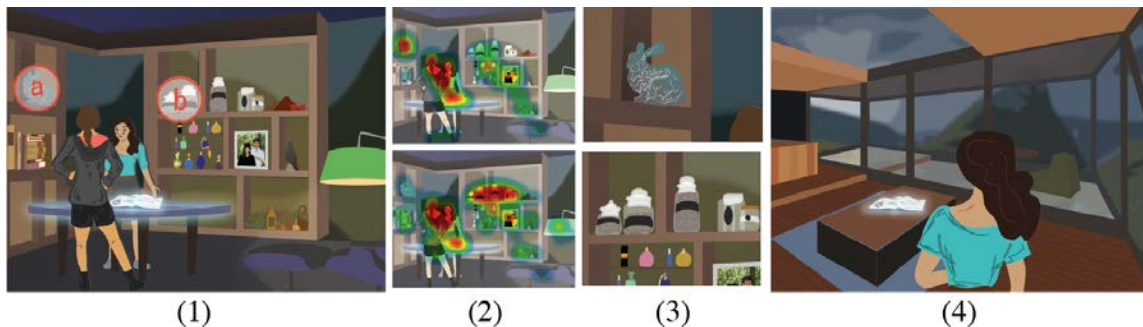
List of Figures

INTRODUCTION AND PREVIOUS WORK

Figure 1: (1) type A frame with two attention elements. (2) Heat maps of user eye fixations. Top illustrates fixation on attention element a. Bottom illustrates fixation on attention element b. (3) S type frames depend on varying eye gaze. (4) K frame displays regardless of S frame

Attention is a limited resource that intrinsically dictates our perceptions, memories, and behaviors. Further, visuospatial attention correlates highly with user engagement, heart rate, and arousal. Artists and interactive game designers alike strive to maximally capture and consciously direct the viewer's attention. This is perhaps most true of comic art, or the art of telling stories through a sequence of pictures arranged in space [McCloud, 1994]. Indeed, Will Eisner stated that the comic artist's primary goal is to "secure control of the reader's attention and dictate the sequence in which the reader will follow the narrative" [Eisner, 1985]. In a study of variability in static image eye paths, Jain and colleagues found that skilled comic artists achieve Eisner's objective. Comics, particularly those with less text (e.g. Watchmen), have far less variability in viewer eye paths than any other type of static image. However, even in these directed scenes, viewer eye paths can vary by up to 20 percent [Jain et al., 2012b]. It follows, that user attention varies in even the most carefully crafted graphic narratives. In FrameShift, we use eye tracking to measure reader attention and change text and visual elements later on in the story accordingly.

While attention and perceptual consciousness are distinct neurobiological processes, we only consciously perceive the stimuli we attend to [Koch and Tsuchiya, 2007]. Attention can be allocated in a voluntary, goal-directed manner as well as captured in an involuntary,

stimulus-driven manner. In both cases, eye tracking serves as a good proxy for attention [Hoffman and Subramaniam, 1995]. Indeed, tracking eye movements is a much better probe for attention than more indirect manual reaction time measures [Duc et al., 2008].

A considerable amount of work in Human-Computer Interaction focuses on user engagement. Designers manipulate color, lighting, reward systems, environment design, characters, etc. to elicit emotions and in turn, engagement, in their users [El-Nasr, 2007]. Nasr, et al. assessed a wide spectrum of measurement methods for validating measures of arousal in user studies. The researchers used qualitative surveys and self-reported emotions to correlate with psychophysiological measures. First, they found that eye fixation times and saccades are good probes for engagement and correlate well with heart rate and electromyography measures of arousal. Additionally and less obviously, users were more engaged and reported more enjoyment when they made implicit decisions than when they had to choose from a list of explicit decisions. One of the goals of FrameShift is to explore this realm of implicit subconscious decision-making.

Through FrameShift we have introduced a novel system which eliminates the need for explicit decision-making in a branching narrative. Instead, we have built an extensible framework for using attention as an indirect method of interaction and to introduce perceptual changes in narratives. The framework comprises a series of graphic panels or frames. A subset of the frames are defined as key panels that appear to the reader regardless of their attention. Using a set of dependency rules, further subsets of frames are set up by the author to be inserted in the narrative based on patterns of attention. If a reader spotted a clue in one frame, they may be presented with further frames that pertain to the subplot associated with that clue.

Critical to this narrative structure is that the complete set of all subplots forms the one true plot. The readers' attention modulates the various subplots. Rather than constructing an extensive branching storyline, FrameShift modifies the user's perception of the plot based on the input of their own attention. Attention is limited resource [Sternberg, 2011]

and FrameShift both exploits this fact and also attempts to elicit attentional awareness. Attentional disorders are common in modern society [Mandell et al., 2005]. Individuals with attentional deficit disorders struggle with consciously allocating their attention on a daily basis, a process those without attentional deficits often take for granted. Through FrameShift we attempt in a small way to bring awareness to that struggle. Our system utilizes a mild flashlight effect to highlight the users gaze, in part as a foil to explain the presence of the hardware eye-tracker, and in part as feedback to highlight where the user is looking. At the end of each reading experience, we inform the reader that their attention drove the narrative and how their perception of the true plot was impacted simply by how they read through the frames.

In our user studies we first validate the use of attention as a method of interaction and decision-making and establish a threshold for gauging user attention in any given panel. We then present and test a proof of concept, authored using FrameShift, for variation in belief state and user engagement. Finally, we introduce the longer length narrative currently in development.

IMPLEMENTATION

Hardware and Apparatus

FrameShift uses the Tobii EyeX eye-tracker for measuring gaze and fixations. This eye-tracker is a hardware device that attaches to the bottom of a screen or tablet. The device emits near infrared light patterns that reflect off of the users' cornea and pupil to detect motion and orientation relative to the head [AB, 2015].

Our ultimate goal is to create an accessible and enjoyable system for both authors and readers. Most graphic novels are intended to be published physically and designed for a physical medium. Existing digital comic book readers simply offer user interfaces which mimic the physical reading experience with the addition flashy animations [Made-

Fire, 2015]. We designed FrameShift from the ground up instead to be as accessible and enjoyable as possible for the lay reader.

We initially investigated several software based eye tracking options which purport to make use of the built-in cameras on tablet devices to measure eye gaze and fixations. However, none of these software options provided the accuracy required. In contrast to each of the hardware options we discuss below, the software options had accuracies upwards of 10 onscreen centimeters. This accuracy is sufficient for determining which general region the user is looking at but not specific objects.

Consequently, we investigated affordable, off-the-shelf, hardware-based eye-trackers instead of more high-end systems commonly used in psychological research. Our apparatus of choice, the EyeX, retails at $99 [Tobii, 2015] and Tobii has a line of compatible eye-trackers specifically for video games [SteelSeries, 2015]. Further, our framework is compatible with comparably priced EyeTribe eye-tracker.

The trade-off for accessibility was slightly lower accuracy and precision (though still much more accurate and precise than the software based options). Affordable eye-trackers such as this provide an average accuracy of $1°$ of visual angle and an on screen error of around 1 cm at 60 cm viewing distance [EyeTribe, 2015]. In informal tests we discovered that attentional elements need to have a minimum onscreen diameter of 3.5 cm to record reliable fixations. For each user study presented below, we define each attentional element with this diameter in mind.

Framework

Platform Our software implementation uses Unity3D as the graphics rendering platform. This provided us with 2D UI component rendering as well as basic shading and lighting for the scenes. All of our framework code is written in C# for Unity3D's .NET compatible compiler. This allows us to run on multiple platforms (with various eye-tracker support) and allows for fast gaze data event processing. This speed is critical as the the eye-tracker

hardware generates a stream of events that need to be handled in real-time.

Tracking and Attention   The Tobii EyeX Unity3D SDK provides several different event types including: gaze location, direction, eye position, and fixation start and end events. However, our experimental and framework design necessitate fixation durations and this was not provided. To overcome this we wrote a gaze event manager state machine. This state machine processes the raw gaze and fixation data streams and generates recordable fixation duration data. For the flashlight effect, we smoothed the raw gaze locations by geometrically averaging the current location with the previous one.

Narrative Framework Representation   The narrative framework is constructed from a series of graphic panels. Each panel is represented as a node in a dependency graph which in turn is represented as a *Frame* object in the Unity3D development environment. Each *Frame* loads a 2D texture for the main image which represents the graphic novel panel. Additionally each *Frame* object contains metadata about its visibility, fixation data, and *AttentionalElements*. We define *AttentionalElements* as physics-based colliders which represent regions in the 2D texture that influence the narrative. These elements can be objects or characters in *Frame* which reveal some information in the narrative as a whole. We took care to ensure the screen size of each *AttentionalElement* had a minimum diameter of 3.5 cm, in line with the accuracy of the Tobii EyeX and other comparable, affordable eye trackers.

We designate *Frame* nodes to have 3 possible properties:

- K is a key frame and **must** appear in the narrative
- S is a shiftable frame and **may** appear in the narrative
- A is a frame which has attentional elements with dependent S nodes

S frames appear in the narrative based on the state of an attentional element in a previously appearing A frame. K and S nodes can both also be A nodes.
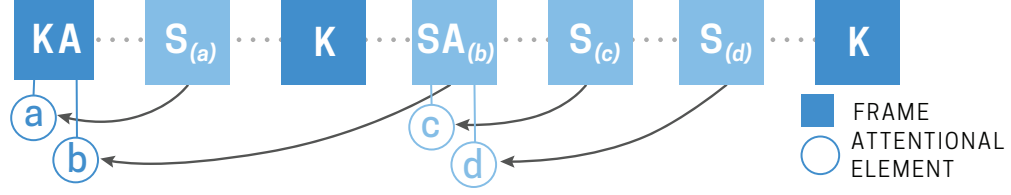
Figure 2: Dependency graph showing attentional elements and their dependent shiftable nodes.

Branching and Belief States   Based on our measurements of the reader's attention, they may or may not be shown some S nodes, thus altering their belief state in regards to the plot. Say an S node contains an unsavory fact about a character which casts doubt on that character. The reader's perception of that character changes based on the presence of that S node in the narrative that they participated in simply by attending or not attending to some elements.

AttentionalElements are further organized into *AttentionalGroups* providing a great deal of flexibility for potential narratives authored using FrameShift. Each *AttentionalGroup* contains one or more *AttentionalElements* and can be categorized by the relationships between the elements. This relationship determines whether a frame that references this *AttentionalGroup* is inserted into the narrative or not. There are three possible relationships between attentional elements: ALL, ANY, or NONE. For example, if elements are governed by the ALL relationship, a user must attend to ALL of those elements for the dependent shiftable node to display. This allows for a much greater variety of dependency behavior. Additionally each frame can have multiple *AttentionalGroups* with the same rule applied between groups. This allows for fairly complex narratives where a single shiftable frame can depend on several *AttentionalElements* being noticed across many different frames. This flexibility is demonstrated in our second User Study presented below.

Recording   For both experimental setups we record each user's eye gaze data as well as various session data such as survey questions. We record every fixation and use raycasting to determine if the user's gaze collided with any of the *AttenionalElements*. If so, we record
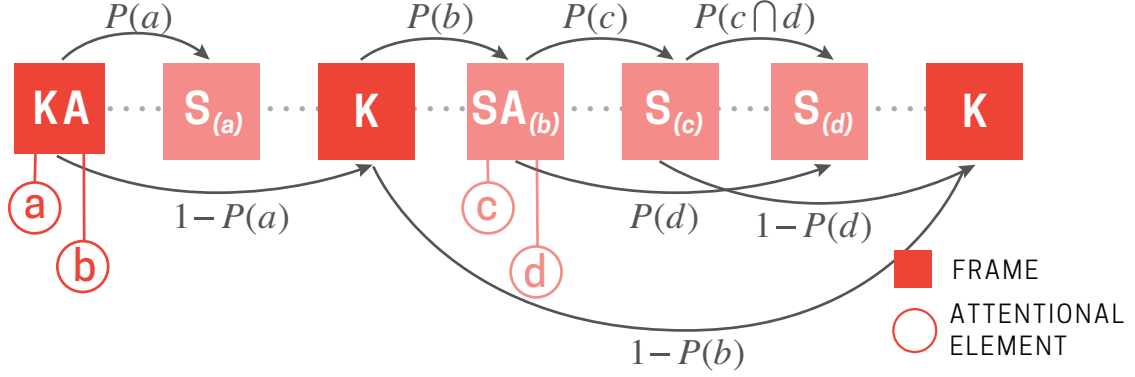
Figure 3: Example narrative paths with probabilities based on gaze data.

the *AttenionalElement* along with the fixation. We serialize every session into JSON format which we then import into MongoDB for easy analysis.

Extensibility and Authoring Ease FrameShift is intended to be an authoring tool and as such we designed many of our components to be easily extended and expose many of the narrative design tools as simple object arrangement and linking tasks that can be done in the Unity Editor. Additionally we have opened some of the framework logic to optional PlayMaker FSMs with events. This allows authors to add simple interactions and animations as well as extend the functionality of the some of the narrative behaviors.

METHODS

User Study 1: Fixation Threshold

Design Our approach necessarily relies on a total fixation duration threshold for visual recall to predict and shift user belief states throughout the narrative. We require a metric to determine with some certainty if a reader has consciously attended to any given element of a scene. Here, there is a critical distinction between conscious perception, recognition, and recall. While conscious perception and recognition are primarily determined by the neural pathways for visual short term memory, conscious recall can be used as a proxy for

visual long term memory [Luck, 2008]. Early on in the visual pathway (in area V1), the brain identifies and stores objects in short term and working memory. This is referred to as precategorical visible persistence and is available early in the visual processing stream. This visible persistence is temporary and is maintained for only several seconds [Vogel et al., 2001]. In contrast, postcategorical information, processed further down the ventral visual pathway, persists for longer durations and contributes to long term memory.

Figure 4: Example stimuli in User Study 1

To account for this visual short term versus long term memory problem, we adapted a task which requires users to process and categorize visual cues into discrete, semantic objects, an extension of Henderson's stimulus-based gaze control tasks [Henderson, 2003]. Our study replaces the real-world stimuli in Henderson's tasks with illustrated graphic panels with text cues. We hypothesize that there exists a correlation between the total fixation duration on an attentional element and the user's persistent memory of the attentional element. This task is designed to place a significantly greater cognitive load on the user than reading the ultimate graphic narrative. In a continuous narrative, attentional elements are repeated and can be placed in the context of a larger story. In contrast, the images shown in our study bore no narrative continuity with one another and showed a wide variety of environments. Informally, users reported feeling very challenged by the study – as if they were taking a memory test. We presented each image for a variable amount of time in order

to garner a wide range of total fixation durations on each element. Each image was curated and constructed manually by an artist to check for visual consistency.

Participants   Thirty participants affiliated with Dartmouth College took part in this experiment (15 women, 15 men). All participants had normal or corrected-to-normal vision and were naive with respect to the purposes of the research.

Stimuli   Twenty illustrated scenes were used as stimuli (see Appendix A). Twenty percent of the images shown were created entirely by the authors. The remaining eighty percent of the images were selected and modified from concept artwork similar to the authors' artistic style. Each scene was analyzed for feature congestion using the Rosenholtz metric for visual clutter to ensure consistency [Rosenholtz et al., 2007]. Each image fell within a range of 1.67-3.32 (mean 2.0814, standard deviation 0.437). For comparison, a cluttered real life visual scene was marked with a score of 4.4
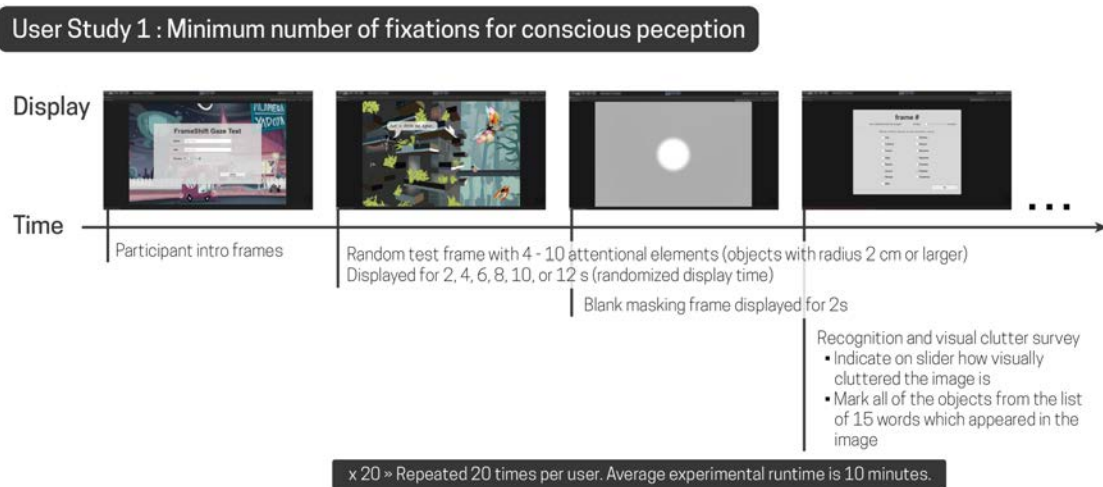

Figure 5: Fixation Threshold Experimental Setup

Procedure   Participants were first calibrated with the eye tracking software and asked for their name, age, gender. Calibration comprised having the user fixate on 9 blinking

markers at the edges and center of the display area. Calibration was considered accurate if the estimate of the user's current fixation position was within $0.5° - 1°$ degrees of visual angle of each marker. Throughout the instruction screen and the remainder of the study, a subtle flashlight was presented on the screen as feedback of their foveal gaze. If the user reported a lag or discrepancy between the flashlight cue and their actual gaze, they were re-calibrated (this occurred once among 30 users). The participants then read a description of the study along with a set of instructions. Here, they were informed that their eye movements would be monitored while they looked at scenes that they would be asked to analyze in terms of visual clutter and in terms of content. They were verbally told that the presentation time of each image would vary between 2 and 26 seconds, that it was normal to not remember much information about an image, and that there would be a questionnaire following the presentation of each image. The visual clutter data was collected but not used and primarily served as a mask of the experiments true purpose.

After the instructions were reviewed and the eye tracking software was calibrated for the user, the participant was shown 20 experimental trials. A trial consisted of the following events. First the participant fixated on the center of a blank screen to indicate he or she was ready for the trial to begin. The fixation display was then replaced by the trial scene which remained visible for a random even integer amount of time between 2 and 26 seconds and then replaced by the blanking screen for 2 seconds. After each image and blanking screen, the user was presented with a questionnaire comprising 15 words. These words consisted of the true attentional elements in each scene (4-10 elements per scene) and false attentional elements generated randomly from a word bank. The user was instructed to choose any of elements they remembered seeing in the scene. The user was allowed to remain on each questionnaire screen for as long as they needed. The order of image presentation was determined randomly for each participant. The entire study lasted approximately 15 minutes for each user. We omitted the data from the first three users (not counted toward the thirty participant total) as we altered the design of the study after they took it. We clarified

the instructions presented to the participants at the beginning and altered the wording in the trial questionnaire.

As a follow up to the study, users were informally asked to spend some time making pairwise comparisons of the images in the study in terms of visual clutter (http://tinyurl.com/opuejak). This resulted in a user-reported confirmation of the feature congestion scores previously determined by the Rosenholtz algorithm. As of this publication we have not formally analyzed the data for correlations between these feature congestion scores and accuracy scores, informally the ordering appears consistent.

User Study 2: Belief State Variation

For our second user study we wished to examine the variability in belief states engendered by our shifting panel framework described above. This study primarily served as a proof of concept and minimal viable narrative for the FrameShift authoring tool. We required a short, simple, self-contained graphic narrative with enough ambiguity to adapt into a shifting story. We discovered a children's story titled *Chomp*, authored by Melissa Mattox and illustrated by Mark Chambers [Mattox, 2014], about a young shark's first day of school. In the original narrative, Chomp the shark struggles to fit in at school because his smaller fish classmates are scared he will eat them. When it is revealed at lunchtime that Chomp is in fact a vegetarian, the day gets much better for Chomp. Because Chomp is intended for a young audience, the narrative is extremely straightforward and the illustrations are pointed. In the original illustrations, Chomp is portrayed as a very timid and scared shark who just wants to be accepted and make friends.

In the present study, Chomp's facial expressions were altered throughout the 15 panel story to be more neutral. We hypothesized that if the readers missed the critical information that Chomp is a vegetarian, they might come to a very different conclusion about Chomp's first day at school. In this way, the user belief state can range between two extremes: Chomp is a harmless vegetarian who wants to make friends or Chomp is a normal,

carnivorous, shark who wants to eat his classmates for lunch. We marked three objects in the illustrations as attentional elements and introduced four shiftable nodes dependent on those elements. This plot structure is illustrated in Figure 6 below. Based on the results of the first user study (discussed below), we set the minimum total fixation durations for our attentional elements to be 1 second.

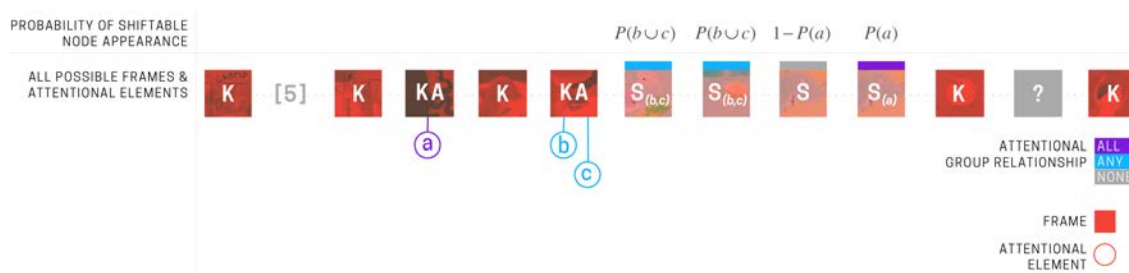The three attentional elements were in two attentional groups of differing relational types

Figure 6: Plot Diagram of User Study 2: Chomp

(Fig.6). Each attentional group (and resulting shiftable node) represents a possible shift in belief state continuum. We number the frames in terms of the order the images were presented in the original narrative. See Appendix B for the original version of the story as well as the altered, shiftable panels. The first attentional element appeared on frame 8. This element was in a group by itself with relation type ALL. If a user attended to this element, one version of frame 13 would appear, supporting the pescetarian belief state. If a user did not attend to it, a different version of frame 13 would appear, supporting the vegetarian belief state. The second two attentional elements appeared in frame 10 and were in an attentional group with relation type ANY. If a user attended to either of the elements, frames 11 and 12 would appear. If a user didn't attend to either element, an alternate version of frame 11 would appear. It follows that there exist four possible versions of this simple narrative.

Participants   Fourteen unique participants affiliated with Dartmouth College took part in this experiment. All participants had normal or corrected-to-normal vision. Thirteen of the participants were naive with respect to the purposes of the research. The fourteenth

participant took the study multiple times and was not naive to the purposes of the research. We have omitted the fourteenth participant's data from our quantitative results but not from our discussion of the user feedback.

Stimuli   Fifteen illustrated narrative panels were used as stimuli (see Appendix B). Each panel was scanned in color at a resolution of 600 dpi from the children's story *Chomp*, authored by Melissa Mattox and illustrated by Mark Chambers. Each panel was then vectorized in Adobe Illustrator, simplified, and modified to maintain visual consistency with the fixation threshold data set in User Study 1. The text was replaced with similar text in the same font as the text presented in User Study 1.

Procedure   Participants were first calibrated with the eye tracking software and were not asked for any additional demographic information. As in User Study 1, eye-tracker calibration was performed first. Throughout the instruction screen and the remainder of the study, a subtle flashlight was presented on the screen as feedback of their foveal gaze. Participants were informally told to read the story as they would naturally and to progress forward with the next button user interface element in the shape of a shark tail. At this point, participants were informed that they could not go backwards in the story or navigate with arrow keys. There were no time limits or blanking screens between the images presented. The participants progressed freely until the second to last panel when they were presented with a questionnaire screen. Here, they were asked to use adjectives to describe Chomp, what they believed would happen next, and to indicate on a slider whether they agreed or disagreed with the statement "Chomp is a killer." Participants were allowed to take as much time as needed on the question screen and were not allowed to progress to the final panel in the narrative without answering each question. Following the question panel, participants were informed that they might have missed some panels along the way based on what they did or did not attend to. If users were curious about the study, they could submit their email for a follow-up to the study.

DISCUSSION AND RESULTS

User Study 1: Fixation Threshold

Raw data files consisted of time and position values for each fixation sample. Fixations are defined previously in the Implementation section under *Hardware*. Fixation events corresponded to a frame (1-20), attentional element (null or word from list), and frequency (total number of times user fixated on the element). We consider each data sample to be a single attentional element. Each attentional element was defined by constructing a rectangular box or ellipse shaped collider around the target object. Each fixation in the scene was determined to be within or outside of the bounds of each attentional element. A typical heat map of a trial is shown in Figure 7 for a subject looking at a scene for 14s (mean presentation time). The viewing pattern shown here is consistent with the results of previous gaze fixation studies in that the majority of fixations land on or near objects or what we define as attentional elements [Henderson and Hollingworth, 1999].
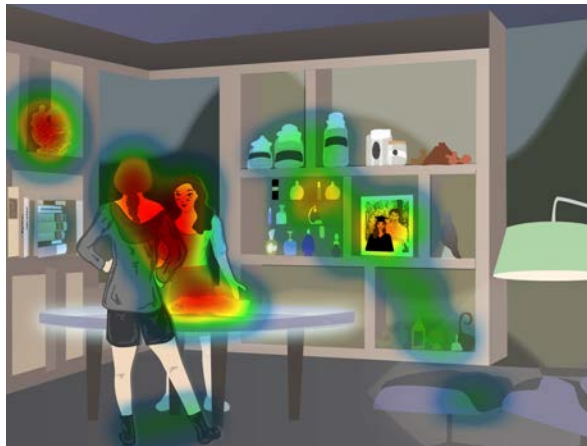

Figure 7: Heat map of a single trial

It follows that there are four cases that can occur:

- **True Positive:** the user remembers seeing the element and the eye-tracker data shows they fixated on that element

14

- **False Positive (Type I error):** the user remembers seeing an element but the eye-tracker data shows they did not fixate on the element

- **True Negative:** the user does not remember seeing an element and the element does not appear in the scene or the eye-tracker data shows they did not fixate on the element

- **False Negative (Type II error):** the user does not remember seeing an element but the eye-tracker data shows they fixated on the element

We consider our user report to be our measure of consciously attending to a location or attentional element. Using this report, we can split our population into two sub-populations of report yes and report no. We are interested in determining the probability of the user consciously attending to an element (report yes) given they fixated on it. We use posterior inference to determine this probability for the two sub-populations.

Let A = conscious recognition or report yes and B = fixation duration in seconds. We are interested in $P(A \mid B)$. From Bayes' theorem, we know that $P(A \mid B) = \frac{P(B|A)P(A)}{P(B)}$. These data are presented in Table 1 and graphed in Figures 8 and 9. From these data we can reject the null hypothesis that total fixation duration on an element has no effect on conscious recognition. The mean total fixation duration for the report yes sub-population is 1.056 seconds while the mean total fixation duration for the report no sub-population is 0.7415 seconds (Graph 1). We posit that if a user fixates on an attentional element for a duration of 1 second, there is a 63% likelihood they will consciously recognize the element. In contrast, if the user fixates on an attentional element for a duration of 0.74 seconds, there is only a 48% likelihood they will report yes to having perceived the element.

While our initial analyses are encouraging, we determined a flaw in our experimental design after the fact. In particular, our measures of false positives and true negatives are problematic. Here we map a correlation between total fixation duration and recall. We group the fixation durations into half second bins. However, there is no way to classify a

Total Fixation Duration on Element(s)	Report Yes	Report No	Total
$> 0.0 - 0.5$	523	542	1070
$> 0.5 - 1.0$	452	265	723
$> 1.0 - 1.5$	214	127	341
$> 1.5 - 2.0$	125	62	187
$> 2.0 - 2.5$	72	27	99
$> 2.5 - 3.0$	48	12	60
> 3.0	75	16	91
Total	1509	1051	2571

Table 1: Analysis of User Reports in Fixation Duration Study
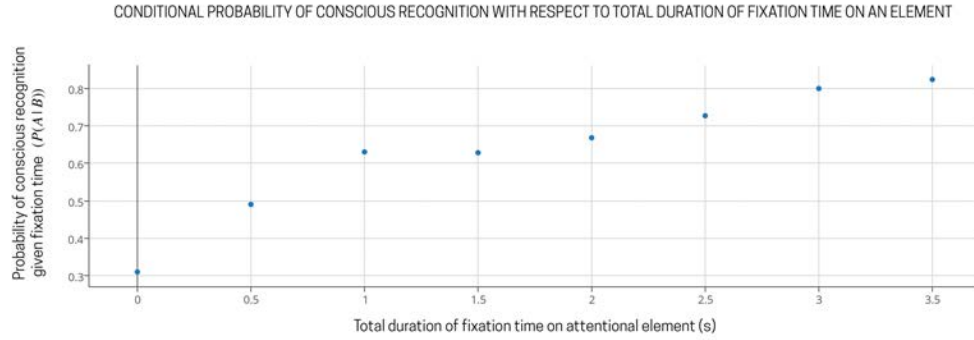
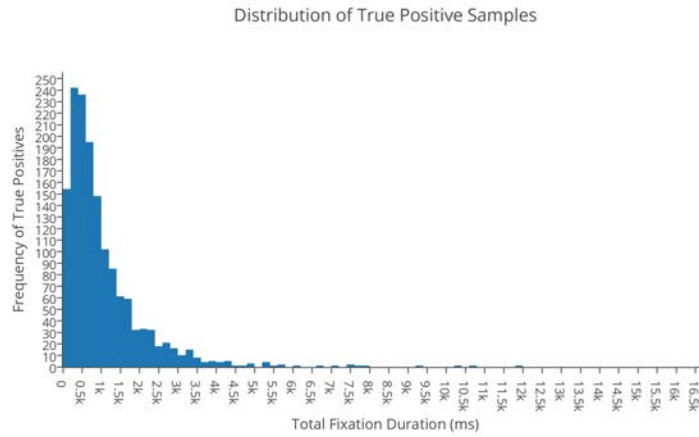Figure 8: Graph of Fixation Duration Study Results

Figure 9: Distribution of True Positive Results

negative response, i.e. the participant did not fixate on an element, in a nonzero fixation duration bin. It follows that if every user simply reported yes for every attentional element,

our probability of conscious recognition given fixation duration, would have a perfect value of 1.0 for every fixation duration. While none of participants acted in this fashion, we must consider the results above with a grain of salt.

The present research was designed to investigate eye movement behavior and attention during complex illustrated graphic novel viewing. Our primary focus was determining a threshold for total fixation duration on any given attentional element for use in our larger framework. Here we extend the work of Henderson, et al. who investigated eye movement behavior during real-world scene viewing. We posit a minimum total fixation duration of 1 second serves as a sufficient threshold for the framework and use this duration in our second user study presented below.

User Study 2: Belief State Variation

For our second user study we collected users' opinion of the shark with the question "Is Chomp a killer?" with a slider ranging from Strongly Agree to Strongly Disagree. We separated the sessions by the nodes visited. We had four potential paths but in practice the one *AttentionalElement* that lead to two of the variations (seeing a fish skeleton) was missed by 100% of the participants. Thus we only had two potential belief state variations:

- *None*: where they did not attend enough to any elements and were thus presented with the base case scenario.

- *Greens*: where they attended to either the greens or the salt *AttentionalElement* and thus saw the frames that mentioned that Chomp was vegetarian.

We split the sessions by which belief state variation they experienced and conducted an unpaired t-test between the two populations. Even with this simple narrative we had statistically significant results with a two-tailed P value of 0.0432 and thus passing the P value < 0.05 (t: 2.2602, df: 12) statistical significance test. The mean difference was -0.42 in affect

17

from the group that missed the frames that confirmed Chomp as a vegetarian to those that saw him as a nice guy.

Group	None	Greens
Mean	0.329182223909	0.747958949667
Standard Deviation (SD)	0.310879429427	0.047694993705
Standard Error of the Mean (SEM)	0.093733674768	0.027536717454
Population Size (N)	11	3

Table 2: Chomp: belief state affect between the two possible narrative paths

The adjectives and freeform reports that we also collected can be found in Appendix B.

As of publication we have not yet analyzed the data from the fixation threshold study to determine any possible correlations between feature congestion and reported recognition. However, we were careful to ensure visual consistency both within and among our data sets. We used the Rosenholtz algorithm to analyze each trial image and determine the mean feature congestion score for each user study data set [Rosenholtz et al., 2007]. Their measure of feature congestion takes into account local variability in certain key features, e.g., color, contrast, and orientation. We determined there was no statistically significant difference between the two populations with a two-tailed P value of 0.6378 (t: 0.4751, df: 33). These data are presented below in Table 3.

Group	Fixation Threshold Data Set	Belief State Data Set
Mean	2.081410000000	2.137400000000
Standard Deviation (SD)	0.437084417600	0.145996090400
Standard Error of the Mean (SEM)	0.097735046966	0.037696028449
Population Size (N)	20	15

Table 3: Average Feature Congestion Within and Among Data Sets

In this second study, we investigated the question whether a shiftable node-based narrative framework, using user attention as an implicit method of interaction: (a) provides enough flexibility as an authoring tool and (b) generates enough variation in user belief states. While the threshold we used in this second user study was contingent on our slightly

flawed experimental design in our first user study, we surpassed our expectations by generating a statistically significant variety of user belief states in our relatively small sample size of $n = 14$ thus indicating a strong effect.

Informally, we asked participants in this study their opinions regarding the attention as a method of interaction. Specifically, we asked how their reading experience compared to reading traditional graphic novels as well as to choose your own adventure novels. The majority of users preferred the implicit interactions to the explicit interactions in choose your own adventure novels. We presented participants in the study with a physical choose your own adventure graphic novel with 3,786 possible endings: *Meanwhile*, authored by Jason Shiga [Shiga, 2010]. We chose this novel because the author developed an iOS app which allows users to read it on a tablet. Overwhelmingly our participants reported more confusion while reading *Meanwhile* than while reading *Chomp*. We choose to present these findings as an informal aside given the obvious bias of users preferring the work of the present researchers over that of an unknown stranger. As of this publication, we continue to collect data in our second user study to further motivate our future work presented below.

CONCLUSION AND FUTURE WORK

The data from the present research, in conjunction with those from previous studies, converge on the following conclusions regarding user attention as an implicit method of interaction and decision-making in a shiftable node-based narrative:

- there exists a significant variety in user gaze fixation patterns and, in turn, in their ability to recognize and recall attentional elements

- there is a strong likelihood (67%) that given a user fixation duration of at least 1s on an attentional element of at least 3.5cm in diameter onscreen, presented in a complex, illustrated, scene, the user will consciously register the attentional element

- the shiftable node structure allows for a sufficient amount of variability in user belief-state

- implicit decision making in a branching narrative is preferred to explicit decision making.

Taken together, these results support the viability of FrameShift as an engaging and extensible authoring and reading framework.

FrameShift has been accepted for presentation at the ACM SIGGRAPH 2015 conference later this year. For this conference we are authoring and developing a lengthier narrative, incidentally also titled FrameShift. The artwork for this narrative is complete and several excerpt panels from this larger story are presented in Appendix D. We are on schedule to complete this lengthier narrative within the next month. In this lengthier narrative, shiftable nodes primarily serve as subplots which occur concurrently with the main timeline. As we observed in our second user study, even in a simple fifteen panel narrative, our shiftable framework introduces a significant amount of variability in user belief state. In our simple test study, we mapped this belief state variability on one axis, namely the degree to which a user agrees with the statement "Chomp is a killer." However, as narratives grow in complexity, so do the possible user belief states. After we complete the lengthier narrative, we will design possible visualizations of these complex belief states. Additionally we are adding in a cloud component as well to store readers paths and probabilities both to assist the author in determining relevance of attentional elements and to give the readers some statistics about their own reading experience as compares to others.

A further extension of the FrameShift structure could be as a resource for individuals with attentional deficit disorders. Specifically, the subtle flashlight effect which provides the user with eye gaze feedback, could be modified to increase in intensity as the user pays more attention to an element in the scene. Additionally, this could be used to illustrate the nature of attention as a limited resource to lay readers.

Figure 10: Trial images from User Study 2, ctd.

Appendix B: User Study 2

Full Chomp Study Results:

| Group	Affect	Adjectives	Next Day Plans
none	0.0516810976	"boodle boodle"	"poodle doodles"
none	0.636986	"devious"	"chomp will eat the other fish"
greens	0.8030303	"adorable, heartwarming, calming, Nemo-like, colorful, whimsical"	"ill be another shark stereotype misunderstanding that is also resolved"
greens	0.720838249	"Active, Outgoing, Determined"	"He will make more friends"
greens	0.7200083	"nervous, eager, determined. lonely, friendly, motivated, unrelenting"	"ill have a picnic/lunch party (something where they all eat together)."
none	0.74491477	"timid, lonely, scared, shy, cautious, unsure"	"thing new and make even more friends. probubly go on an adventure "
none	0.5	"Friendly"	"he will make a new friend "
none	0.0566625558	"Naive, friendly, innocent, slow"	"They will have to hunt for food / eat things"
none	0	"Smiley, Nice, Hungry"	"Chomp's going to eat more fish."
greens	0.9906599	"sad, lonely, misunderstood, nice, friendly"	"Chomps will have a better day with his new sea friends"
none	0.156289	"Shy, ruthless, deceptive"	"Chomp will eat more of his classmates"
none	0.5	"Friendly"	"He struggles with his natural hunting instincts"
none	0.114778049	"accidental fish eater"	"i think he's going to eat more fish by accident"
none	0.0525114536	"crafty"	"Chomp will eat more of the "children""
none	0.807181537	"Friendly, shy, and loyal"	"more friends and enjoy Ms Pufferfish's lesson plan, no matter what it is.
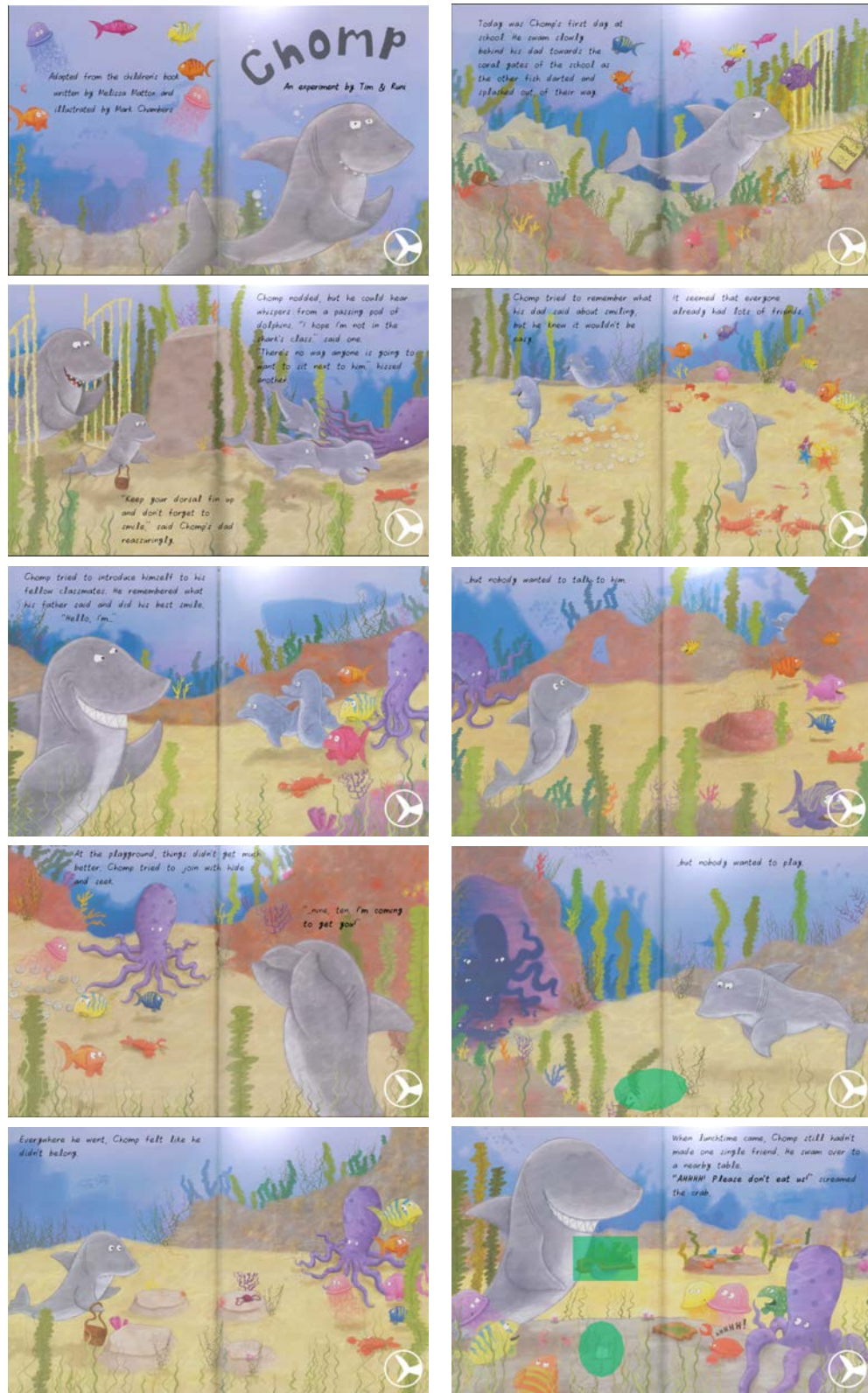
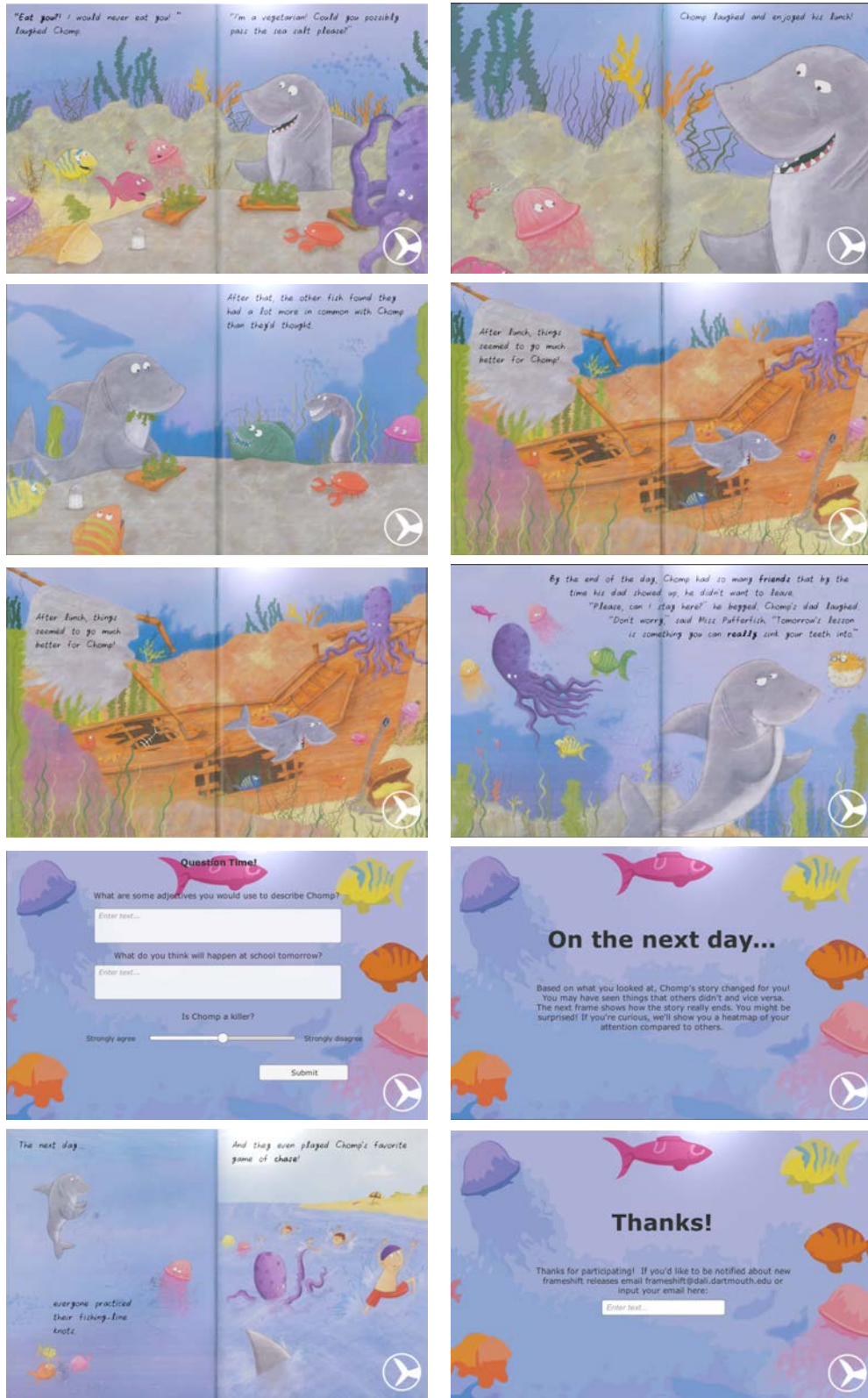Figure 11: Trial images from User Study 2

Figure 12: Trial images from User Study 2, ctd.

Appendix C: Distribution of Work

As part of the requirement for a joint thesis this sections identifies the different work performed by each of the two authors.
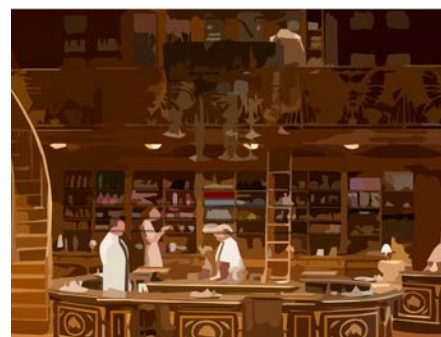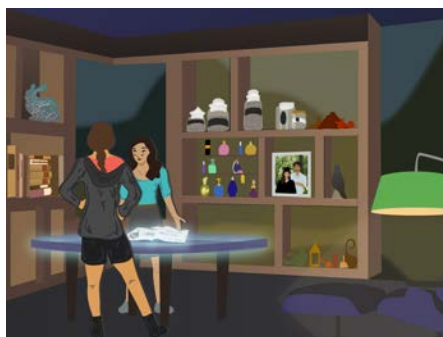
Rukmini Goswami took the lead on:

- user study design and process
- analysis of results
- experimental setup in Unity3D
- EyeTribe initial experimentation
- statistical methods
- all artwork for the future work

Tim Tregubov took the lead on:

- scaffolding out the Unity3D C# classes and framework.
- designing the narrative framework structure
- Tobii SDK integration
- gaze and fixation data processing
- stack architecture and code structure

Appendix D: Preview images from FrameShift the novel

Bibliography

Magy Seif El-Nasr, J Morie, and Anders Drachen. A scientific look at the design of aesthetically and emotionally engaging interactive entertainment experiences. 2010.

S. McCloud. *Understanding Comics*. HarperCollins, 1994. ISBN 9780060976255.

Eakta Jain, Yaser Sheikh, and Jessica Hodgins. Inferring artistic intention in comic art through viewer gaze. In *Proceedings of the ACM Symposium on Applied Perception*. ACM, 2012a.

Will Eisner. *Comics and Sequential Art*. Poorhouse Pr, 1985. ISBN 0961472804.

Eakta Jain, Yaser Sheikh, and Jessica Hodgins. Inferring artistic intention in comic art through viewer gaze. In *Proceedings of the ACM Symposium on Applied Perception*, SAP '12, pages 55–62, New York, NY, USA, 2012b. ACM. ISBN 978-1-4503-1431-2. doi: 10.1145/2338676.2338688.

Christof Koch and Naotsugu Tsuchiya. Attention and consciousness: two distinct brain processes. *Trends in cognitive sciences*, 11(1):16–22, 2007.

James E Hoffman and Baskaran Subramaniam. The role of visual attention in saccadic eye movements. *Perception & psychophysics*, 57(6):787–795, 1995.

Albert Hoang Duc, Paul Bays, and Masud Husain. .5-eye movements as a probe of attention. *Progress in brain research*, 171:403–411, 2008.

Magy Seif El-Nasr. Interaction, narrative, and drama: Creating an adaptive interactive narrative using performance arts theories. *Interaction Studies*, 8(2):209–240, 2007.

Robert J. Sternberg. *Cognitive Psychology*. Cengage Learning, 2011. ISBN 1133313914.

David S. Mandell, William W. Thompson, Eric S. Weintraub, Frank DeStefano, and Michael B. Blank. Trends in diagnosis rates for autism and adhd at hospital discharge in the context of other psychiatric diagnoses. *Psychiatric Services*, 56(1):56–62, 2005. doi: 10.1176/appi.ps.56.1.56. PMID: 15637193.

Tobii Technology AB. *An introduction to eye tracking and Tobii Eye Trackers*, 2015. http://tinyurl.com/tobii-whitepaper [Accessed: 2015].

MadeFire. *MadeFire*, 2015. http://www.madefire.com [Accessed: 2015].

Tobii. *Tobii EyeX Tracker*, 2015. http://www.tobii.com/en/eye-experience/ [Accessed: 2015].

SteelSeries. *SteelSeries Sentry*, 2015. http://steelseries.com [Accessed: 2015].

EyeTribe. *EyeTribe*, 2015. http://dev.theeyetribe.com/general/ [Accessed: 2015].

Steven J Luck. Visual short-term memory. *Visual memory*, pages 43–85, 2008.

Edward K Vogel, Geoffrey F Woodman, and Steven J Luck. Storage of features, conjunctions, and objects in visual working memory. *Journal of Experimental Psychology: Human Perception and Performance*, 27(1):92, 2001.

John M Henderson. Human gaze control during real-world scene perception. *Trends in cognitive sciences*, 7(11):498–504, 2003.

Ruth Rosenholtz, Yuanzhen Li, and Lisa Nakano. Measuring visual clutter. *Journal of vision*, 7(2):17, 2007.

Melissa Mattox. *Chomp (Paperback book and Audio CD)*. Scholastic, 2014. ISBN 0545832004.

John M Henderson and Andrew Hollingworth. The role of fixation position in detecting scene changes across saccades. *Psychological Science*, 10(5):438–443, 1999.

Jason Shiga. *Meanwhile: Pick Any Path. 3,856 Story Possibilities. (Top Ten Great Graphic Novels for Teens)*. Harry N. Abrams, 2010. ISBN 0810984237.