Sensemaking Strategies with Immersive Space to Think

Lee Lisle*
Center for Human-Computer Interaction
Department of Computer Science
Virginia Tech

Kylie Davidson†
Center for Human-Computer Interaction
Department of Computer Science
Virginia Tech

Edward J.K. Gitre‡
Center for Human-Computer Interaction
Department of History
Virginia Tech

Chris North§
Sanghani Center
Department of Computer Science
Virginia Tech

Doug A. Bowman¶
Center for Human-Computer Interaction
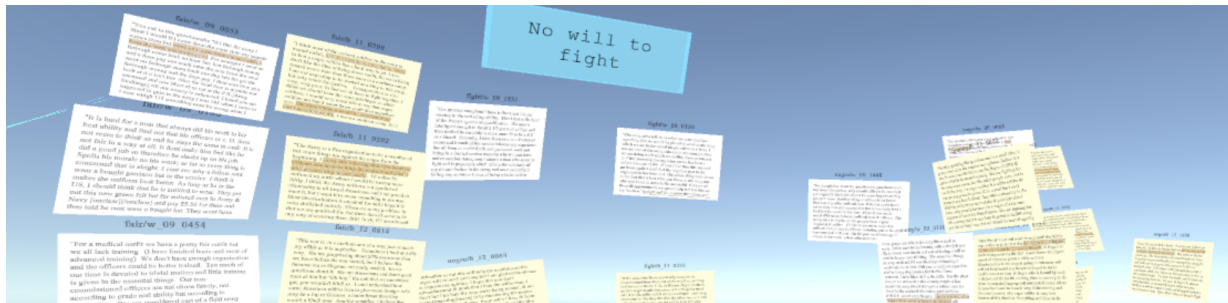Department of Computer Science
Virginia Tech

Figure 1: A labeled cluster of related documents created by a participant.

ABSTRACT

The process of sensemaking involves foraging through and extracting information from large sets of documents, and it can be a cognitively intensive task. A recent approach, the Immersive Space to Think (IST), allows analysts to browse, read, mark up documents, and use immersive 3D space to organize and label collections of documents. In this study, we observed seventeen novice analysts perform a historical analysis task in order to understand how users utilize the features of IST to extract meaning from large text-based datasets. We found three different layout strategies they employed to create meaning with the documents we provided. We further found patterns of interaction and organization that can inform future improvements to the IST approach.

Index Terms: Human-centered computing—Visualization—Visualization techniques—; Human-centered computing—Human Computer Interaction (HCI)—Interaction Paradigms—Virtual Reality

1 INTRODUCTION

Sensemaking is a cognitively difficult task when it involves foraging through large amounts of data to find meaningful items and inferring how those items relate to one another [18, 23]. For example, intelligence analysts might have many different sources such as phone records, video recordings, or pictures from a possible terrorist cell, and they need to uncover potential plots. Or perhaps a historian needs to understand the motivations of the people of a small town during the American Civil War to see how their choices changed the town by examining personal letters, newspaper clippings, and banking records. Both of these examples involve large multimedia

*e-mail: llisle@vt.edu
†e-mail: kyliedavidson@vt.edu
‡e-mail: egitre@vt.edu
§e-mail: north@cs.vt.edu
¶e-mail: dbowman@vt.edu

datasets and the task of discovering a story within the data. When an analyst begins to tackle a large dataset, they need to read and organize the dataset, corroborate themes between artifacts, and synthesize a story from the evidence into a coherent account [35]. Single data artifacts can only show small snapshots of understanding, so analysts interpret large datasets and combine artifacts to tell a complete story.

Immersive Analytics has created a new pathway to employ virtual reality (VR) and augmented reality (AR) technologies to assist with sensemaking practices. A recent approach, Immersive Space to Think (IST), aims to assist analysts by providing a virtual environment that allows users to organize virtual documents and artifacts to extract meaning. Examples of these objects can be seen in figure 1. While studies that have already been conducted suggest there is merit to this approach [3, 20], they have been limited by small dataset size, lack of advanced features to support sensemaking, or small number of participants. In the work presented in this paper, we seek to understand how novice analysts will use an advanced version of IST with a relatively large set of source documents to perform a complex sensemaking task. In particular, we are interested in how participants will make use of 3D space for organizing information and what sensemaking strategies they will employ. From our results, we expect to find ways to refine IST to better support sensemaking, and to develop hypotheses about the benefits of the IST approach that can be tested in future experiments.

We ran an exploratory study to better understand how novice users interact with the data objects and create structures of meaning within IST. The study recruited seventeen participants to perform a historical analysis task with a large dataset of 100 text documents. They were given a large virtual four by eight meter room for exploring and analyzing the documents, and markup tools (such as a highlighting tool and a note tool) to offload their cognition onto the environment and reinforce their recollection of what they had learned.

Through the experiment, we found three primary organization methods: semicircular arrangements, environment-based arrangements, and planar arrangements. All three of these arrangements are essentially two dimensional (2D) structures situated in a three dimensional (3D) space. We further analyzed how our new initial

529

document layout, where all the documents are visible at the start of the program, affected users' choices in their sensemaking. While we expect that there are multiple effective ways of using IST, these results provide a baseline for future studies.

Our study is one of the first to observe usage of a full-featured immersive analyics tool that assists users in the process of sensemaking with a large document set. We found that participants could complete a historical analysis task with IST and were more effective when they used organizational and annotation interactions within the tool. Further iterations of IST could be improved by what we learned in this study, for example by providing ideas for semantic interaction techniques that can intelligently aid the analyst.

2 RELATED WORK

2.1 Sensemaking

As many previous authors have established, sensemaking is a cognitively intensive and difficult task [1, 23, 29, 31]. The process of sensemaking involves "structuring the unknown" through the organization of data in ways that enable people to better understand and extrapolate from data [1]. Through these structures or frameworks of data extrapolation, people can have more complex understanding of the underlying data [31]. Each piece of evidence strengthens the understanding of a situation, and being able to weave these together into a story affords the ability to spread the newly gained knowledge.

Often, decisions are made with inaccurate or incomplete representations of the world, and people have to complete those representations with strategies such as storytelling to share their findings with others or fill in gaps in understanding [11]. As a corollary, when evidence is presented that challenges previous understanding, the new evidence can be taken with previous data and reframed into a more complex understanding [18]. We envision that IST will be able to assist users in such tasks, since it enables them to explain relationships through understanding the stories between various data artifacts.

Pirolli & Card defined the various stages of sensemaking and how they continually loop together allowing an analyst to better understand a set of data [23]. Their model proposes two main loops: foraging and sensemaking. The foraging loop involves analysts gathering relevant data sources and meaning, while the sensemaking loop involves the analyst understanding how those data sources fit together collectively to tell a story.

Other tools have been developed to assist with the sensemaking process. For example, Sacha et al. created a design process to assist with the creation of visualizations that convey information discovered during a sensemaking process [26]. Endert et al. made the observation that most sensemaking tools and software focus on either the foraging loop or the sensemaking loop [13]. However, we envision IST as a way to address both the foraging loop and the sensemaking loop by providing an abundance of data sources for the user to search through as well as the tools to develop an understanding of the data's relationships. Furthermore, we expect that there is more than one strategy for sensemaking with IST, and personal preferences can affect how an individual analyzes data with this approach.

2.2 Immersive Analytics

Immersive analytics is built upon the fields of data visualization, visual analytics, VR/AR, and human-computer interaction [7, 21]. Skarbez et al. argue that most current immersive analytics applications support user sensemaking through abstract data visualizations [28]. Nim et al., for example, leveraged immersive analytics to better understand and visualize bird migrations [22]. Their design allowed users to see the bird migration patterns with the geospatial data and provided context to what the birds had to face in their travels, such as environmental conditions or weather patterns. Analysts could take multiple data streams and navigate them naturally, and

could thereby perform more complex analysis. Cordeil et al. used the immersive environment to give users more control over what they were seeing through embodied data axes [10]. Users could place axes like physical objects, which changed the visualization of data in novel ways. Batch et al. followed up this work with an ecological study where they observed experts using the ImAxes tool in their normal work routine [4].

While examples such as this one use immersive analytics to view and manipulate quantitative datasets, an alternative approach is to use immersive space to explore and organize non-numeric data such as text-based documents [28]. IST's design aims to address this alternative approach to immersive analytics, while our study lets us understand how IST might assist with the sensemaking process.

The Space to Think approach has been studied in several iterations in previous work. Originally, Space to Think was a sensemaking approach that involved large-scale 2D displays that would partially wrap around the user [2]. In their study, Andrews et al. compared their large 2D display to a standard 17-inch display. They found that users managed documents in completely different ways depending upon display size. They further found that, while using the large display, users would create clusters of documents that shared meaning and explained the story behind the documents. They further found that users would leverage spatial memory to quickly find those clusters which enabled users to add to their understanding. Bandyopadhyay et al. built upon this with an Immersive Space to Think prototype [3]. They analyzed how IST would open new avenues of performing sensemaking of text-based documents in virtual reality, and found that users would create working sets of documents that they would build understanding upon. Furthermore, they created subspaces and clusters of documents to further create meaning.

2.3 Embodied & Situated Cognition

Wilson defines embodied cognition as the concept that one's cognitive processes are rooted in their interaction with the world [34]. Wilson particularly points out that part of embodied cognition is that we offload cognition onto the environment. We do this in various ways, such as writing notes in the margins of pages, highlighting key phrases or sentences in papers, or setting a timer as a reminder to perform a certain action. Wickens continues this thought process and defines these actions as ways of externalizing memory out of the mind and onto the environment [33]. Offloading cognition can also appear in the way objects are arranged. Kang et al., for example, found that participants would organize physical documents in ways that denoted additional meaning, such as timelines [17]. From these concepts, many different immersive analytics or sensemaking tools have incorporated some form of offloading cognition onto the environment, through annotation or other means [5, 9, 15].

Wilson also makes the point that cognition is situated in the environment in that we use our environment to assist in solving problems. Brown et al. further define situated cognition as "cognition that takes place in the context of task-relevant inputs and outputs," which means that we process information within the context of our surroundings and that affects how we understand concepts [6]. Both of these views are relevant to the design of virtual environments that can nurture the analytical process.

3 THE IMMERSIVE SPACE TO THINK APPROACH

IST's main contribution is the ability to interact with a large set of documents in various ways in order for the user to both extract and convey additional meaning. In Bandyopadhyay et al.'s work, users wore a tethered head-mounted display (HMD) and interacted in a small two-meter by two-meter tracked area [3]. In their virtual environment, users interacted with a set of virtual text documents that they could move around the space to organize their thinking, but the system had no other features to support sensemaking. Furthermore,
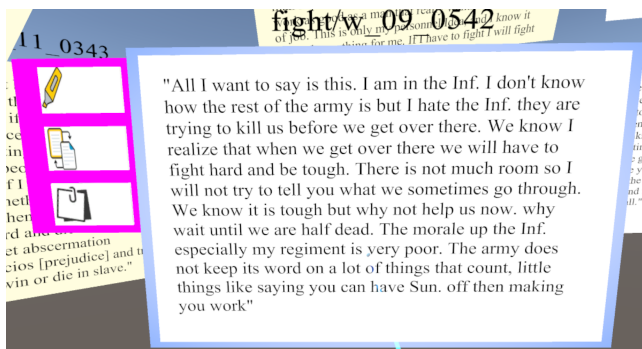
Figure 2: An example of a text document in IST. The UI panel on the left features document-specific interactions that can manipulate the document through highlighting words, copying the document, or taking notes on the document.

users complained about the blurriness of the text and the tethered HMD limiting their movement.

3.1 Improved Display and User Experience

We improved upon IST's design with new features to better capture what an analyst needs to do when performing a sensemaking task with a large document set. First, we used an untethered HMD: the HTC VIVE Pro with wireless attachment. This HMD provided high resolution and allowed users to move freely around a larger area. Previous work has shown that physical walking in VR has many benefits such as increased recall, which would prove useful for analyst tasks [19, 25]. Since we used a large tracked area, the virtual space needed to inform the user where the boundaries were. Therefore, we added a virtual floor to show users where they could safely walk. Lastly, we added a virtual bulletin board to "hang" the documents on so users could scan multiple documents at once (see figure 3).

3.2 Interaction Methods

We consulted with experts in historical analysis to design the various required interaction methods for IST. In doing so, we added ways of externalizing thought processes to offload cognition onto the virtual environment based on the feedback from the experts. Both of these groups of features are described in more detail below.

Single-hand Movement of Documents: Users can move documents by ray-casting to a document and pressing the trigger button. The document can be moved at the fixed initial distance, or be translated forward or backward along the ray by pressing on the VIVE Pro controller's trackpad in the corresponding direction.

Multi-hand Resizing and Moving of Documents: Users can also "grab" documents with both controllers simultaneously, which allows the user to resize the documents using a "pinch-to-zoom" metaphor. The user can move the controllers away from one another to make the documents larger, or towards one another to make them smaller.

Text scrolling: Text documents are often larger than the allotted visible space, so we used a scrolling metaphor that allowed the user to view additional text. To scroll a document, the user points the controller at the document, then drags their finger on the trackpad to scroll up or down in a similar fashion to scrolling on a smartphone web browser.

Text Highlighting: A key part of analyzing text documents is to make themes, keywords or quotes more salient for later use. Our solution was to allow users to highlight words. Users can point a document and press the "menu" button on the VIVE Pro controller. The word or words that intersect with the controllers ray are then selected. The user can then hold down the menu button to select entire phrases or sentences. Once they are pleased with their

selection, the user can press a software button on a UI panel that appears to the left of a document as seen in figure 2. This button makes the selection permanently highlighted.

Document Copying: Often during sensemaking, a single piece of evidence supports multiple arguments. To address this in IST, we allow for documents to be copied through pressing a software button the the document's UI panel. This will create a copy of that document that is attached to the ray extending from the controller, so the user can put the copy in a different place.

Note Taking: It is difficult to remember every insight one finds while performing sensemaking. We wanted to support external memory by allowing users to offload cognition onto the environment through notes. These notes are activated through a software button on the document's UI panel. In our implementation, notes are recorded through a Wizard-of-Oz approach where an experimenter manually types in the dictated note.

Label Creation & Keyword Search: Similarly, labels can offload cognition to the environment through a key word or phrase that represents a theme or idea. An example of a label can be seen in figure 1. Like the note-taking, entering a label is also implemented in this study through a Wizard-of-Oz technique. Labels can also be used to search for documents that contain the word or phrase on the label. Each label has their own UI panel with a button for this feature. When the user presses this button, all documents with the label's text will have their title bar highlighted. This allows the user to quickly identify the documents that contain the desired word or phrase.

4 EXPERIMENTAL DESIGN

4.1 Goals & research questions

Our study was designed to address four questions about the IST approach in order to better understand how it might assist with sensemaking.

RQ1: How do novice users use 3D immersive space in order to extract meaning from qualitative data artifacts?

We want to know what effective organization strategies users will employ with IST. This will allow us to better design interaction methods that can support these structures, as well as allow the implementation to understand what the participants might be doing and learn how to assist them in their process.

RQ2: How does our implementation of the IST concept affect users' strategies?

We have updated IST's feature set with many new additions such as highlighting, note-taking, and search, among others. We want to know if and how these new additions have affected the processes observed in previous studies and if they were effective additions to the implementation overall.

RQ3: How effective are the different strategies that we observe for sensemaking with IST?

We wish to know whether particular sensemaking strategies in IST lead to better analytic performance than others. If there are strategies that perform relatively better, we could design the IST interface and implement new features to support those strategies.

RQ4: How can we further improve the design of IST to support sensemaking?

This question seeks to better understand what might be missing from IST that could support users make better inferences from their datasets.

4.2 Apparatus

In our implementation, we use an HTC VIVE Pro HMD with a wireless attachment running on a desktop PC with an Intel i7-8700k processor and an NVIDIA 1070 graphics card. The user holds two VIVE Pro wireless controllers to interact with the documents. User movement is tracked by a SteamVR 2.0 Lighthouse tracking system covering a four-by-eight meter space that was kept clear of obstacles.

Figure 3: This is the initial state of IST when the user first enters the virtual environment. The floor represents the tracked area the user could traverse, and all the documents are initially displayed in their categories on the bulletin board.

4.3 Experimental Task

The goal of our study was to understand how novice users organize and make sense of a large set of discrete data artifacts. We decided to use a historical analysis task that would be completed by students taking a history class at our university. We selected a set of 100 responses to an open-ended question from a survey of American soldiers conducted by the US military during World War II. The survey responses were available through The American Soldier Project [14]. The survey asked soldiers about race relations in the armed forces. The theme of the survey and racial identification of the responses also informed the prompts we asked participants to answer as the historical analysis task:

1. According to the soldiers who responded, what ought to be the overriding consideration in the Allied war effort, pragmatism or principle? Should America be fighting for the principles of democracy wherever those principles are threatened or violated, even in America itself; or should the country focus solely on winning the war to end it as soon as possible?

2. How do these views differ based on the respondent's racial identification? What do they indicate about the state of racial relations across the armed forces?

We designed these prompts to require close reading of the survey responses, as well as have the participants interpret themes across multiple responses. Since we wanted to limit their time in VR to around 40 minutes to prevent fatigue, we did not require participants to actually create a written response to the prompts; instead, we asked them to deliver a verbal outline at the end of their analysis session.

Since we wanted to simulate a typical historical analysis process, we pre-selected 100 documents from the database of over 2500 responses. These documents were chosen based on searches for five keywords that were recommended by a subject specialist. This represents a typical process where the analyst prunes irrelevant documents before more closely analyzing the remaining documents. Within the documents chosen for each category we selected ten responses from white soldiers and ten from black soldiers. The five keywords were: White, Negro, Fight, Fair, and Country. Note that "Negro" was used since that term was commonly used in the source documents, by soldiers in the World War II era, to describe people of African descent. However, because of the offensive nature of this term, for the remainder of the paper we refer to this category as "Black."

Each response was displayed in IST as a virtual document, as seen in figure 2. Each document's title had the keyword it was associated with combined with the race of the respondent (either w or

b for white or black) and the ID number for the response derived from cataloging data. The race of the respondent was also encoded in the background color of the document: we used a background color of white for white soldiers and yellow for black soldiers. In the example from figure 2, the keyword was "fight," the response was from a white soldier (as seen in both the title containing "w_" and its background color being white), and the id number of the document was 09_0542. We should note that the respondent encoding is different than the categories "White" and "Black." For example, a yellow document in the White category denotes a black soldier's response (document color) which talks about white soldiers (category). Background colors were used so the user could quickly identify a key attribute of each document.

4.4 Participants

We recruited 24 participants for our study from an undergraduate history course on World War II. However, due to software and hardware issues we lost audio recordings for seven participants. This left us with seventeen participants (3 female) with a mean age of 20.3 (standard deviation of 0.8). Seven of these participants were majoring in political science, while the rest were from various other fields. Nine of them had no previous VR or AR experience, while seven of the remaining eight had only used VR or AR once or twice. Three wore contact lenses during the experiment, while four used glasses, and the remaining ten had good uncorrected vision. The experiment was approved by the university's institutional review board.

4.5 Procedure

To address our goal of understanding how novice users utilize space in IST, we split our study into five phases: a pre-study phase, a training phase, the main phase, an in-VR interview phase, and a post-experiment interview phase. These are described in detail below.

The pre-study phase involved welcoming the participant, presenting them with an informed consent to read and sign, and getting them comfortable in the physical space. We also had the participants answer a brief background questionnaire to better understand their experiences with VR/AR, their fatigue level, their field of study, and demographic information. We also introduced the concept behind IST to the participants. This phase typically lasted five to ten minutes.

In the training phase we introduced the participants to IST's environment (as seen in figure 3) and taught them how to use the controls of IST. We used a set of CNN article transcriptions in this phase, so as not to reveal the main document dataset. We showed the participants the boundaries of the tracked area. We showed participants each feature as described in section 3.2 and allowed
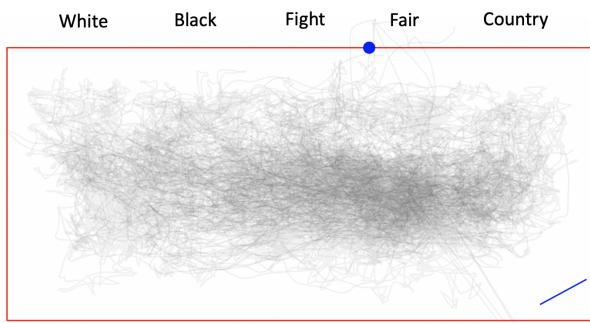
532

Figure 4: A plot of camera position data across all participants. Transparency is used so that darker regions represent more time spent in those regions. The red box indicates the tracked area, the blue line located in the bottom right corner indicates the starting position of the prompt, and the blue marker by "Fair" indicates the position of the experimenter. Participants traveled freely within the tracked space except for the areas near the boundary.

the participants time to explore the environment and practice the controls. This phase also typically took five to ten minutes.

The main phase of the study involved the participants performing the experimental task, as described in section 4.3. During this phase, an experimenter was in the room at all times to enable the Wizard-of-Oz features of label creation and note taking. This also served to reassure the participant, as VR is often isolating. This phase was designed to be split up into two sections of 40 minutes each rather than one long session. However, all of the participants opted to complete this phase in a single continuous session that lasted a mean of 51.7 minutes.

After indicating that they had completed the main phase, the participants would then start the in-VR interview phase. This involved asking a series of open-ended questions to better understand their sensemaking process while using IST. These included an outline of their answer for the two prompts, describing the clusters they formed, and asking them to describe any spatial relationships they used during their sensemaking process.

Finally, the post-experiment interview phase was designed to gather feedback about the general concept of IST and specific features of our implementation. We asked the participants about their user experience, what features they liked, and what could be improved. We also asked participants what their impression of the approach was, if they wouldd use IST in their work (and why), what features were useful or were distracting, and if they had any more comments on the system. This phase took between five and ten minutes.

4.6 Data Collection & Measures

We collected a variety of data in order to measure participants' actions, output, and feedback during the experiment.

We recorded video of the Unity screen (which included the participant's point of view) using Microsoft's Xbox Game Bar App. This allowed us to review what the participant was doing, as well as any console error messages and Wizard-of-Oz actions that were performed during the experiment. To gather insights from participants on what they were thinking while using IST and analyzing documents, we also recorded audio from the HTC VIVE Pro microphone. This also recorded the in-VR interview, which we later transcribed for further analysis.

We coded IST to output a log file of all user actions. These included camera and controller positional data and user actions as described in section 3.2. In particular, document, camera, and controller movement were recorded in the log file up to 10 times a

second to understand the user's full range of movement rather than simply their start and end locations.

Final document layouts were recorded using a save feature that recorded document positional data and all annotations like labels or notes. This allowed us to see the final clusters or structures of meaning that each participant formed during the experiment and see any similarities between participants. Furthermore, it allowed us to see what documents the participants interacted with, what documents they did not interact with, and what documents they deemed important.

The post-experiment interview was recorded using Apple's Voice Memos app that was running on an Apple iPhone. These interviews were also transcribed for further analysis.

5 RESULTS & DISCUSSION

To address RQ1, we analyzed how participants moved and interacted with documents in IST (section 5.1) and how they used the immersive space to organize documents during their analysis process (section 5.2). To address RQ2, we compared these findings with findings from a prior study on the IST concept [3], in particular analyzing how the initial document layout may have biased participants' selection of documents (section 5.3). For RQ3, we analyzed the analytic performance of our participants and attempted to correlate this performance with different IST usage strategies (section 5.4). Finally, for RQ4 we analyzed the results of our post-experiment interview (section 5.5).

5.1 Participant & Document Movement Patterns

We tracked movement data of both the participants and the documents they used through the log files generated by IST. Participants completed the main task and in-VR interview in a mean of 51.7 minutes (standard deviation of 11.4). However, there were only three complaints about fatigue in the post-experiment interview, and all three were related to eye strain. This was surprising, as users often get fatigued or affected by simulator sickness while experiencing a virtual environment for that long [24]. Furthermore, the participant movement map seen in figure 4 also shows that participants tended to stay generally in the center of the area, with a slight bias to the right side. This right-side bias may be due to the participants typical starting location, which was below the blue marker in figure 4. Participants still used a significant portion of tracked area over the course of the experiment. As users got more comfortable being in the virtual environment they spread out and used more of the available space.

One pattern of interaction that we observed was in how documents were processed. Participants would choose a document, read it, the process the document. Processing the document would involve offloading cognition onto the document through highlights or notes, and/or placing the document in an intermediate or final position. This includes documents that were placed back on the bulletin board or in a trash pile, when the participant deemed the document unnecessary to their outline. Often, participants would place these documents in their final overall position at this stage.

5.2 Clusters & Final Layouts

Participants interacted with a mean of 31.4 documents (standard deviation of 7.35) that they placed all around the virtual environment. Final placement for all of the documents that participants interacted with can be seen in figure 5. The final layouts for each participant represent how they understood the relationships among the documents. This is reflected in the structures of meaning the documents create, or how they have externalized their thought patterns with their notes, highlights, labels, and relative positions. When asked about their layouts, participants mentioned that they did group the documents meaningfully.
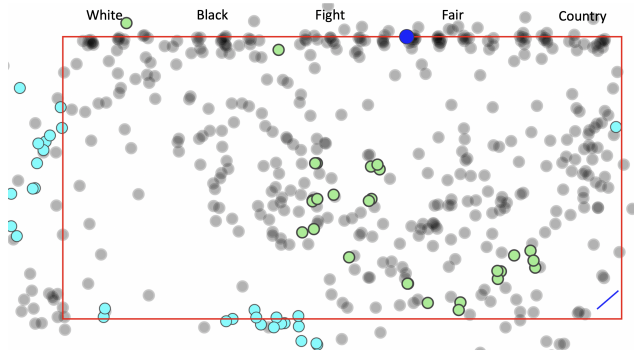
533

Figure 5: A scatterplot of all participants' final document layouts, removing documents that the participants did not interact with during the study. The tracked area had a roughly even distribution of the documents, with hotspots on the bulletin board (the dark border at the top of the plot) and the edges of the virtual floor (represented by the red rectangular outline). The environmental layout of participant 3 (seen in figure 6) is marked with cyan points, while the semicircular layout of participant 4 (seen in figure 7) is marked with green points.

While our participants had many different organizational structures for their documents, we observed three common types of overall layouts; we have termed these *semicircular*, *environmental*, and *planar*.

Semicircular layouts, as the name suggests, were arrangements of documents in a rough semicircle. All of the documents faced towards the same general center point, where the user would stand. This afforded the ability to see all documents at once or by turning the head slightly. Labels could also be dispersed within to denote themes and sub-clusters. Participant 4's semicircular structure can be seen in figures 1 and 7. This participant stacked documents vertically, and while some documents obscured others, all the documents could be seen with a little movement. As participant 22 phrased it, their layout was "as if I'm on a stage and [the documents] are the crowd, they're all circular around me as if we're in a theater and I'm on the stage." This metaphor speaks to how all the documents are visible from the user's point of view at all times. Seven participants utilized this kind of arrangement.

Environmental layouts used structures and cues from the virtual environment when placing documents and clusters. In our implementation, these cues include the bulletin board, the separators on the bulletin board, and the virtual floor. Participant 3, for example, used the left and back edges of the floor (when facing the bulletin board) to arrange their clusters, as seen in figure 6. Seven participants in all used these environmental cues as ways of organizing documents. This, combined with the feedback given in the post-experiment interview, suggests that users wanted some sort of existing structure to build upon. For example, participant 16 wanted the ability to create their own bulletin board structure in order to arrange documents. They thought it would be a way to organize documents around common themes and then move the entire bulletin board when constructing the outline to their essay. Several other participants shared this desire; further, they wanted the floor and existing bulletin boards to be impenetrable so that documents couldn't pass through them. This would be analogous to a cork board that some analysts use during sensemaking with paper artifacts.

Planar layouts involve multiple clusters of documents where each cluster forms a linear wall-like structure. Participant 15 had a planar layout where they placed documents roughly along five planes that were perpendicular to the bulletin board. This participant seemed to be pulling out documents from each category and organizing them in their own sections to find themes within the category. This

participant also specifically mentioned that they stacked documents vertically if they shared a theme. Participant 11 had a similar arrangement, except they had their planes in various positions and orientations in space that didn't seem to align with any environmental elements. Neither participant 11 nor 15's planes shared a common viewpoint with any other plane.

Participant 21 was the only one that had no discernible organization, and even commented that they were just keeping the documents close to where they were originally placed. However, during the in-VR interview, they noted that they wished they had created clusters away from the bulletin board, and noted what clusters they would have formed if they could do it over again.

In previous work [3], Bandyopadhyay et al. observed similar layouts compared to those seen in this study. Arrangements that were like the semicircular arrangement were seen, though the radii of those layouts were often much smaller. Their study also observed "wall" layouts, which are similar to the planar layouts but consisted of a single plane. In Batch et al.'s work, they found similar semicircular and linear patterns as well [4]. Similarly, Satriadi et al. found egocentric spherical and planar layouts in their work on coordinating multiple map views in VR [27]. However, none of these immersive analytics studies saw layouts that conformed to environmental cues, with Satriadi et al.'s study purposely leaving out environmental cues as part of their study design.

One common aspect of these layouts is that they all use local 2D surfaces in a global 3D space. Even the semicircular arrangement is essentially a 2D cylindrical plane wrapped around the user. This conflicted with our expectation that users would use 3D depth in more meaningful ways. Still, the existence of the 3D space in IST was useful, not for encoding more dimensions of meaning of the documents, but for spreading out the documents, viewing multiple clusters of documents more easily, and for expressing relationships between clusters of documents. Moreover, this finding suggests that we can enhance IST with tools that support the creation and manipulation of 2D surfaces on which documents can be attached and arranged. At the same time, we plan to look at long-term usage of IST in the future, as strategies may change over time to adapt to the 3D environment.

Participants used several methods of organization within clusters. Labels were used as we anticipated in eleven of the layouts, where they denoted cluster or subcluster themes and headings that participants could group documents around. Several participants, however, used a single document's note section to summarize a theme instead of using a label. This was due to the character-count constraints we enforced on labels to support the search capability; these users wanted more than a phrase to describe a theme.

5.3 Document Selection Bias

Prior work on Space to Think presented users with a list of files for the user to browse and select from, affording choice of documents, but only based on the document title. In prior research on the Immersive Space to Think concept [3], documents were presented to the user in a stack. This stack was similar to grabbing a stack of papers off a printer, where only the top document was visible at any one time. Thus, users almost always examined documents in the same order. In our version of IST, we made all documents visible to the user on a virtual bulletin board, indicated a key piece of information about each document (whether it contained the response of a white or black soldier), and pre-sorted the documents into categories. We hoped that this would help users find the most relevant documents to analyze and allow freedom of choice. Due to the changes we made in the way we presented the documents to our participants, we were interested to see if there was bias in the way participants selected documents from the bulletin board.

Figure 8 represents the bulletin board with each document represented as a point. The color of the point indicates how many times
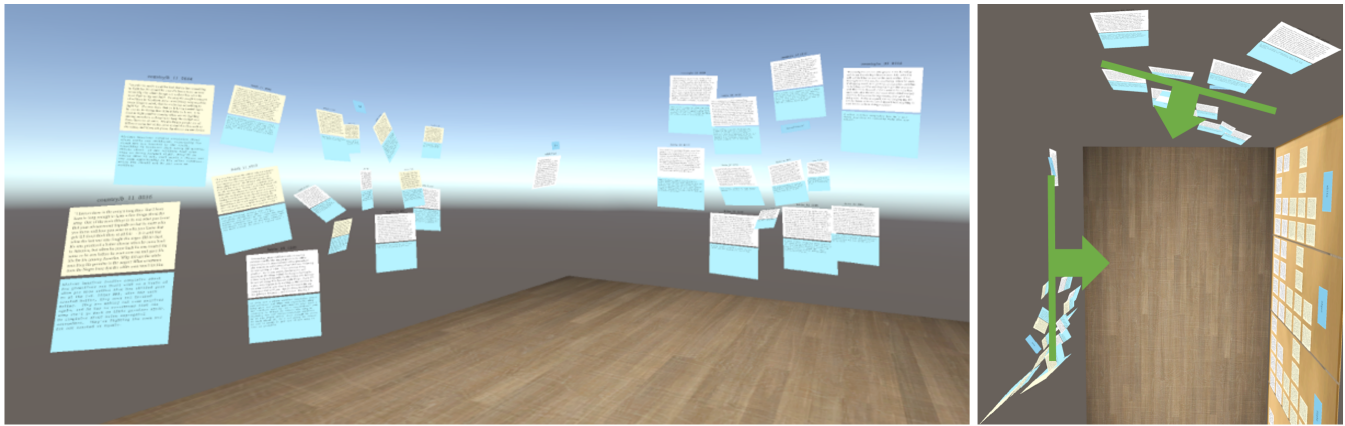
Figure 6: Participant 3's final layout of documents, with a first-person view on the left and a top-down view on the right. The top-down view is annotated with arrows to indicate the direction the cluster is facing. This participant used an *environmental layout*, where the documents conform to an environmental cue. In this case, they are aligned to the virtual environment's floor.

Figure 7: Participant 4's final layout is shown here in a top-down view. An arc overlay was added to highlight that the participant used a *semicircular layout*, where all the documents are visible from a center point where the user stands. This affords the ability to see all the documents at once or by slightly turning their head.

Figure 8: Frequency of document selection by all participants. Each document is represented as a point relative to its location on the virtual bulletin board. Darker colors represent more frequent selection.

participants interacted with that document and removed it from the bulletin board. The darker the mark, the more times the documents were interacted with by the participants. This helps us understand the way that participants selected documents from the bulletin board, and whether certain documents were chosen more often than others.

Looking at the top row of documents in Figure 8, we can see that the participants tended to interact with the top left documents within each category. Additionally, when the board is broken down into the two response types, we can see a trend that the most referenced documents of each response type (Black or White) tend to be the top/left documents for each sub-category. In cognitive psychology, order effects are often seen in selection tasks, where users show a bias towards the order of the items presented for selection [30]. We

believe that these results show a strong order effect in the document selection task that is reflective of the way English text is read (top to bottom, left to right).

We found a similar order bias when we examined the number of times each document was one of the first three documents selected by a participant. Within the categories White, Fight, Fair and Country, the most selected document was the top-left in each respective category. Fewer documents overall among the first three selected were from the Black category. We believe that this could be due to user discomfort with the displayed category name (see section 4.3). Together, these results show a strong order effect on document selection by our participants.

5.4 Historical Analysis Performance

As we stated for RQ3, we wanted to understand whether there were any particularly effective strategies for analyzing documents in IST. We recruited three experts in historical analysis to review our participants' verbal outlines and document layouts, and with their help we designed six ratings to evaluate how well each participant performed historical analysis. These were:

- *R1:* How evident is it that they have read individual documents?

- *R2:* How evident is it that they have read groups of documents?

535

- *R3:* Do they have a clear organizational scheme?

- *R4:* Have they identified ambiguities and contradictions?

- *R5:* Do they show nuance and understanding of the topic?

- *R6:* What is the quality of their answer to the prompts?

Each question was rated on a ten-point scale. We performed an intraclass correlation (ICC) test to see if the expert ratings were consistent. Using a one-way mixed effects model, our ICC value for the sum of the ratings was $ICC(C, 1) = 0.666$, which indicates good agreement between graders. We also checked the individual ratings, and got $ICC(C, 1) = 0.396, 0.364, 0.596, 0.453, 0.691$, and 0.638 for R1, R2, R3, R4, R5, and R6 respectively. This means there was good agreement between graders for R3, R5, and R6, but poor agreement for R1, R2, and R4.

We further performed correlation tests on many different variables to see how they might be related to the ratings. These included layout type, number of documents interacted with, time spent in IST, participant travel distance, total document travel distance, label usage, existence of trash clusters, propensity to use notes or highlights (including neither or both), number of documents participants stacked vertically, and if they emphasized the outline structure in their notes, labels, or clusters. For correlations between categorical variables and scores, we used ANOVA tests, while correlations between continuous variables and scores used Pearson tests [8].

We found two weak correlations: one between R3 and label use $(F(1, 15) = 3.732, p = 0.0725)$; and one between R5 and whether the participant used notes and/or highlights $(F(3, 13) = 2.607, p = 0.961)$. The first correlation indicates that participants had better ratings for organization if they used labels. The second correlation indicates that participants showed more nuance and understanding of the topic if they used both highlights and notes. Both of these correlations indicate that the tools we added to IST can help lead to more effective sensemaking; while 3D space can be useful for organizing an analyst's thinking, this should be combined with additional tools for externalizing the thought process. In addition, IST can be further improved with semantic interaction that intelligently assists users. For example, Wenskovitch et al. designed a feature where user-defined highlights would allow the system to suggest documents of potential interest. They found that participants who used this feature performed significantly better than those that did not [32].

5.5 Post-Experiment Feedback

The post-experiment interview asked the participants how they felt about the IST approach or our implementation. Fifteen of the seventeen participants would like to use IST to write future papers or essays, though three of those answers were conditional upon more practice. These participants particularly pointed out that the virtual environment removed distractions, allowing them to focus on the analysis task. Participant 1 said, "I felt like I had my own space there, to really work ... I was able to concentrate." The remaining two participants didn't like the controllers or preferred physical interaction with real documents.

Participants had varied opinions on what features they preferred. Seven participants mentioned liking the ability to enlarge documents. This could explain why we found relatively little fatigue or eye strain. Six participants discussed how having notes was a key part of their process. Participant 3 stated that they liked how notes "directly attached to the documents. That way, it would be harder to lose track of what notes were with what." This points to the importance of being able to offload cognition and memory onto the environment through note-taking.

Three participants compared using IST to having a browser window with 100 different tabs open. Participant 21 said that "it was

definitely better than sitting on your laptop with 100 PDFs open ... just trying to sift through it ... once you [close one tab], you've gotta go back and figure out which PDF it was to retrace it down." All three stated that it was much easier to view multiple documents simultaneously and compare/contrast them, as compared to a traditional desktop or laptop display.

6 LIMITATIONS

While we were able to see that participants were able to make sense of a large document set with IST, there were several limitations to our study.

The Wizard-of-Oz features likely influenced some amount of the results. To create notes, for example, usually involved a back and forth with the experimenter to ensure that they got across the correct words and themes the participant wanted. This could have been interpreted as some amount of agreement with the participant, even if the experimenter did not mean to.

Another limitation was with our tracked area. Though we did calibrate the tracking, the edges of the area tended to lose tracking for the controllers occasionally. This caused some frustration in the participants, and it was specifically called out as a frustration in the post-experiment interview by five participants.

A key limitation of this study is that we recruited novice users. While novices can reveal some flaws in designs quickly, they likely have different methods and results as compared to expert analysts. We expect to study expert usage of IST in the future.

7 CONCLUSIONS & FUTURE WORK

Our research revealed several different ways that users can utilize IST to perform sensemaking and better understand large datasets. We observed three main spatial organizational structures that can help people find themes and stories within large document sets. These structures suggest that future versions of IST could be designed with explicit support for the creation, manipulation, and use of 2D surfaces for organizing clusters of documents. Beyond this, we have several plans for improving IST, both as an approach and in its implementation.

First, we want to add a method for directly entering text into IST. This will allow participants to offload cognition onto the environment without losing track of their thought process. Second, we hope to integrate semantic interaction into IST. Semantic interaction refers to observation of the analyst's interaction with the data and intelligent assistance based on these observations. This approach has been studied extensively by visual analytics researchers, even in previous iterations of the Space to Think [12, 16]. Third, we plan to create an augmented reality variant of IST. This would allow users to use non-digital tools such as whiteboards or paper documents, and digital tools such as word-processing software and search engines, to enhance the sensemaking process. Fourth, we propose that various strategies could be implemented into future iterations of IST to help mitigate bias. One simple strategy that does not require significant changes would be randomizing documents on the wall. Another strategy to mitigate document selection bias would be to provide alternative techniques for selection, such as a search-based document finder.

REFERENCES

[1] D. Ancona. Framing and acting in the unknown. *S. Snook, N. Nohria, & R. Khurana, The Handbook for Teaching Leadership*, pp. 3–19, 2012.

[2] C. Andrews, A. Endert, and C. North. Space to think: large high-resolution displays for sensemaking. In *Proceedings of the SIGCHI conference on human factors in computing systems*, pp. 55–64. ACM, 2010.

[3] P. Bandyopadhyay, L. Lisle, C. North, and D. A. Bowman. Immersive space to think: The role of 3d space for sensemaking. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*. ACM, 2020.

[4] A. Batch, A. Cunningham, M. Cordeil, N. Elmqvist, T. Dwyer, B. H. Thomas, and K. Marriott. There is no spoon: Evaluating performance, space use, and presence with expert domain users in immersive analytics. *IEEE transactions on visualization and computer graphics*, 26(1):536–546, 2019.

[5] S. Benford, D. Snowdon, C. Greenhalgh, R. Ingram, I. Knox, and C. Brown. Vr-vibe: A virtual environment for co-operative information retrieval. In *Computer Graphics Forum*, vol. 14, pp. 349–360. Wiley Online Library, 1995.

[6] J. S. Brown, A. Collins, and P. Duguid. Situated cognition and the culture of learning. *Educational researcher*, 18(1):32–42, 1989.

[7] T. Chandler, M. Cordeil, T. Czauderna, T. Dwyer, J. Glowacki, C. Goncu, M. Klapperstueck, K. Klein, K. Marriott, F. Schreiber, et al. Immersive analytics. In *2015 Big Data Visual Analytics (BDVA)*, pp. 1–8. IEEE, 2015.

[8] P. Y. Chen, M. Smithson, and P. M. Popovich. *Correlation: Parametric and nonparametric measures*. Number 139. Sage, 2002.

[9] D. Clergeaud and P. Guitton. Design of an annotation system for taking notes in virtual reality. In *2017 3DTV Conference: The True Vision-Capture, Transmission and Display of 3D Video (3DTV-CON)*, pp. 1–4. IEEE, 2017.

[10] M. Cordeil, A. Cunningham, T. Dwyer, B. H. Thomas, and K. Marriott. Imaxes: Immersive axes as embodied affordances for interactive multivariate data visualisation. In *Proceedings of the 30th Annual ACM Symposium on User Interface Software and Technology*, pp. 71–83. ACM, 2017.

[11] W. Elm, S. Potter, J. Tittle, D. Woods, J. Grossman, and E. Patterson. Finding decision support requirements for effective intelligence analysis tools. In *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, vol. 49, pp. 297–301. SAGE Publications Sage CA: Los Angeles, CA, 2005.

[12] A. Endert, P. Fiaux, and C. North. Semantic interaction for sensemaking: inferring analytical reasoning for model steering. *IEEE Transactions on Visualization and Computer Graphics*, 18(12):2879–2888, 2012.

[13] A. Endert, W. Ribarsky, C. Turkay, B. W. Wong, I. Nabney, I. D. Blanco, and F. Rossi. The state of the art in integrating machine learning into visual analytics. In *Computer Graphics Forum*, vol. 36, pp. 458–486. Wiley Online Library, 2017.

[14] E. Gitre and K. Luther. The american soldier collaborative digital archive. 2018.

[15] T. J. Guerreiro, D. Medeiros, D. Mendes, M. Sousa, J. A. Jorge, A. Raposo, and I. H. dos Santos. Beyond post-it: Structured multimedia annotations for collaborative ves. In *ICAT-EGVE*, pp. 55–62, 2014.

[16] E. Horvitz. Principles of mixed-initiative user interfaces. In *Proceedings of the SIGCHI conference on Human Factors in Computing Systems*, pp. 159–166, 1999.

[17] Y.-a. Kang, C. Gorg, and J. Stasko. How can visual analytics assist investigative analysis? design implications from an evaluation. *IEEE Transactions on Visualization and Computer Graphics*, 17(5):570–583, 2010.

[18] G. Klein, B. Moon, and R. R. Hoffman. Making sense of sensemaking 2: A macrocognitive model. *IEEE Intelligent systems*, 21(5):88–92, 2006.

[19] W. S. Lages and D. A. Bowman. Move the object or move myself? walking vs. manipulation for the examination of 3d scientific data. *Frontiers in ICT*, 5:15, 2018.

[20] L. Lisle, X. Chen, E. J. K. Gitre, C. North, and D. A. Bowman. Evalu-

[21] K. Marriott, F. Schreiber, T. Dwyer, K. Klein, N. H. Riche, T. Itoh, W. Stuerzlinger, and B. H. Thomas. *Immersive Analytics*, vol. 11190. Springer, 2018.

[22] H. T. Nim, B. Sommer, K. Klein, A. Flack, K. Safi, M. Nagy, W. Fiedler, M. Wikelski, and F. Schreiber. Design considerations for immersive analytics of bird movements obtained by miniaturised gps sensors. In *Proceedings of the Eurographics Workshop on Visual Computing for Biology and Medicine*, pp. 27–31. Eurographics Association, 2017.

[23] P. Pirolli and S. Card. The sensemaking process and leverage points for analyst technology as identified through cognitive task analysis. In *Proceedings of international conference on intelligence analysis*, vol. 5, pp. 2–4. McLean, VA, USA, 2005.

[24] T. M. Porcino, E. Clua, D. Trevisan, C. N. Vasconcelos, and L. Valente. Minimizing cyber sickness in head mounted display systems: design guidelines and applications. In *2017 IEEE 5th international conference on serious games and applications for health (SeGAH)*, pp. 1–6. IEEE, 2017.

[25] R. Rädle, H.-C. Jetter, S. Butscher, and H. Reiterer. The effect of egocentric body movements on users' navigation performance and spatial memory in zoomable user interfaces. In *Proceedings of the 2013 ACM international conference on Interactive tabletops and surfaces*, pp. 23–32, 2013.

[26] D. Sacha, A. Stoffel, F. Stoffel, B. C. Kwon, G. Ellis, and D. A. Keim. Knowledge generation model for visual analytics. *IEEE transactions on visualization and computer graphics*, 20(12):1604–1613, 2014.

[27] K. A. Satriadi, B. Ens, M. Cordeil, T. Czauderna, and B. Jenny. Maps around me: 3d multiview layouts in immersive spaces. *Proceedings of the ACM on Human-Computer Interaction*, 4(ISS):1–20, 2020.

[28] R. Skarbez, N. F. Polys, J. T. Ogle, C. North, and D. A. Bowman. Immersive analytics: Theory and research agenda. *Frontiers in Robotics and AI*, 6:82, 2019.

[29] J. Stasko, C. Görg, and Z. Liu. Jigsaw: supporting investigative analysis through interactive visualization. *Information visualization*, 7(2):118–132, 2008.

[30] F. Strack. *"Order Effects" in Survey Research: Activation and Information Functions of Preceding Questions*, pp. 23–34. Springer New York, New York, NY, 1992. doi: 10.1007/978-1-4612-2848-6_3

[31] K. E. Weick, K. M. Sutcliffe, and D. Obstfeld. Organizing and the process of sensemaking. *Organization science*, 16(4):409–421, 2005.

[32] J. Wenskovitch, L. Bradel, M. Dowling, L. House, and C. North. The effect of semantic interaction on foraging in text analysis. In *2018 IEEE Conference on Visual Analytics Science and Technology (VAST)*, pp. 13–24. IEEE, 2018.

[33] C. D. Wickens, J. G. Hollands, S. Banbury, and R. Parasuraman. *Engineering psychology and human performance*. Psychology Press, 2015.

[34] M. Wilson. Six views of embodied cognition. *Psychonomic bulletin & review*, 9(4):625–636, 2002.

[35] W. Zhu, P. Cui, Z. Wang, and G. Hua. Multimedia big data computing. *IEEE multimedia*, 22(3):96–c3, 2015.

ating the benefits of the immersive space to think. In *2020 IEEE 6th Workshop on Everyday Virtual Reality (WEVR)*. IEEE, 2020.