Immersive Space to Think: Immersive Analytics for Sensemaking with Non-Quantitative Datasets

Lorance R. Lisle

Dissertation submitted to the Faculty of the Virginia Polytechnic Institute and State University in partial fulfillment of the requirements for the degree of

Doctor of Philosophy in
Computer Science & Applications

Doug A. Bowman, Chair
Chris North
Joseph L. Gabbard
Edward J.K. Gitre
Nicholas F. Polys
Tobias Höllerer

December 8, 2022
Blacksburg, Virginia

Keywords: Sensemaking, Virtual Reality, Augmented Reality, Immersive Analytics
Copyright 2023, Lorance R. Lisle
Immersive Space to Think: Immersive Analytics for Sensemaking with Non-Quantitative Datasets

Lorance R. Lisle

(ABSTRACT)

Analysts often work with large complex non-quantitative datasets in order to better understand concepts, themes, and other forms of insight contained within them. As defined by Pirolli and Card, this act of sensemaking is cognitively difficult, and is performed iteratively and repetitively through various stages of understanding. Immersive analytics has purported to assist with this process through putting users in virtual environments that allows them to sift through and explore data in three-dimensional interactive settings. Most previous research, however, has focused on quantitative data, where users are interacting with mostly numerical representations of data. We designed Immersive Space to Think, an immersive analytics approach to assist users perform the act of sensemaking with non-quantitative datasets, affording analysts the ability to manipulate data artifacts, annotate them, search through them, and present their findings. We performed several studies to understand and refine our approach and how it affects users sensemaking strategies. An exploratory virtual reality study found that users place documents in 2.5-dimensional structures, where we saw semicircular, environmental, and planar layouts. The environmental layout, in particular, used features of the environment as scaffolding for users’ sensemaking process. In a study comparing levels of mixed reality as defined by Milgram-Kishino’s Reality-Virtuality Continuum, we found that an augmented virtuality solution best fits users’ preferences while still supporting external tools. Lastly, we explored how users deal with varying amounts of space and three-dimensional user interaction techniques in a comparative study comparing small virtual monitors, large virtual monitors, and a seated-version implementation of Immersive Space to Think. Our participants found IST best supported the task of sensemaking, with evidence that users leveraged spatial memory and utilized depth to denote additional meaning in the immersive condition. Overall, Immersive Space to Think affords an effective sensemaking three-dimensional space using 3D user interaction techniques that can leverage embodied cognition and spatial memory which aids the users understanding.
Humans are constantly trying to make sense of the world around them. Whether they’re a detective trying to understand what happened at a crime scene or a shopper trying to find the best office chair, people are consuming vast quantities of data to assist them with their choices. This process can be difficult, and people are often returning to various pieces of data repeatedly to remember why they are making the choice they decided upon. With the advent of cheap virtual reality products, researchers have pursued the technology as a way for people to better understand large sets of data. However, most mixed reality applications looking into this problem focus on numerical data, whereas a lot of the data people process is multimedia or text-based in nature. We designed and developed a mixed reality approach for analyzing this type of data called Immersive Space to Think. Our approach allows users to look at and move various documents around in a virtual environment, take notes or highlight those documents, search those documents, and create reports that summarize what they’ve learned. We also performed several studies to investigate and evolve our design. First, we ran a study in virtual reality to understand how users interact with documents using Immersive Space to Think. We found users arranging documents around themselves in a semicircular or flat plane pattern, or using various cues in the virtual environment as a way to organize the document set. Furthermore, we performed a study to understand user preferences with augmented and virtual reality. We found a mix of the two, also known as augmented virtuality, would best support user preferences and ability. Lastly, we ran two comparative studies to understand how three dimensional space and interaction affects user strategies. We ran a small user study looking at how a single student uses a desktop computer with a single display as well as immersive space to think to write essays. We found that they wrote essays with a better understanding of the source data with Immersive Space to Think than the desktop setup. We conducted a larger study where we compared a small virtual monitor simulating a traditional desktop screen, a large virtual monitor simulating a monitor 8 times the size of traditional desktop monitors, and immersive space to think. We found participants engaged with documents more in Immersive Space to Think, and used the space to denote importance for documents. Overall, Immersive Space to Think provides a compelling environment that assists users in understanding sets of documents.
Dedication

To my wife, Dr. Alice Haskins Lisle, who supported me throughout this process, and to our three daughters, Catherine, Birdie, and Evelyn. I love you.
Acknowledgments

Personal Acknowledgments

The completion of my dissertation would not have been possible without the advice, strength, and love of my wife Dr. Alice Haskins Lisle, to whom I attribute my entire Ph.D. She is a constant inspiration in all my work. I would not have made it without her.

To my advisor, Dr. Doug A. Bowman, who profoundly impacted my life in his patience and understanding. He gave me time and space to raise a family and navigate a global pandemic all while I was trying to run participants and get my dissertation completed. He taught me how to reframe arguments and how to think about experimental design and not to give up just because someone thought of something first.

To Dr. Joseph L. Gabbard, who pushed me to apply for graduate school and advised me in so many ways during my academic career. I am proud that he is on my committee.

To Dr. Chris North, who was at nearly every progress meeting and would always give good advice on what to look for in the data.

To the rest of my committee, Dr. Edward J.K. Gitre, Dr. Nicholas Polys, and Dr. Tobias Höllerer, thank you all for your gracious feedback throughout the process.

To all my colleagues in the 3DI group: Dr. Wallace Lages, Dr. Run Yu, Dr. Yuan Li, Dr. Lei Zhang, Shakiba Davari, Feiyu Lu, Leonardo Pavanatto, Kylie Davidson, Cory Ike Ilo, Ibrahim Tahmid, Alexander Giovannelli, Logan Lane, and Dr. Jerald Thomas. Thank you all for the friendship and good times. I’m glad we went out of our way to socialize and help one another. uwu

Lastly, I’d like to thank the Office of Naval Research for the generous funding that supported me for several years.

Attribution

Chapter 5 and section 3.4 were written in collaboration with Kylie Davidson, Dr. Edward J.K. Gitre, Dr. Chris North, and Dr. Doug A. Bowman. Kylie Davidson is a Ph.D. student at Virginia Tech and helped with the analysis of the data, as well as co-developing many of the visualizations. Dr. Edward J.K. Gitre is a Professor of History at Virginia Tech, and provided the dataset and helped guide the user study. Dr. Chris North is a Professor of
Computer Science at Virginia Tech, and has guided the design of IST, the user studies, and the analysis of the data. My advisor, Dr. Doug A. Bowman, has assisted with all aspects of the experimental process for both studies.

Section 5.1 was performed in collaboration with Xiaoyu Chen, Dr. Edward J.K. Gitre, Dr. Chris North, and Dr. Doug A. Bowman. Xiaoyu Chen was a graduate student at Virginia Tech and assisted with the software for the desktop condition. Dr. Edward J.K. Gitre, Dr. Chris North, and Dr. Doug A. Bowman assisted with all phases of the experimental process.

Section 5.2 was performed in collaboration with Kylie Davidson, Leonardo Pavanatto, Dr. Chris North, and Dr. Doug A. Bowman. Kylie Davidson again helped with the analysis and co-developed several visualizations. Leonardo Pavanatto is a Ph.D. student at Virginia Tech and provided the software to create the virtual monitors used in the study. Dr. Chris North and Dr. Doug A. Bowman assisted with all phases of the experimental process.
# Contents

## List of Figures

xi

## List of Tables

xvi

## 1 Introduction

1.1 Background & Motivation ........................................ 1
1.2 Approach ............................................................. 5
1.3 Research Overview .................................................. 6
   1.3.1 Organization .................................................... 7

## 2 Related Work

2.1 Theoretical Foundations ........................................... 9
   2.1.1 Embodied & Situated Cognition ................................ 9
   2.1.2 Sensemaking .................................................... 13
2.2 Sensemaking in Various Domains ................................ 16
   2.2.1 Scientific Analysis ............................................. 16
   2.2.2 Intelligence Analysis ......................................... 17
   2.2.3 Historical Analysis ........................................... 18
2.3 Visual Analytics ................................................... 19
   2.3.1 Concept .......................................................... 19
   2.3.2 Tools ............................................................. 21
2.4 Immersive Analytics .............................................. 22
   2.4.1 Display Technology ............................................ 22
   2.4.2 Tools ............................................................. 25
   2.4.3 Future Directions ............................................... 26
3 Design of the Immersive Space to Think

3.1 Hardware and User Experience Specifications ........................................... 28
3.2 User Needs ........................................................................................................ 29
3.3 Interaction Methods ......................................................................................... 30
   3.3.1 Single-hand Movement of Documents ...................................................... 30
   3.3.2 Multi-hand Resizing and Moving of Documents ...................................... 30
   3.3.3 Text scrolling ............................................................................................ 31
   3.3.4 Text Highlighting ....................................................................................... 32
   3.3.5 Text Copying .............................................................................................. 32
   3.3.6 Document Copying ..................................................................................... 32
   3.3.7 Keyboard Input .......................................................................................... 33
   3.3.8 Menu Interactions ...................................................................................... 34
   3.3.9 Note Taking ............................................................................................... 34
   3.3.10 Label Creation ......................................................................................... 34
   3.3.11 Keyword Search ....................................................................................... 35
   3.3.12 Save & Load ............................................................................................ 35
   3.3.13 PDF Viewing ........................................................................................... 36
   3.3.14 Video Playback ....................................................................................... 36
3.4 Different Realities: A Comparison of Augmented and Virtual Reality for Sensemaking Tasks ......................................................... 37
   3.4.1 Research Questions .................................................................................... 38
   3.4.2 Experiment Design ................................................................................... 39
   3.4.3 Results & Discussion ................................................................................ 43
   3.4.4 Limitations ............................................................................................... 51
   3.4.5 Conclusions .............................................................................................. 52
3.5 Chapter Summary ............................................................................................ 52

4 The Effect of 3D Immersive Space on Sensemaking Strategies .................. 54

4.1 Sensemaking Strategies with Immersive Space to Think .............................. 54
4.1.1 Experimental Design ............................................. 56
4.1.2 Results & Discussion ........................................... 62
4.1.3 Limitations ......................................................... 71
4.2 Chapter Summary .................................................. 72

5 Comparison of IST to Other Sensemaking Tools ............... 73
5.1 Initial Comparative Study ......................................... 74
  5.1.1 Methods ......................................................... 75
  5.1.2 Results & Discussion ......................................... 78
  5.1.3 Limitations ....................................................... 81
5.2 Comparison of Small, Large, and Immersive Displays ........ 82
  5.2.1 Condition Design .............................................. 84
  5.2.2 Experimental Design ......................................... 87
  5.2.3 Results & Discussion ......................................... 93
  5.2.4 Limitations ....................................................... 105
  5.2.5 Conclusions ..................................................... 106
5.3 Chapter Summary .................................................. 107

6 Conclusions & Future Work ....................................... 108
6.1 Conclusions ........................................................ 108
  6.1.1 RQ1: How can we design analytic tools for 3D immersive space for sensemaking with large non-quantitative datasets? 108
  6.1.2 RQ2: How does 3D immersive space affect users performing sense-making tasks with non-quantitative datasets? ....... 108
  6.1.3 RQ3: What are the relative benefits for 3D Immersive space as compared to large scale, high-resolution displays and traditional methods of performing sensemaking tasks? ....................... 109
6.2 Future Work ........................................................ 110
  6.2.1 Scalability of Immersive Space to Think .................... 110
  6.2.2 AI-powered Assistance for Sensemaking in IST ............ 110
<table>
<thead>
<tr>
<th>Section</th>
<th>Title</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>6.2.3</td>
<td>Effect of Real-World Environment</td>
<td>111</td>
</tr>
<tr>
<td>6.2.4</td>
<td>Best Practices for the use of Immersive Space to Think</td>
<td>111</td>
</tr>
</tbody>
</table>

Bibliography

Appendix A Sensemaking Strategies with Immersive Space to Think Appendix

A.1 Informed Consent Form | 124
A.2 Questionnaires & Interviews | 128
A.2.1 Background Questionnaire | 128
A.2.2 Post-session Semi-structured Interview | 130
A.3 Post-Study Semi-structured Interview | 130
A.4 IRB Approval Letter | 131

Appendix B Different Realities: A Comparison of Augmented and Virtual Reality for Sensemaking Tasks Appendix

B.1 User Consent Form | 134
B.1.1 Questionnaires | 138
B.1.2 Post-Session Questionnaires | 140
B.1.3 Post-Session Semi-structured Interview | 143
B.2 IRB Approval Letter | 143

Appendix C Comparison of Small, Large, and Immersive Displays

C.1 User Consent Form | 146
C.2 Questionnaires & Interviews | 150
C.2.1 Background Questionnaire | 150
C.2.2 Post-session Semi-structured Interview | 155
C.2.3 Post-Study Semi-structured Interview | 156
C.3 IRB Approval Letter | 156
List of Figures

1.1 The sensemaking process as described by Pirolli & Card. It involves an iterative pipeline of stages that can be looped between each other [94]. 2

1.2 An example of a fictional detective’s workspace shows how they are organizing data to perform sensemaking from HBO’s True Detective [95]. 3

1.3 The initial state of IST when the user first enters the virtual environment is a floor, bulletin board, and multimedia documents. The floor represents the tracked area the user could traverse, and all the documents are initially displayed in predefined categories on the bulletin board. 5

2.1 Klein and Moon’s Data/Frame sensemaking process. It involves a frame of understanding that is adjusted or reframed as additional data is considered by the analyst [63], [102]. 14

2.2 Sacha et al.’s Knowledge Generation Model that incorporates a computer counterpart that the analyst works with to better understand a dataset using visual analytics [102]. 15

3.1 An example of a text document in IST. The UI panel on the left features document-specific interactions that can manipulate the document. These are highlight text selection, copy text selection, copy text selection with a citation, and add a note to the artifact. 31

3.2 The keyboard that participants used to enter text in IST was on a tracked wheeled desk, as seen in the image on the left. Once integrated into IST, our initial design was to represent the keyboard with a desk with a white keyboard in the same position as the physical keyboard, as seen in the center image. Feedback from pilot studies evolved our design, leading to a pass-through “Desk Portal” design as seen in the image on the right. Furthermore, the IST menu can be seen above the portal. These buttons are general purpose buttons that assist the user with the IST environment. In this implementation, the user can save their layout, create a new note, create a new label, or search the artifact dataset. 33

3.3 An example of a label in IST. It contains a key word or phrase and can denote meaning to a group of documents or an area in the virtual environment. 35
3.4 IST can handle many different file types, including images, text files, *pdfs* and videos (seen on the bottom row). Additionally, these are the resources the participant provided for phase three. .......................... 36

3.5 On the left is the physical space that users work in. When performing tasks in VR, they see the virtual environment to the right. The environment has a floor-with-railings setup so users know where they can travel safely when considering physical obstacles and the tether to the XR-3. .......................... 39

3.6 The left boxplot shows R1 grade performance (How evident is it that the participant has read individual documents?) for AR and VR, and the right boxplot shows R2 grade performance (How evident is it that they have read groups of documents?) for AR and VR. .......................... 45

3.7 The significant UEQ results from experiment II showed that participants found AR more physically safe (Secure vs. Insecure), Dependable, and Attractive than VR. .......................... 46

3.8 Left: 3D scatterplot of where B8 organized documents in AR. The red line emphasizes the semicircular layout, with the observation point being at the center of the arc. Right: B8’s layout in VR. The blue rectangle is the floor; documents are organized around three sides of the floor in an environmental layout. .......................... 48

3.9 Participant B11 wrote down quotes and thoughts for their analysis on the whiteboard, separating thoughts from white and black soldiers by using different color markers. .......................... 49

3.10 User movement and document layout for B15’s VR (on the left) and AR (on the right) sessions. The original plane that documents appeared on is shown in red, while the whiteboard is shown as a green line. Movement is visualized with the blue line. .......................... 50

3.11 Participant B15 utilized a strategy for analysis called a T-Chart. Using the whiteboard in the AR condition, they wrote about the varying experiences of white and black soldiers, filling the board with quotes, themes, and ideas, as seen on the left. They attempted to recreate a T-Chart in the VR condition, using the note feature, as seen on the right. .......................... 51

4.1 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. 57
4.2 A line graph of camera position data across all participants plotted with transparency. 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" indicated the position of the experimenter. The green diamond indicates the rough starting position for all participants. The darker regions indicate more movement in that space. Participants traveled freely within the tracked space except for the areas near the boundary. The area near the starting position and the prompt was used most often.

4.3 Example of a cluster of annotated documents. The documents have highlights, nearby labels, and two of them have short notes that were dictated by the user.

4.4 A heatmap 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 bottom of the heatmap) and the edges of the virtual floor (represented by the rectangular outline). Both the environmental and semicircular layouts can be seen in this aggregate data, with a rough semicircle around (0, 0) and higher numbers of documents near the edges.

4.5 Participant 3’s final layout of documents, with a first-person view on the left and a birds-eye view on the right. The birds-eye view is annotated with arrows to indicate the direction the cluster is facing. This participant used an environmental layout, where the documents all conform to an environmental cue. In this case, they are aligned to the edges of the virtual environment’s floor.

4.6 Participant 4’s final layout is shown here in birds-eye 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.

4.7 A labeled cluster of related documents created by a participant.

4.8 Example of Participant 20’s emphasis on paper structure. On the left, they have black respondents’ views on the war. On the right, they have the white respondents’ views on the war. The sides have labels for pragmatism or principle, and many notes are annotated for the order they want the quotes to appear in the paper. In the center of the arc is the prompt, where they could reference it quickly.
5.1 The final layout of documents arranged by the participant for writing essay two (top-down view). The layout was a rough dome shape, where the participant sat in a chair in the center of the documents where they could reference documents while writing their essay. 79

5.2 The final layout for the participant for writing essay two, as seen by the participant while they were writing the essay. From this angle, the bulletin board is seen on the left side of their view, while the right side contains documents they wanted to reference during writing. The central document in the foreground is the window they used to write the essay. 80

5.3 This is a test subject using the seated version of IST. They are in a rotating chair with a desk attachment using the Varjo XR-3. A VIVE tracker tracks the location of the desk where a keyboard with numpad allows the user to type. They are also using a Valve Index controller to interact with documents. 83

5.4 We created a file browser implementation of document selection. This allowed us to mimic Windows in that documents could be previewed (on the right) before opening them. It is opened via the virtual menu bar that’s attached the the desk and can be closed via the ‘X’ button in the top right corner. . . 85

5.5 An example of the virtual monitor conditions shown to users. The left is the small condition, which represented a 24-inch 4:3 monitor with 1600x1200 pixels. The right is the large condition, which represented an array of 4x2 of those monitors. This resulted in an 85-inch monitor with 6400x2400 pixels. 86

5.6 These boxplots show the UEQ results for the aggregate categories of Attractiveness, Efficiency, Stimulation, and Novelty as well as significant differences between the results. IST and the large condition are significantly more attractive, efficient, and stimulating over the small condition, and IST is significantly more novel than either large or small, as well as the large condition being significantly more novel than small. 93

5.7 These boxplots show the NASA-TLX results for self-assessed performance and frustration. Participants found IST significantly less frustrating than the small condition. 94

5.8 These boxplots show the results of score on each dataset (left), and score with each condition (right). There was significant differences in scoring for each dataset, but no significant differences in score for each condition. . . . . . . . 97

5.9 This set of boxplots shows the number of notes created in each condition, the number of documents that were left open at the end of the session for each condition, the time taken in each condition, and the number of grabs in each condition. There was significance found in the time taken between small and large, as well as the number of grabs between IST and large and IST and small. 98
5.10 This ridgeline plot shows the number of grabs for all users over time normalized over each session.

5.11 These are various examples of strategies seen in the small condition. P8 (left) used multiple small notes next to the document they were currently reading in order to organize their notes. P9 (middle) utilized the document preview window on the file browser instead of opening files. P2 (right), like most participants, had to layer windows on top of each other in the small condition.

5.12 These are various examples of strategies seen in the large condition. P4 (left) created a note for each document in their dataset and organized them into clusters. P2 (right) kept all the documents open and arranged around the screen so they could refer to them while writing their report.

5.13 These are various examples of strategies employed in the IST condition. P13 (upper left) used timelines, and you can see they highlighted the dates in each document. P5 (upper right) used a strategy where they had different "working areas" and the area shown here is where the sorted documents, while they had other areas for their notes and a last one for their report writing. P17 (lower left) created a wall of notes that they referred to (instead of the document set) during the report writing synthesis. P1 (lower right) arranged documents in a 360-degree arc above their head.

5.14 Participants used several different strategies in order to find certain documents in the interview stage. On the left is a count of the number of times each strategy is used in each condition. On the right is a boxplot of the strategy type compared to how long it took for each strategy to find documents.
List of Tables

5.1 This table details the various file system features for the small and large conditions as well as IST. 84
Chapter 1

Introduction

1.1 Background & Motivation

Sensemaking is the “ongoing retrospective development of plausible images that rationalize what people are doing,” [119], or, more simply, the process of “how people make sense out of their experience in the world” [30]. The process is iterative and repetitive, where people will often loop back to repeat steps to gain a greater understanding [62]. As humans, we do this everyday when we research new office chairs to purchase or when we prepare our children’s schedule for the summer. However, the process can be difficult as datasets can be incomplete or large and complex [6].

For example, parents are bombarded with different opinions, research, and options on how to potty train their child. Some research may show that pull-up style training underwear may be effective, while others can state that they impede the learning process. Further research may show that potty training experts state that the process should begin as early as fifteen months, while other articles state that it should start at thirty. Parents need to perform sensemaking to gather information, evaluate the sources, compare and contrast the various methods, and finally make a decision and present it to their client, the toddler. This process can be more difficult when more sources are considered or the decision is more impactful.

Pirolli and Card developed one of the foundational explanations of sensemaking, and created a framework that describes the sensemaking process as a series of stages [94]. This can be seen in figure 1.1. The process they described involves a pipeline of stages that are performed one by one, where the analyst can move between stages as they need. The process involves foraging for data sources, placing relative importance upon each source, creating hypotheses based on these sources, and then presenting the findings. The analyst can move freely between stages as they add context to their understanding of the dataset in aggregate.

Analysts perform the sensemaking process to better understand complex datasets that are often overwhelmingly large [3]. For example, a historical analyst may have thousands of records from a town in France that was destroyed during the First World War, and they want to piece together what life was like for the people in that town. They would have to compare and contrast records to build understanding and meaning, as well as infer the relationships between the various data artifacts. Organizing this many documents in ways that they can be easily accessed is a difficult problem.
Analysts could use physical documents and spread them out on a bulletin board, using twine to show their relationships and annotations to summarize documents or label clusters. This has often been depicted in movies and television shows as ways detectives solve cases or how conspiracy theorists understand “what’s really going on.” An example from HBO’s *True Detective* can be seen in figure 1.2. While these methods can prove useful, the sheer number of documents required in some analyses can clutter work spaces and prevent collaboration. With these limiting factors, physical documents can be difficult to work with.

Often, data analysts turn to visual analytics to better organize data, understand relationships, or summarize their findings to share with others [26]. This field is often split between two types of analysis, however. Quantitative datasets can help analysts understand relationships with numerical data through charts such as heatmaps [66] or geographic location information [41]. Non-quantitative datasets are often filled with text documents and pictures, such as documentation of code [60] or past conference paper abstracts [45]. These two general areas require different visualizations as well as different ways to process and share that data. A significant portion of visual analytics research has focused on quantitative
datasets, as abstractions for numerical data can be extremely effective, but the support for non-quantitative datasets is less common. In this work, we focus on the latter.

Researchers and analysts need to be able to view non-quantitative datasets (images, text documents, videos, etc.) and often use either desktops or laptops. However, the traditional computer experience is limited by its display technology. In Valve’s May 2020 Hardware & Software survey, 86.91% of respondents have a primary display resolution of 1920 x 1080 [1]. This resolution limits users to seeing only one or two documents simultaneously, which can make it difficult to understand relationships in larger document sets. One potential way to overcome this limitation is through virtual navigation through multiple “virtual” desktops or through pan and zoom interfaces. However, users tend to forget where documents or objects of interest are located using these methods and spend a significant amount of time and cognitive overhead navigating the interface searching for them [10, 97]. Therefore, we are exploring ways of improving our display technology.

One visual analytics tool that specifically assists analysts with large non-quantitative datasets is the Space to Think [8]. Andrews & North’s tool involved using a large, high-resolution
display (made out of eight 30” monitors, each with a resolution of 2560x1600) that wrapped around the user. They then conducted a study that compared how participants used built-in Microsoft Windows software (e.g., WordPad, Windows Picture Viewer) on both a traditional 17” display and their space to think display to solve a sensemaking task. They found many behavioral differences between the two display technologies, including how the two conditions had differences in how users took notes and externalized memory. Their second study had professional analysts perform the same task with only the large display condition. They found that analysts used common organizational styles that would not have worked on a small display. Furthermore, they observed interaction methods that mimicked those performed with physical documents. While the displays could nearly surround the user, there is still room for some improvement.

First, while large, high-resolution displays are large enough that they force users to turn their heads to see the entirety of the displays, they are still inherently a two dimensional (2D) experience. Beyond overlapping windows, there is no sense of depth with the system or display. With a more rich three dimensional (3D) environment, the user could adjust document layouts to be viewable from different angles or use other positional cues like pitch, yaw, or roll to indicate meaning.

Second, Space to Think was able to leverage 2D spatial memory, but humans reside in the 3D world where we can ascribe meaning to 3D space. For example, we can relate certain figures to landmarks in a scene, such as a corner or edge of a wall or floor. Previous work has found that humans are easily able to use space to understand and convey relationships between objects in 3D space [61], and more work has shown that physical travel can enhance recall [97].

Third, Space to Think was built using the 2D interaction methods, while immersive environments allows for richer 3D interaction. The move to 3D interfaces allows for continuity in how users interact with these documents as compared to the real world. For example, a user could change a modeling tool in a VR program by a drop down menu like they would see in a desktop application, or through making a virtual tool palate where they can change tools as easily as picking up a different paint brush. The latter kind of interaction is described as a natural interaction technique where the user leverages their knowledge of the real world to better understand how to interact with virtual objects [116]. Furthermore, 2D interaction methods do not take into consideration many embodied input streams, such as head, eye, or body tracking. In our previous example, a user would not be able to pick up a 3D tool using a 2D interaction technique easily. An immersive environment allows for more natural interaction techniques that users should be able to grasp easily.

Recently, virtual environment displays have become more affordable, with the Meta Quest 2 and the Playstation VR 2 headsets and controllers being comparable in price to modern video game consoles. As more people purchase these devices, users have more familiarity with both virtual environments and 3D user interactions. Similarly, AR devices have also had several mass-market releases such as the Magic Leap Magic Leap 2 and Microsoft’s
1.2. Approach

HoloLens 2. AR devices and interactions have been heralded as the future of productivity, and are predicted to become as ubiquitous as smartphones in the future [86]. AR allows an easy way for the user to context switch between the sensemaking process, writing up reports of what they discovered, or other necessary activities.

Immersive analytics has also been used to assist analysts to better understand and share their data through large immersive environments. The field has been born through the combination of visual analytics, augmented reality (AR) and virtual reality (VR), and human-computer interaction [81]. Immersive analytics has been shown to effectively assist users in understanding complex quantitative data such as bird migrations [87] or multivariate data [27]. However, immersive analytics has not been heavily explored to represent non-quantitative datasets in virtual environments.

1.2 Approach

We conducted an initial pilot study to understand the role of 3D immersive space in performing sensemaking tasks [11]. In this study, we had a basic version of IST that showed documents as rectangles with text on them that the user could move around the environment. We ran eight participants through a sensemaking task, but found several design challenges that had to be overcome. First and foremost, we found that the HMD that we used for that study, an Oculus Rift, limited the area that users could utilize due to its short tether. This resulted in users standing in one place for most of the study. Furthermore, we found that users leveraged embodied cognition to complete the sensemaking task. Embodied cognition
is the idea that our thinking process is rooted in how we interact with the world. In this case, participants tended to collect similar or important documents in certain locations to denote additional meaning. In our design, we try to support users’ ability to use embodied cognition through additional ways to externalize their memory onto the environment. We can learn from this pilot study and design a better approach for performing these tasks in virtual environments.

We designed an immersive analytics approach that focuses on large sets of non-quantitative data that we call the Immersive Space to Think (IST). As an extension of Space to Think, IST can take a large set of multimedia documents and display them in a large, tracked, 3D immersive space. Users can manipulate these documents through grabbing and placing them in the nearly unlimited interaction space and appoint meaning to the documents’ relative position or rotation. Examples of these documents and the virtual environment presented to the user can be seen in figure 1.3. Like Space to Think, analysts can use IST to extract meaning from these document sets by leveraging various sensemaking strategies and processes such as storytelling or reframing [6, 32, 62].

IST also allows for its users to use various forms of embodied cognition to offload meaning onto the environment. Physical navigation, for example, has been shown to perform better than virtual navigation seen in traditional desktop interfaces by both increasing user performance and user satisfaction [10, 68]. Furthermore, IST affords externalizing memory by allowing users to directly annotate documents through notes and labels, which is similar to using sticky notes in the real world or keeping notes in a text document [53]. This allows the users to recall their previous thoughts on the documents more easily when they return to them, or to assist others by communicating summaries of document clusters.

We want to better understand how the coupled system of user and IST can best work together. Much like the original Space to Think paper, we want to understand what virtual representations of data are created with IST, and how can those representations be better supported by IST itself [8]. Furthermore, we want to better understand how virtual and augmented reality implementations of IST perform compared to both large, high-resolution display implementations of Space to Think as well as traditional single monitor desktop implementations.

1.3 Research Overview

**RQ1: How can we design analytic tools for 3D immersive space for sensemaking with large non-quantitative datasets?**

While quantitative information has been extensively researched in immersive analytics, non-quantitative datasets have seen less focus. It is expected that the two forms of data will be assisted by different affordances, as well as different ways of looking at and representing the
1.3. Research Overview

data. This, in turn, means that non-quantitative datasets need different design principles in order to better support the act of sensemaking with these datasets. From our initial pilot study described earlier, we have a good point of reference to design the IST approach. Furthermore, we can add more complex and nuanced interactions from our collected observations. However, as we collect more data we will likely find more user needs that need to be addressed. Therefore, we want to work on improving upon IST’s design at every available opportunity.

**RQ2: How does 3D immersive space affect users performing sensemaking tasks with non-quantitative datasets?**

While there are many ways that IST and immersive environments can assist users, this work aims to figure out how users leverage different aspects of the 3D space around them. While we offer the immersive space and interaction methods with the documents, we don’t know how these aspects of IST will affect users’ sensemaking process or strategies. The hypothesis is that users will employ common strategies that can be extrapolated and then the design of IST can be iterated upon to further complement and support these strategies. Furthermore, it is expected that users will create structures of documents that connote additional meaning through proximity, relative position, or rotation along one or more axes.

**RQ3: What are the relative benefits for 3D Immersive space as compared to large scale, high-resolution displays and traditional methods of performing sensemaking tasks?**

While Space to Think has the limitations of being linked to a 2D display and leveraging large, high-resolution monitors that are still relatively small compared to room-scale immersive environments, it is not understood if the additional space and interaction methods are of benefit to the user. Does IST actually afford a better, deeper understanding of a document set over the original Space to Think? How do either of these tools compare to a traditional desktop setup? These are questions that aim to understand which tool actually performs the best in the hands of analysts.

1.3.1 Organization

This dissertation is structured to address the research questions directly, but not how the studies were performed chronologically. This is due to how the research questions inform considerations on how to address each other. To address RQ2 and understand how people will use immersive space in order to perform sensemaking tasks, we need an initial design that RQ1 addresses. However, there is a feedback loop where once we better understand how immersive space is used from RQ2, we can improve the design of IST to better answer
CHAPTER 1. INTRODUCTION

RQ1. Once we have answered both of those research questions and have a refined design and interface for IST, we can address RQ3 and look into if this approach has relative benefits as compared to other tools for sensemaking.

Chapter two covers related work, including theoretical foundations for this dissertation, sensemaking in various domains, visual analytics, and immersive analytics.

The third chapter addresses the first research question and describes how the IST approach is designed and how the design evolved as studies were evaluated. It then discusses a study that addresses where on Milgram-Kishino’s virtuality continuum IST best supports the sensemaking process [83]. It was the second main study performed in Fall 2021. In this mixed methods study, we found that participants favored aspects of AR but generally preferred VR and felt it better supported focus. Furthermore, through the feedback and designing the interaction techniques, we recommend IST adopts a blend of AR and VR known as augmented virtuality to support users’ focus while also providing them the ability to interact with real-world tools.

The fourth chapter addresses the second research question and details a study of strategies employed by novice users answering a historical analysis prompt. This was the first main study performed in Fall 2019. In this mixed methods study, we found participants create 2 or 2.5D structures in semi-circular, environmental, or planar layouts while performing the sensemaking task in IST. They further utilized most of the tracked area provided to them and offloaded cognition onto the environment through notes, labels, and highlights in varying amounts.

Chapter five deals with comparing immersive space and 3D user interaction techniques with more traditional 2D desktop computer technologies to address the third research question. Two studies are then discussed. In the first, a small study performed in Fall 2019, a single participant used both IST and a traditional desktop setup to perform historical analysis tasks where they generated essay responses. Professional assessment of those essays found that the essays generated in IST were of a higher quality with a more nuanced understanding of the source material. The second study was performed in the summer of 2022 and compared different amounts of space and 2D vs. 3D interaction methods. We found that participants had a higher engagement with the documents in IST and used strategies to mitigate the lack of space in the small and large 2D conditions.

Chapter six concludes by discussing each research question and how the studies addressed them, as well as future directions for this research.
Chapter 2

Related Work

2.1 Theoretical Foundations

Before we delve into sensemaking tools - for both traditional computer systems and virtual environments - we first cover a baseline understanding of how humans process data. We cover the theoretical foundations of embodied and situated cognition to better understand how our minds process activities, and then we cover how we perform the act of sensemaking.

2.1.1 Embodied & Situated Cognition

Embodied cognition is the idea that cognitive processes are rooted in the body’s interaction with the world [122]. Our thought processes extend to both what we do with our body and what we interact with. Wilson detailed six implications of this theory:

1. Cognition is situated
2. Cognition is time pressured
3. We off-load cognitive work onto the environment
4. The environment is part of the cognitive system
5. Cognition is for action
6. Off-line cognition is body-based.

While these vary in importance for IST, each has relevance to its design. Situated cognition is “cognition that takes place in the context of task-relevant inputs and outputs,” and situations can help generate understanding through activity [20]. In other words, how we perceive the environment and the context surrounding it affects how we perform tasks and generate knowledge. This will affect the design of IST significantly. The environment should be designed such that it can help the user focus and perform the sensemaking task. However, this doesn’t just mean virtual objects; the physical world around them also needs to be helpful. For a VR variant, this might mean changes in the tracked area so the user doesn’t
have to spend cognitive resources to avoid obstacles, while for AR variants, this might mean that they have additional real-world tools handy. Therefore, care must be taken in the design of the environments for IST.

The third item on the list, that we offload cognition onto our environment, is also relevant to the design of IST. While performing sensemaking analysis, analysts often take notes to remember key insights into what they are trying to understand so they don’t have to keep it in their internal memory stores. This is also called externalizing memory, where a person’s thoughts are morphed and represented in the environment [121]. Also, this externalization doesn’t necessarily come in the form of annotations. How the data is structured can also indicate meaning. For example, in Kang et al.’s visual analytics tool Jigsaw evaluation study the non-Jigsaw participants often used physical documents and clustered them to form meaning [59]. The process of creating these clusters was also a form of offloading cognition onto the environment. This plays into the fourth point, that the environment is part of the cognitive system. Zhang & Norman expand upon this point, where they found that people not only use the environment as memory aids, such as using fingers or written numbers to remember amounts, but that they will break down complex tasks into component levels to analyze independently [124]. We can then return to the ways we have offloaded and externalized cognition and reframe our understanding of each individual component using those externalizations. Once we understand a key insight, we can then apply that understanding to the rest of the data.

Distributed cognition is related to embodied and situated cognition. The concept posits that the representation of knowledge is distributed in the minds of individuals and through their environment in artifacts [85]. A cognitive system or activity includes both the people performing the cognition and the artifacts they use. This is similar to the idea of embodied cognition in that it encompasses the first, third, fourth and fifth points by Wilson. However, distributed cognition takes into account that multiple minds may be working on the same activity [47]. Therefore, a single cognitive system can be said to be distributed amongst an environment, minds, and artifacts.

From these points, it’s clear that annotation can assist with the sensemaking process. There have been many systems designed for virtual environments that support different types of annotations [14, 23, 46]. Clergeaud and Guitton developed a method for extracting annotations on objects, locations, or tasks in virtual environments and recording them for later use outside the virtual environment [23]. Guerreiro et al. make the point that with asynchronous collaboration in virtual environments, simple text-based annotations may not be enough to accurately convey meaning [46]. They designed and implemented a system for adding text, audio, or video annotations to virtual environments to enhance a design review process for collaborative engineering projects.
2.1. Theoretical Foundations

Spatial Memory and Organization

Robertson et al. define spatial memory as “the ability to remember where you put something,” and this can be supported digitally in many ways [98]. They went on to explore how 3D virtual environments affected spatial memory with a dataset of 100 web pages. In their study, they found that a 3D virtual environment with passive landmark cues performed better in terms of how fast participants found particular web pages as well as fewer incorrect retrievals as compared to a web browser’s text-based favorites system. Wickens and Hollands continue this argument, stating that designers should create these landmarks when creating synthetic environments to assist with navigation and spatial memory tasks [121]. Mann et al. looked into the idea of spatial navigation through a virtual environment and how it impacts recall [80]. They had participants perform a noun memorization task through a mental method of loci, a desktop virtual environment method of loci, an immersive virtual environment method of loci, or however the participant usually completed memorization tasks. The two virtual environment conditions involved virtual navigation through the environment with landmarks with the words they needed to memorize. They found that participants performed better with the two virtual environment methods over the mental method.

Physical navigation through a 3D space involves understanding the structure of the space, which has been shown to increase user performance with large, high-resolution displays. Ball et al. found that increased physical navigation (as opposed to virtual navigation) increased performance during a real-estate visualization task [10]. Relatedly, Andrews et al. looked at the effect spatial memory had on Space to Think and its large, high-resolution displays [9]. They found that users would offload cognition using the displays through taking advantage of spatial memory to remember where information was stored in the visualization. Comparatively, the smaller display, where participants had to leverage virtual navigation such as panning and zooming, required longer search times to find their information groupings and clusters. Andrews and North expand upon this, stating this effect occurs through embodied cognition and that natural human spatial abilities afford this form of distributed cognition [7]. Interestingly, they think this effect is not limited to large, high-resolution displays, but likely extends to other forms of immersive displays.

Egocentric body movements, as well as interaction techniques, have also been shown to assist with spatial memory [97, 114]. Rädle et al. performed two studies evaluating a pan and zoom interface on a large, high-resolution display using egocentric body movements as compared to traditional multi-touch panning and zooming [97]. They found that their new method that leveraged embodied cognition through their egocentric interaction technique was faster and more efficient than the traditional method. Furthermore, participants who used their new technique performed significantly better on a long-term spatial memory task. Similarly, Tan et al. found egocentric interfaces that leveraged the awareness of the body’s position relative to its environment assisted with spatial memory tasks [114].

Spatial memory and its effect on sensemaking has been looked at in traditional 2D monitor
settings as well. Cockburn and McKenzie focused on evaluating the effectiveness of 2D and 3D models (via 2D, 2 ½ D, and 3D variables) for performing a sensemaking task in both virtual and real-world settings [24]. Their experiment looked at confidence in their organization as well as recall in location of specific documents. They found that recall and confidence both fell as dimensionality increased. However, their experiment did have limitations. Both the screen they used and the real-world space were small in size, which limited the amount of space they could utilize for spatial memory. Furthermore, the physical mockup was created using fishing line that could physically impede the participants’ movement when organizing documents, as well as limited the actual 3D placement to wherever there was a fishing line. This caused groupings to have at least 5cm of space between objects, which was a significant portion of the total space (90cm x 90cm x 75cm area). Lastly, participants could not resize documents or add annotations in either condition, which likely impacted the results.

Spatial memory can be expressed in organizational strategies as well. Robinson ran a study where domain experts performed a collaborative intelligence analysis sensemaking task using note cards as data artifacts [99]. They found that participants would use spatial metaphors in offloading cognition such as topic clustering or social networks in order to complete the task. In addition, participants would use localized areas for certain ideas. Hypotheses, for example, were typically grouped together even if they concerned different topics. Other metaphors would also use spatial understanding to convey meaning. For example, timelines are inherently spatial; later events would be placed to the right while earlier events would be on the left. We expect similar encodings in IST’s 3D environment. Relatedly, Liu et al. performed 2 studies investigating how document layout curvature in 3D immersive space affects spatial memory and user experience [77]. In their first study, they found users perform significantly better at recall and spatial memory tasks using flat planar layouts over circular wrap-around layouts. For their second study, they replaced the circular wrap-around layout with a semicircular wrap-around layout. With this change, they found no significant differences in spatial memory performance, but eleven of the twelve participants preferred the semicircular layout. This may indicate that users will gravitate towards using a semicircular layout within their own organizational schemes.

Spatial memory has also been explored when used in conjunction with immersive environments. Lages and Bowman performed a study comparing physical locomotion to virtual navigation to study how users perform in a demanding visuospatial task [68]. They found that participants with high spatial ability but low or middling gaming experience performed significantly better with the physical locomotion interface regardless of their measured spatial ability, while high gaming experience found no significant difference between the interfaces. This indicates that our experiences in a 3D world can overcome inexperience in other realms such as experience in controlling camera angles.
2.1.2 Sensemaking

Pirolli and Card developed one of the most well-known models of sensemaking from an intelligence analyst’s perspective [93, 94, 101]. As seen in the image from the introduction, figure 1.1, sensemaking is a series of steps performed in two main loops: the foraging loop and the sensemaking loop.

As its name implies, the foraging loop concerns itself with how the analyst finds and evaluates new data sources. The analyst has to search and filter through their data sources creating a “shoebox” of good sources that are relevant to the question the analyst is trying to solve. This filtering process involves only skimming the data. The sources in the shoebox are then carefully read to find details that support or disprove the initial theory, which creates an evidence file. At this stage, the analyst moves into the sensemaking loop.

It is during the sensemaking loop that the analyst takes the evidence file to fully understand what the dataset means in relation to their original argument. The remaining documents are then organized schematically to create a visualization. For example, they could be ordered in a general timeline of events, or clustered by main topic. From these schema, a case is built and hypotheses are created to explain their relationships to the initial theory, and the work is presented to the client.

The analyst can go back to previous stages at any point in this pipeline in order to refine their findings and create a better overall argument. For example, details in data sources can reveal new questions that will, in turn, need their own data sources to create a better argument, or a reevaluation of hypotheses could trigger a new search for supporting details. The client could even give feedback at the presentation stage that could require more examination.

While this work focuses on Pirolli and Card’s model, there have been criticisms that their model is too linear [57]. Competing models have also been developed that capture the non-linearity of the process. These models represent unique spins on sensemaking that can further assist with the design of IST.

Klein and Moon describe the process of sensemaking as the formulation of a mental model of what happened and a mental simulation of what may happen next [62]. They further created a model for sensemaking that revolves around the concept of data frames, which is depicted in figure 2.1 [63]. In their model, an analyst formulates a hypothesis and creates a data-frame or understanding about a situation. As more data sources come in, the analyst can elaborate on the frame adding new details or question and doubt the frame as its currently constructed. If the incoming data contradicts what is currently understood about the theory, that data source is evaluated for quality. If the new data is judged to be of a low quality, it is tracked in case it is relevant again, but the frame and hypotheses are preserved. If the new data is of a higher quality, however, the analyst’s understanding is “reframed” to incorporate the new data and relationships. It is through this constant questioning of the current understanding that allows the analyst to fully understand the dataset.
**Figure 2.1:** Klein and Moon’s Data/Frame sensemaking process. It involves a frame of understanding that is adjusted or reframed as additional data is considered by the analyst [63]. [102].

This model of sensemaking lends itself to immersive space and its ability to leverage spatial memory and organization. As additional data artifacts are considered, the analyst must consider how it interacts with other artifacts. This, in turn, requires them to remember where they put related artifacts to compare and contrast the sources simultaneously [98]. Immersive space affords better recall through physical navigation, which would allow the analyst to sort documents efficiently and recall their organization quickly [10]. This also supports tracking of specific data artifacts that the analyst deems anomalous. Part of the Data/Frame model shows that analysts must track questionable data and continually re-evaluate how that fits in with the rest of the frame. This can similarly be assisted through the use of 3D space and landmark cues, which have been shown to have increased performance on finding particular documents through leveraging spatial memory [98]. Since this model relies heavily on recalling particular documents, it is well supported through immersive space and egocentric user interaction techniques.

Sacha et al. developed a model of knowledge generation for visual analytics where the human and the computer work together and create a better understanding of a dataset [102]. In other words, it’s a model on how visualizations can assist analysts to perform sensemaking on certain datasets. As depicted in figure 2.2, the model explicitly states what the computer, and thus the visualization, should do for the analyst. The computer should collect input
data and organize it via provenance data (details on how the input was created, gathered, selected, or preprocessed) and/or metadata (second order data or “data about the data”, which is chosen by the interest of the visualization creator). The computer creates a model for the data, which is either a set of descriptive statistics or a data mining algorithm that can assist the analyst in determining information about the data in aggregate. Both of these are used in a data visualization that allows the analyst to detect patterns in the dataset. The analyst in this model traverses the exploration, verification, and knowledge generation loops. In the exploration loop, the analyst can either interact with the data to provide a different viewpoint or other setting (as determined by the visualization designer) or find an interesting observation about the dataset from the visualization. This leads to the verification loop where the analyst takes the findings from the visualization and comes up with an insight about the dataset as a whole. This can lead to a hypothesis that can be tested - either by the visualization or through statistical means. When the hypothesis is tested and proved accurate, knowledge is then generated. In our proposed IST approach, the computer will create visualizations of multimedia documents that the user can interact with in various ways, allowing for findings and insights to be generated by viewing the documents more closely.

Elm et al. created a support function model for intelligence analysis that focuses on repeatedly narrowing and broadening the scope of the analysis until a decision is reached [32]. Intelligence analysts face a data overload problem in that there is often too many data sources in an effort to not miss any critical events. However, the analyst must also be able to incorporate information in a broader sense. They created a model to support decision analysis that requires analysts to shift between narrowing and broadening the focus of their sensemaking. They argue that the narrowing helps assist the analyst come to a final conclusion, but is in constant conflict with broadening which helps the analyst consider all options. Furthermore, allowing the focus to broaden occasionally can help the analyst
redirect attention to previously unexplored ideas. Finally, the key argument the authors make is that analysts need to be able to tell a story to tie together their data sources and create plausible hypotheses that explain the dataset as a whole. There is often not a clean relationship between all data sources, so hypotheses need to fill in the blanks between them.

We envision that IST will be able to assist users in creating stories to explain relationships, since it enables them to sort and organize the documents into structures or orderings. Through these structures or frameworks of data extrapolation, people can have more complex understanding of the underlying data \[119\]. 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. Similarly, Ancona discusses what strategies analysts should use during sensemaking \[6\]. They liken sensemaking to cartography, where analysts must create a map or story with the data sources to understand what actually occurred.

An important point in sensemaking is that the analyst does not work in a vacuum. Humans belong to various organizations such as a citizen of their country, a worker in their company, or even as small as a member of a family. Furthermore, Taylor and Van Every make the point that sensemaking “is not context-free networking” \[115\]. An analyst has many overarching and competing pressures from the organization they work for, their country, and their community that affects how they understand the world and their data sources. Weber and Glynn further this argument, stating that the sensemaking process is impacted by the context of the individual or the organization and their social norms, social mechanisms, and underlying setting \[118\]. For example, an African-American man in the United States will view interactions with police or governmental officials in a different way than a white person would. Our understanding of the world is inherently affected by our past and the relationships we have within our organizational structures. IST can be augmented to address these limitations and present opposing viewpoints in ways that can combat our prejudices.

2.2 Sensemaking in Various Domains

Sensemaking isn’t limited to a single topic. We as humans employ the process when we are choosing which laptop to buy or when writing a proposal. There are many professional domains that leverage the process regularly, including science, the intelligence community, and history researchers. In this section we cover how each of them engage in the process.

2.2.1 Scientific Analysis

Scientists engage in scientific analysis in order to make sense of the natural world. In particular, this sensemaking process must be systematic and focused. Passmore et al. define how scientists perform this process in detail and define the functional unit of scientific
thought as “models” of science [90]. They argue that these models must be integrated into science education such that they offer a more authentic experience of the science sensemaking process than the ubiquitous, and somewhat misleading, “scientific method.” Furthermore, they point out that the process of scientific sensemaking is not radically different from the sensemaking performed by people in general activities. Fleck and Simon also worked at passing on teaching science through scientific models [39]. Their AR tool allowed elementary school students to interact with celestial bodies and test hypotheses in a simulation of how the earth and moon cast shadows and reflect sunlight. In a study with French 10-11 year olds, they compared their AR tool to a traditional physical model. Fleck and Simon found that more students had better mental models of the relationships between the three celestial bodies when using the AR tool. These results indicate that virtual environments may assist with engagement of the sensemaking process.

Often, sensemaking tools are developed to specifically assist scientists to manage vast quantities of data. For example, Dahshan et al. created a tool that helps scientists manage, search, and organize complex multi-variate simulations [28]. Their tool, GLEE, uses a blend of semantic interaction and visualizations to assist scientists find interesting relationships in their data. One way it performs this is through the scientists’ own discoveries. If the scientist sees a potential relationship between two simulations, the tool will attempt to build a model of similarity between the two simulations and provide other similar simulations for the scientist to process. The authors ran three studies with GLEE: the first with an elementary particle dataset and ten scientists from different fields, the second and third with oilfield wastewater disposal dataset and five geoscience domain experts. They received positive feedback from their participants on the quality of the design and its ability to support analysis and sensemaking.

In some cases, the tools available to scientists can help define the future of that field. Gahegan and Harrower detailed a plan on how to use visualizations and geocomputational tools in order to better understand geological data, and they use that plan to chart the research challenges in the field of geology [41]. They, as in other fields, discuss how there is an abundance of data in databases that needs to be mined for relationships. They also point out that, while relationships may be discovered by machine learning algorithms, it takes a human to verify that the relationships exist in a meaningful sense. In other words, the machines cannot work alone. Furthermore, they state that the future of the field stands to move in several directions, including through improvements to the human computer interface design and visualization techniques.

### 2.2.2 Intelligence Analysis

Intelligence analysis rapidly expanded in importance and scope in the wake of the September 11th attacks on the World Trade Center [26]. Sensemaking tools are also often created and evaluated using intelligence analysis as a baseline task [18, 58, 74, 99]. In particular, there
became a large focus on distilling the huge amount of data that analysts have to deal with. Johnston performed an ethnographic study of intelligence analysts in order to understand their needs and develop ways to improve training and quality of analysis [56]. He posited that intelligence analysis is an inherently careful collaborative process due to the fact that reports were deliberated in groups and there were political ramifications if they were inaccurate. He details the analysis cycle in his work. It goes from planning and direction to data collection to data processing to analysis and production to dissemination. This is similar to the Pirolli and Card sensemaking model, but it also delves deeper into the planning and data collection process. Johnston’s final recommendations were to build a better training program, work to combat ethnocentrism, and improve team cohesion and socialization to allow more voices to be heard in the process of analysis.

There has been some effort to create decentralized sensemaking tools. Li et al. developed a way to create sensemaking microtasks to get crowdworkers to solve intelligence analysis tasks asynchronously [74]. They designed a process that creates tasks based on each of the five steps in the sensemaking loop from Pirolli and Card’s sensemaking model (see figure 1.1). One key issue they discovered in the design of their tool was that partial analysis needed to have a defined beginning and ending in order to communicate sensemaking results. Since analysts often have a “black-box” understanding of the data and their layouts, it can be hard to transfer their results to another analyst without significant explanation. They solved this by creating well-defined context slices that required the crowdworkers to defend their choices for each step, and having a “majority-vote” method of resolving conflicts. Majority-vote entails several crowdworkers performing the same task and the result that occurs most often is chosen as the most-likely answer. They then evaluated their tool using three sensemaking task datasets: an easy mystery puzzle, a medium difficulty “Clue” game puzzle, and a difficult intelligence analysis task used to test professional analysts. They found that crowdworkers were able to effectively solve each task.

2.2.3 Historical Analysis

Historical analysis typically involves taking an overwhelming amount of sources and distilling them into stories that help us better understand what happened in the past [40]. These data sources are, for the most part, non-quantitative in nature. For example, journal entries of presidents can help a historian understand their thought process, or newspaper clippings can help build a narrative on what the public opinion was on a series of events. Abbot argues that digital tools that help researchers make sense of the myriad sources available to them are required and, in concert with physical collections, afford a greater connection with the data [2].

Teaching how to perform historical analysis and the sensemaking of historical data can be difficult. Hicks and Doolittle developed a system for helping teach elementary school students how to take several disparate multimedia sources to understand a historical event
2.3. VISUAL ANALYTICS

[51]. They developed the SCIM-C strategy to help teach students how to perform historical analysis and focuses on five phases: summarizing, contextualizing, inferring, monitoring, and corroborating. Summarizing is the skill of quickly distilling a document into its key attributes, such as type of source, subject, author, purpose and other facts or ideas that the source presents. Contextualizing involves understanding the general world that this source debuted in. For example, a letter from one person to another may hold a different meaning if it was written in peacetime or in a war-torn country. Inferring involves taking the context and looking at the sources through that lens to see what subtext the source conveys. The student must then monitor their reflections on the source. In other words, they must gauge whether the source makes sense in its environment and evaluate the quality of the source. The last phase of the strategy is to look for corroborating evidence. This stage involves taking multiple sources of data and create an understanding of what they mean in aggregate. This process can be seen as a way for students to learn the process of sensemaking, as it shares many similarities to Pirolli and Card’s sensemaking model [94].

Hicks and Doolittle followed up on their strategy by developing a software tool that teaches the process and evaluating it with a group of 77 undergraduate students [49, 50]. They had the students engage in three 30-minute tutorials on the SCIM-C strategy, testing their ability to apply what they learned after each session. Hicks and Doolittle found that the students could appreciably recall the strategy and could apply the first three stages well on new historical sources. However, they did not find significant application of the monitoring step by the students towards new sources. The authors explain that it’s difficult to measure the amount of monitoring performed by people as it’s a mainly internal process relying on cognitive and metacognitive abilities. Altogether, their study showed an appreciable ability to give students an understanding of how to perform historical analysis.

2.3 Visual Analytics

The field of visual analytics helps create tools that assist analysts in quickly finding insights into large and complex datasets. In this section, we cover the basic concepts behind the field and cover a few ways it specifically helps with the sensemaking process. We then cover a sample of the modern tools created by the field.

2.3.1 Concept

In 2005, Thomas and Cook defined visual analytics as “the science of analytical reasoning facilitated by interactive visual interfaces [26]. They noted that analytical reasoning is a process that allows us to understand large complex datasets. This is quite similar to the sensemaking process, and they use interactive visuals to better understand and communicate insights on these datasets. Thomas and Cook further argue that visualizations allow the
underlying data - that can be large and unwieldy - to be easily understood and communicated to others. They perform this through abstract representations or transformations. One important aspect they state is that these visualizations must be able to support continually added data. As the amount of information grows, the visualization must be able to handle it and update to reflect the changes to the analyst. Data visualizations such as these are widely used in the intelligence analysis community.

In another vein, Endert et al. made the observation that most visual analytics tools and software focus on either the foraging loop or the sensemaking loop [34]. In this case, tools focusing on the foraging loop are ways of finding new sources to add to an analyst’s collection, while tools focusing on the sensemaking loop deal with datasets that have already been pruned of less helpful sources. However, we envision IST as a way to address both the foraging loop and the sensemaking loop by providing an abundance of data sources that the user can easily manage foraging through as well as the tools to understand the data’s relationships and perform the sensemaking loop.

One of the more important aspects of visual analytics is how they affect the mental model of the data for the user. Liu and Stasko’s work on this issue highlights how the user internally understands the data and further that the data visualization is primarily used for external anchoring, the process of foraging for information, and cognitive offloading [78]. External anchoring means understanding how data is linked to the visualization. For example, if the visualization creates circles based on population sizes in countries, you externally anchor each circle to a country based on an attribute of that circle. This could be abstract, such as using the flag of that country as the coloration for the circle, or be as direct as a label on each circle stating the country it depicts. Data visualizations help with the foraging process through restructuring how the data is displayed. Using the previous population example, a table of population numbers makes it more difficult to gauge relative sizes, while seeing the size difference of various circles abstracts the data in a way that’s easier to understand. Lastly, cognitive offloading for data visualizations is performed through creating pointers that refer to particular save points for the visualization. In this way, we can capture parts of the visualization that have particular meaning for our purposes and store that state for later usage.

In particular, the idea of a mental model is similar to that of the sensemaking process. Using Klein’s model of sensemaking as example, as we use a tool we create a frame of reference for how its used and what the tool is good for. However, once we are challenged by the tool by a surprising interaction, we may need to reframe our mental model of that tool. However, mental models can be inaccurate as long as they functionally allow us to use the tools they are formed around [88]. It doesn’t matter if we don’t fully understand a tool or are misguided about its underlying apparatus as long as the model attains the results we expect.
2.3. VISUAL ANALYTICS

2.3.2 Tools

There are many visual analytics tools of different varieties that tackle many different aspects of the sensemaking process. While a large number of these focus on quantitative datasets, we are going to focus in this subsection on tools that work with non-quantitative or multimedia datasets.

The previously mentioned Space to Think was a workstation designed to support the sensemaking process with large, high-resolution displays [8]. The main benefit of the display was that it could allow users to organize documents using built in applications and windows and leverage spatial relationships to denote additional meaning. Andrews et al. ran a comparative study against a workstation with a 17” monitor where the participants had to solve a sensemaking task from a visual analytics contest. Their data sources were multimedia assets (including pictures, text documents, and a spreadsheet). Participants often had to spend considerable amounts of time searching for documents they had opened on the smaller display and created many physical notes on paper. This indicates that the large, high-resolution display offered a way of externalizing memory by creating enough space that users could find relevant documents easily and keep their annotations visible simultaneously.

Another visual analytics system that specifically looks at non-quantitative data is the Jigsaw tool [108]. Stasko et al. developed the tool to allow analysts to coordinate multiple views of document data while highlighting connections and relationships between the documents. Changes in one view updated across the other views as well. In the list view, entities were listed on the left and right, and relationships were shown through lines connecting the two columns. The graph view would show color coded entity nodes that are specified by the user and any relationships between them are represented by edges. The scatterplot view shows pairwise connections between entities. Lastly, the text view shows the actual reports that generate the other views. These various views assist the analyst with the sensemaking loop in Pirolli and Card’s sensemaking process and can particularly help the analyst generate reports from their findings. Kang et al. evaluated Jigsaw’s design with a user study; asking participants to use paper documents, desktop tools, Entity (a limited version of Jigsaw that only allowed the Text view), or Jigsaw to solve an intelligence analysis task [59]. While overall performance measurements were not the goal, they found that participants using the full version of Jigsaw performed the best at solving the task. Furthermore, the study found that the Jigsaw users leaned heavily on List view to find entity relationships between people and documents, and Jigsaw users were able to employ several different strategies using the tool. Of key note, however, is that the participants in non-Jigsaw settings often tried to group documents into clusters and make connections similar to those that the tool would do automatically.

Khaloo et al. developed a visual analytics tool for viewing codebases called Code Park [60]. Their tool enabled users to see each code class as a room, with members of the class (methods, variables, etc.) displayed on the walls of the room like wallpaper. Their goal was to create a tool that allowed novice users to understand and make sense of the relationships between
code objects. In two studies with their tool, participants reported higher engagement and understanding of relationships with Code Park as compared to a traditional integrated design environment.

2.4 Immersive Analytics

Immersive analytics is built upon the fields of data visualization, visual analytics, virtual and augmented realities, and human-computer interaction [21, 36, 81]. Data visualization and visual analytics deal with how information is presented, while virtual and augmented reality involve the display technologies that the visualizations are seen through. Human-computer interaction is the glue that holds the two together and affords the analysts the ability to interact with the data. These fields are integrated in such a way that most current immersive analytics applications support user sensemaking [107].

2.4.1 Display Technology

Augmented and Virtual Reality has been touted as “the ultimate display” since the early work of Ivan Sutherland in the 1960s [112]. Typically, these displays are worn close to the eyes of the user as a head-worn display (HWD) to offer a freedom of movement that is not offered by stationary displays [70]. These displays are tracked such that the user can see virtual objects and move around or interact with them as if they were real. Furthermore, while VR allows the user to see expansive and entirely virtual worlds, AR affords the ability for the user to stay present in the real world with no time delay and be able to navigate obstacles that would be difficult to detect in virtual reality [103]. It is therefore “augmenting” the real world. In addition, while there has been an effort to use integrated cameras in VR headsets to detect and display obstacles, the ability to see and understand them is still greater in AR [84]. However, AR displays, especially optical see-thru displays, face difficulties in implementation and are often hindered by significantly smaller field of views (FOVs) and rely on less accurate tracking technologies [70].

User interfaces in both VR and AR have been researched heavily, with the display of information being a particular focus [70]. Early work defined several different techniques for how windows display in the world. Feiner et al. detailed three ways for windows to be displayed in augmented reality: head-fixed, surround-fixed, and world-fixed [38]. Head-fixed involved using the AR display as a heads-up display; in other words, the windows are always visible at a fixed point on the display ignoring any head-movement by the user. Surround-fixed meant that the windows are at a fixed rotation around the user but would move with the user as they navigated their environment. Lastly, world-fixed windows were displayed at fixed points in the environment. Users could walk around them and view them from multiple angles. In IST’s case, we use world-fixed data objects that users can move freely. Bell et al.
continued information display management by studying how objects’ view could be manipulated in AR and VR [13]. They developed ways of tagging windows in order to meaningfully communicate how those windows should behave in different contexts. For example, while working collaboratively, windows shouldn’t occlude the collaborators from each other. This idea will become relevant as IST is developed for AR.

While the focus in immersive analytics is on augmented or virtual reality headsets, we should also consider the benefits provided by large, high-resolution displays as a competing option. Some applications of these displays have qualities of immersion, such as wall displays being large enough that it takes some effort to look away from the screen [18, 70]. However, this is also one of it’s largest tradeoffs. Since it doesn’t offer a 360 degree field of regard (the angle at which the virtual environment is viewable around the user), these displays cannot completely immerse the user in a virtual environment. Furthermore, large, high-resolution displays do not support stereoscopic viewing to provide higher quality depth cues to the user. There are some benefits to these displays as compared to VR. As in AR, large, high resolution displays allow for the user to remain acclimated to the real world, but also without being encumbered by a headset.

Collaboration is much easier with a large, high-resolution display as there’s no networking or communication needed between headsets or implementations. For example, a teacher using a wall display can easily communicate to students what they’re trying to point out to discuss by using simple gestures. Bradel et al. used a wall display to study how analysts collaborated on sensemaking tasks while using the same display [18]. They found that participants were able to effectively solve the sensemaking task, though the teams that shared space effectively performed better at the task. Jakobsen and Hornbaek also studied collaboration on a multi-touch wall display [54]. In their study, they had pairs of participants solve a sensemaking task on a shared 2.8m by 1.2m wall display. They found that, while participants spent over 50% of the time working in parallel, teams were able to fluidly move from area to area on the screen and share space with each other without becoming territorial. The authors argue that their display was large enough to not create a scarcity of work-space. In other words, once a work-space is large enough for users to maintain their own areas, fighting over their area becomes less of an issue. Furthermore, the edge of the display nearest a wall and the bottom of the display were used significantly less than the remaining space, which might indicate that users didn’t want to work with documents below their chest level and the physical characteristics of a space, such as a wall, severly impact how a display is used. These considerations will be important for the design of IST.

Even with the benefits of large, high resolution displays, the hardware itself can be difficult to procure, maintain, or store. McGill et al. designed a way to render large, high-resolution displays on AR/VR headsets and make them more ergonomically comfortable [82]. They achieved this through manipulating the relative position of the virtual displays as the user turned their head, with or without “deadspots” where the virtual displays would stop moving. In other words, as the user turned, the virtual displays would rotate in the counter direction at the same rate, pausing at certain intervals (the deadspots) so that the user could use
CHAPTER 2. RELATED WORK

the virtual displays as a normal physical monitor. The authors ran two studies in VR to compare their methods to traditional means. In the first study, they found that users preferred a single deadspot located at the position of the display directly in front of them. In the second study, they implemented a way of switching to different displays being shown through either keyboard input or by gazing at the boundaries of the primary display. They found that the implicit control of the gaze technique was preferred. Their studies imply that we could recreate the original Space to Think using these virtual displays and then compare it to IST using the same equipment.

Often, display technologies are compared against each other to ascertain relative strengths and weaknesses. Steffen et al. compare the affordances between AR, VR and the physical reality in order to provide guidelines for user experience design when designing for AR or VR [109]. For example, while both AR and VR grant the ability to enhance the physical world, AR retains the affordance of understanding the physical context of the user. However, another affordance that VR has is to approximate AR, and some studies use VR to approximate AR HMDs and evaluate interaction methods for proposed tools [67, 71, 72].

Other comparative studies more directly evaluate the effect of mixed reality style on performance metrics. Ping et al. designed a depth perception task for both VR and AR and compared users’ performance in the two modalities [92]. They found that users performed depth estimations better in AR. Suso-Ribera et al. compared different therapy techniques for people with phobias of small animals, including an AR implementation, a VR implementation, and using in-person small animals to help people combat their fear [111]. They found that all three implementations proved to be similarly effective, but their qualitative comparisons between the AR and VR implementations only discussed cost issues. In another comparison, Park & Kim looked at motivations while varying AR and VR for users who were shopping for clothes [89]. They found that motivations impacted the effectiveness of the style of mixed reality, with users who were only browsing for clothes more likely to buy with the VR deployment, but people actively searching for clothes preferring the AR. Voit et al. performed an empirical evaluation study of smart artifacts using five different methods including in-situ, lab studies, AR simulation, virtual reality simulation, and online surveys [117]. While AR scored lower on several usability scales than VR in their findings, it should be noted that they were comparing deployments to a Microsoft HoloLens and an HTC Vive, which have several confounding factors in the comparison such as the tether, weight, field of view, and other variables. Our studies attempt to remove as many confounds as we can, to better evaluate the two styles of mixed reality as fairly as possible. Similarly, our study attempts to recreate the same interactions with the same devices such that we are limiting the differences to the ability to see the real world.
2.4. IMMERSIVE ANALYTICS

2.4.2 Tools

One way of creating immersive analytics tools is to adapt 2D visual analytics tools into 3D variants that could benefit from the third dimension. For example, Kraus et al. developed visualizations based on 2D heatmaps for both a desktop visual analytics scenario and an immersive environment for virtual reality [66]. They found that error rate improved significantly with both the desktop and virtual environment variants over the original 2D heatmap. Furthermore, the VR condition received significantly higher satisfaction, and in qualitative feedback participants reported that peaks were more easily discerned in VR.

Naturally, immersive analytics is often used to take abstract datasets and create a visualization to better understand them. Nim et al. created a VR tool to better visualize and understand bird migrations [87]. 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. For another example, Gold et al. combined Martian rover and satellite data to visualize terrain geometry with the accompanying mineral composition [44]. They found that users had a clearer understanding of the layout of geological data with their visualization. Analysts can use immersive analytics to take multiple data streams and navigate them naturally, and can thereby perform more complex analysis.

Cordeil et al. came up with a way of embedding multivariate data axes in embodied interactions [27]. Their tool, ImAxes, is a VR application where data axes are abstracted into virtual objects. Users can move these axes around to affect how the data is displayed. The system supports the creation of the traditional visualizations of multivariate quantitative data: histograms, scatterplots, scatterplot matrices, parallel coordinate plots, and their 3D counterparts. In addition, the 3D immersive space allowed for the generation of new visualizations, such as a circular parallel coordinate plot, or viewing of traditional plots from different angles. Batch et al. followed this work up by performing several user studies with ImAxes with domain experts [12]. In the primary study, they had twelve economists or data scientists holding a master’s degree or doctorate perform an analysis task with their own datasets in a large 3m x 3m tracked area. They found that participants generally laid out the graphs egocentrically, either in “gallery” or semicircular egocentric arrangements, but none of the participants used the entire tracked area. Furthermore, seven of the participants reported a high level of engagement with their dataset.

There have been some previous efforts into creating immersive analytics tools for non-quantitative datasets. Benford et al. created a virtual environment for viewing documents that is similar to IST [14]. Their tool would present documents based on an initial query using distance to show how relevant the document was to that query. Benford et al. later adapted this tool to create a collaborative virtual environment for browsing the internet [15]. It enabled two or more people to view webpages as windows in a virtual environment while using embodied avatars to communicate where they are in the virtual space. Recent work has focused on how these implementations can assist with workspaces [64, 73, 79]. Kobayashi
et al. explored using HMDs and virtual workspaces for users to perform tasks with multimedia documents to replace large high-resolution display and physical workspaces [64]. They found that participants tended to create spherical workspaces and that their implementation could provide a similar or superior experience to traditional workspaces. Similarly, Luo et al. explored AR sensemaking implementations where they varied collaboration and office furnishings and backgrounds to see the effect on user placement and understanding of non-quantitative multimedia datasets while performing a card-sorting task [79]. They found that users leveraged environmental landmarks, such as furniture or walls, to provide a scaffolding to organize their artifacts and create additional meaning.

2.4.3 Future Directions

Skarbez et al. defined the current research agenda for immersive analytics as well as proposed some of the future directions the field may go [107]. They felt that these five areas were the main areas of focus in immersive analytics: combining human and computer intelligence, the utility of immersion, designing immersive analytics systems, facilitating collaboration through immersion, and changing the process of analysis with immersion. Of particular note, IST has applications in all of these areas.

Combining human and computer intelligence has seen many successes in visual analytics. Endert et al. conducted a survey of the state of visual analytics implementations that used machine learning algorithms to enhance their capabilities [34]. They created four categories of machine learning techniques that were used in combination with visual analytics: dimension reduction, clustering, classification, and regression/correlation models. The authors further specify some future direction for machine learning models in analytic tools. Of particular note for IST, they suggested automated report generation through the algorithm inferring links between concepts, data and analytical results, which could be useful for analysts using IST. Furthermore, they suggest machine-learning classification of analyst annotations through leveraging natural language processing, as well as identifying low-frequency, high-value data sources which could be valuable future additions to IST.

Semantic interaction is one way to leverage machine learning into analytic tools with non-quantitative datasets, which is a machine learning algorithm that changes the data model based on inferences on how the analyst is interacting with the data [18, 19, 33, 45, 120]. For example, if the algorithm detects the user is creating a rough timeline from non-quantitative data sources, the semantic interaction algorithm may suggest additions to their timeline based on the timing of other data sources or the keywords in the documents being sorted. For one example, Endert et al. integrated semantic interaction into their ForceSPIRE visualization tool that took textual documents and laid them out based on the user’s pinning, linking, or dragging of one or more documents [33]. Users could also highlight terms which would then create a label of that term and drag documents that were related to that term towards the label. These interactions could be transferred into IST to assist with the foraging
process for analysts.

Billinghurst et al. detailed types and ways to support collaboration in immersive analytics [16]. They define different types of collaborative immersive analytics as functions of space and time: either the collaboration is co-located or distributed and its either synchronous (the joint task is performed at the same time) or asynchronous (work is performed at different times). The authors further make the point that these systems should prioritize communication between collaborators either in the form of direct communication or annotations and other externalization tools. They further argue that there should be enough space such that users can have their own space as well as shared space to work together.
Chapter 3

Design of the Immersive Space to Think

Designing an immersive analytical tool is a difficult problem that lies at the heart of RQ1. In fact, the design of IST has gone through several iterations along the way in order to better support the process of sensemaking with multimedia documents. After each main study covered in this dissertation, the IST approach was tweaked in order to better support users needs. Additionally, we consulted expert historical and intelligence analysts to refine our design and better support various professional analysts’ needs. How we identified user needs, designed the approach, and how the approach evolved over the course of this dissertation is detailed below.

We also performed a mixed-methods study comparing AR and VR against each other to understand the tradeoffs of each for the sensemaking process. This study was the second main study of the dissertation, performed in the Fall of 2021. It used a version of IST where users could move documents around the space and offload cognition onto the environment using highlights or an integrated keyboard to create notes and labels. Through user feedback, we found that an augmented virtuality approach, as defined by Milgram & Kishino [83], provides a better user experience than either AR or VR. Virtual environments can cut off distractions from the users, and augmenting those with portals to the real world affords the ability to use real-world tools such as keyboards, whiteboards, pen and paper, or non-digitized data artifacts.

3.1 Hardware and User Experience Specifications

In our earlier prototype, users wore a tethered head-mounted display (HMD) and interacted in a small two-meter by two-meter tracked area [11]. In the virtual environment, we provided users with a set of virtual text documents and interaction techniques for grabbing and moving those documents, but no other features to support sensemaking. All of the documents were initially displayed in a stack, such that users could only see the document at the top of the stack until they moved that document. Users complained about the blurriness of the text, as well as the tethered HMD limiting their movement.

Based on this feedback, we improved IST’s design with new features to better capture what
3.2. User Needs

From our pilot study and interviews with professional analysts, we discovered a set of user needs required for the performance of sensemaking tasks.

First, while document movement was already integrated into the prototype, we discovered that users needed to be able to resize those documents. This served two purposes: it made documents easier to read, and it also allowed users to encode relative importance to documents. The larger a document is, the more important it was to the overall task. Conversely, if the document is small, it can be considered less important. Relatedly, the documents should be relatively uniform at the start of a sensemaking task. This is to avoid giving the user undue cues, such as document length denoting importance. Another key observation from the pilot study was the need to support embodied cognition through offering ways of externalizing memory in the environment. When a user discovers an insight, they may have several ways of expressing that discovery so we developed a few ways of allowing participants to encode information. Another need is a way of recording their process. Often analysts will take breaks, so we designed a way to save their progress to continue their sensemaking or review their results. Lastly, the document sets analysts work with often are multimedia-based. Therefore, our approach should be able to support multiple different mediums of non-quantitative data, such as videos, pictures, text documents, and audio recordings.

Taking it from a top-down approach, we further identified several tasks analysts need to perform. These (and their corresponding stages in the Pirolli & Card sensemaking model) are:

1. View multimedia files such as audio, video, and text (Foraging Loop).
2. Search through documents for keywords or phrases (Search & Filter).
3. Organize documents (Shoebox, Evidence File, Schematize).

4. Extract direct quotes or aspects of a document as evidence for an argument (Read & Extract, Search for Support).

5. Support externalizing memory (Sensemaking Loop).

6. Present Results (Presentation).

We need to support each of these activities through our designed interaction methods. Each of the interactions below either directly address one of these tasks or supports other interaction methods to address these tasks.

3.3 Interaction Methods

We designed additional features for IST by consulting experts in historical and intelligence analysis to better understand their processes. We added ways of externalizing memory to offload cognition onto the virtual environment based on the feedback from the experts. These features are described in more detail below.

3.3.1 Single-hand Movement of Documents

Users can move documents by raycasting to a document and pressing the trigger button. Now the document is “grabbed” at the initial point of intersection of the ray and the document. At first, the document could then be translated or rotated at the end of the ray at the fixed initial distance, or be translated towards or away from the user along the ray by pressing on the wireless controller’s trackpad or joystick in the corresponding direction. This supports the third item on the identified task needs list through allowing users to organize documents in 3D space and was how IST was evaluated for the studies in sections 4.1 and 5.2.

Participant feedback from that study evolved the design further, as they wanted greater control over pitch, yaw, and roll of data artifacts. Therefore, we changed the 3DUI manipulation from the ray-casting and reeling technique to the HOMER technique as described by Bowman & Hodges [17]. This also gave users control over the speed that documents moved egocentrically in addition to greater flexibility for data artifact rotation. This was the technique used for the study in section 3.4.

3.3.2 Multi-hand Resizing and Moving of Documents

Users can also “grab” documents with both controllers simultaneously. This allows the user to resize the documents based upon the two initial points from the two controllers in a “pinch-
3.3. Interaction Methods

Figure 3.1: An example of a text document in IST. The UI panel on the left features document-specific interactions that can manipulate the document. These are highlight text selection, copy text selection, copy text selection with a citation, and add a note to the artifact.

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 while the grab points are kept fixed relative to the document. The user can also translate/rotate the documents using this method. This also supports the third item in the identified task needs list, but it further supports the fifth item as users can externalize memory through relative size or rotation of documents. However, after we added keyboard input for the studies in sections 3.4 and 5.2, we switched to a single controller to allow for greater flexibility and ease switching from artifact interaction to typing and vice-versa.

3.3.3 Text scrolling

Text documents are often larger than the allotted visible space on a document, so we designed a scrolling metaphor much like the scrollbar on a traditional desktop window. To scroll a document, the user needs to point at a document, then drag their finger on the trackpad to scroll up or down in a similar fashion to scrolling on a smartphone web browser.
3.3.4 Text Highlighting

A key part of analyzing text documents is to make themes, keywords or quotes more salient for later use. To address this, we added highlighting capabilities to IST. Users can point a document and press the secondary action button on the controller. The word or words that intersect with the controller’s ray are now selected. The user can then hold down the secondary action button and move the ray across the artifact to select phrases or sentences. Once they are pleased with their selection, they can press a software button on a user interface (UI) panel that automatically appears to the left of a document when a controller ray is intersecting the document, as seen in figure 3.1. This button makes the selection highlighted. This supports the fourth item on the identified user tasks list, as highlights can quickly identify direct quotes that the user can integrate into their arguments.

3.3.5 Text Copying

In addition to highlighting, a user often takes direct quotes as support for their arguments in deliverables such as reports or presentations. Our solution for this was to enable quick copying of given selections of text documents via the operating system’s clipboard. Once a user selects a given passage as described above, they can then use one of two software buttons on the UI panel. The first copies only the text selected into the copy buffer, while the second copies the selected text and adds a citation of the source. For example, if text was copied with citation from the artifact seen in figure 3.1, the selected text would have “-citizen/b_12_0463” appended and then put in the copy buffer. These interactions support the fourth item on the identified user tasks list, so users can use direct quotes and excerpts as evidence. This interaction was used for the studies with keyboard input detailed in sections 3.4 and 5.2

3.3.6 Document Copying

Often during sensemaking, a single piece of evidence supports multiple arguments. To address this in IST, we allowed for documents to be copied through pressing a software button the the document’s UI panel. This created a copy of that document that is immediately “grabbed” by the user, so the user can put the copy in a different place. This is also a part of organization, and supports the third item on the identified user tasks list. However, after we explored user strategies in section 4.1, we found this feature was mostly unused and therefore removed it for the following studies.
3.3. Interaction Methods

Figure 3.2: The keyboard that participants used to enter text in IST was on a tracked wheeled desk, as seen in the image on the left. Once integrated into IST, our initial design was to represent the keyboard with a desk with a white keyboard in the same position as the physical keyboard, as seen in the center image. Feedback from pilot studies evolved our design, leading to a pass-through “Desk Portal” design as seen in the image on the right. Furthermore, the IST menu can be seen above the portal. These buttons are general purpose buttons that assist the user with the IST environment. In this implementation, the user can save their layout, create a new note, create a new label, or search the artifact dataset.

3.3.7 Keyboard Input

While we wanted to get direct input in earlier versions of IST, it was clear from participant feedback and interviews that they would prefer to input their own annotations into IST directly rather than through a Wizard-of-Oz technique. Furthermore, several participants wanted to start actually writing their own essay during the study in section 4.1. Both of these actions required that users have access to text-based input methods.

We implemented user-controlled text entry through putting a keyboard on a tracked, wheeled desk. The representation of this desk underwent several revisions, as seen in figure 3.2. Initially, the stand was represented by a desk with a model of a keyboard on top of it in the same position and rotation as the physical keyboard. This was revised after we received overwhelmingly negative feedback from a pilot study. The homing interaction, as described by Giovannelli et al., was simply too difficult with that implementation [42]. We therefore designed and implemented a “Desk Portal” concept that affords the ability to put down and pick up 3DUI controllers and type on the keyboard quickly. The portal further affords the ability to reference real-world data artifacts and tools that can be seen on the desk. Having robust text input enables offloading cognition onto the environment in several varied ways that are detailed below.
3.3.8 Menu Interactions

Often in user interfaces, users need to interact with an overarching menu that can be accessed in the environment in general. IST is no exception, with such examples as note creation or saving their layout as potential actions the user may need to take. While these actions were initially Wizard-of-Oz’d in the exploratory study in section 4.1, we implemented a world menu affixed to the keyboard seen in figure 3.2. This menu was used to save layouts, create notes, create labels, and search the artifact dataset for the study detailed in section 3.4. We added a “Create Report” button for the study in section 5.2, which was a variant of the create note button.

3.3.9 Note Taking

It is difficult to remember every insight one finds while performing sensemaking. We wanted to support this by allowing users to offload cognition onto the environment through notes. Initially, these notes were activated through a software button on the document’s UI panel, and what the user said was recorded through a Wizard-of-Oz approach where an experimenter manually types in the dictated note. This was used for the first exploratory study described in section 4.1. After adding keyboard input, this was done through the keyboard implementation described above, using a virtual edit field that appeared above the keyboard. Also, we added the ability to create notes that weren’t attached to artifacts through the environment menu attached to the keyboard. These were used in the second two studies in sections 3.4 and 5.2. Externalizing memory and offloading cognition onto the environment through notes supports the fourth and fifth items on the user tasks list.

3.3.10 Label Creation

Similarly, labels can offload cognition to the environment through a key word or phrase that represents a theme or idea. Just like note taking, it was initially performed through a Wizard-of-Oz technique for the exploratory study. The user requested a label and what they wanted on it, and the experimenter created the label one meter in front of the user. For the other two studies, we implemented a create label button on the environment menu that created a label for the user. They were then free to edit it using the keyboard edit field. Creating labels is another form of externalizing memory, which addresses the fifth item on the identified user tasks list.
3.3. INTERACTION METHODS

3.3.11 Keyword Search

Often during sensemaking tasks, users needs to find certain names, words, phrases, or other kinds of alphanumeric strings in the entire dataset. We therefore designed a searching mechanism so users could easily search the entire corpus quickly. For the VR exploratory study, we attached this interaction through labels, allowing the user to search for the string the label specified. Each label had their own UI panel with a button for that feature, as seen in figure 3.3. However, after we added keyboard input for the other two studies, we moved this feature to the environment menu. That way, the user could type the string they wanted to search for instead of having to create a label every time they wanted to perform a search.

For all searches, all documents with the desired string’s input have their title bar highlighted, and the text highlighted in red. This allows the user to quickly identify the documents that contain the desired word or phrase, as well as the context the search string appears in. This assists with the foraging process, and addresses the second item on the identified user tasks list.

3.3.12 Save & Load

Once a user lays out documents and wants to stop the program - either because they have finished their task or to rest and return later - we wanted them to be able to save their layout for future usage. This also requires a way to load their layouts at a later time, so we implemented both features into the approach. This allows users to quickly return to a previous state or save a state that contains particular insights that need more study. Furthermore, users can create different layouts to present their findings in 3D immersive space to clients. This addresses the sixth item on the identified user tasks list.
CHAPTER 3. DESIGN OF THE IMMERSIVE SPACE TO THINK

3.3.13 PDF Viewing

The user can also view PDFs while using IST. These act like linked images where each image represents one PDF page and only one page is available at a time. Users can switch between pages through swiping left or right on the trackpad to go to the previous or next page respectively. An example of a PDF object can be seen in figure 3.4.

3.3.14 Video Playback

Videos are displayed with their starting frame visible, but playback is paused. Users can point the controller ray at a video and press the secondary action button on the controller to play or pause the video. Users can also swipe left or right on the trackpad to rewind or skip fifteen seconds in the playback. Clearly, multimedia support addresses the first item on the identified user task list.
3.4 Different Realities: A Comparison of Augmented and Virtual Reality for Sensemaking Tasks

In initial explorations of IST, participants commented that traditional tools or methods were not supported in the digital realm, such as whiteboards to create charts or other physical tools [75, 76]. While these could be recreated in the virtual environment, they would likely have a lower-quality user experience compared to their real-world counterparts. Furthermore, participants stated that hearing other people move around in the same physical office space but being unable to see them due to using VR reduced comfort levels. This experiential data suggests that AR would be a better fit for this type of immersive analytics application or approach, as it affords a view of the real world and use of real-world objects and tools while still providing the same interaction methods with virtual data artifacts.

As Milgram and Kishino stated in their Reality-Virtuality continuum, designers can vary the level of mixed reality of an application by changing the ratio of real to virtual imagery [83]. Designers could choose a purely virtual environment, include more real-world elements to make it an augmented virtuality experience, or sprinkle in virtual objects to the real world to create a more traditional AR experience. The designer must then consider what tradeoffs each style gives their application. For example, in VR users are secluded from the real world; they can only attain visual information the designer wants them to know. This could help them focus and not be distracted by objects in the real world. However, this removes some benefits of the traditional AR affordances, such as the ability to use non-digital objects like a pen and paper or understand physical obstacles around them. Another question that arises from this continuum when considering immersive analytics research is where on the continuum would best support understanding large complex datasets [107]. To improve IST and other sensemaking approaches, we should explore how moving away from virtual environments to a more AR environment can affect sensemaking performance and analysis.

To look into this, we re-implemented IST in AR and an augmented virtuality form to gauge the impact on the user experience and performance. While AR allows users to freely interact with all real-world objects, it can create a more visually cluttered workspace that can distract users from their given tasks. Augmented virtuality can give users a space that is cleaner and can keep their focus on the task, but each physical tool that users may want to interact with has to be registered by the system in some way. As IST affords users the ability to move data artifacts around in virtual space, we designed and performed two studies with a historical analysis task and a text-based dataset to evaluate how each adjusted approach can leverage the real world while still providing a place for users to connect with and focus on their work.

We found various user experience challenges in both studies. For a pilot study ($N = 21$), we found users vastly preferring the AR implementation (20 of 21 participants), but every user found issue with the text-input methods in the VR experience. After adjustments, where we added more augmented virtuality in the form of a “Desk Portal” that allowed participants
to see a real-world tracked desk with accompanying keyboard for text-entry, user perception nearly flipped in the second study ($N = 16$) to prefer the VR implementation (9 of 16 participants). However, in contrast, users felt physically safer in the AR implementation. They also found AR more attractive and dependable than the VR implementation as well, despite the increased focus.

Our data would suggest, then, that the IST approach as well as other immersive analytics applications would benefit from a middle-ground augmented virtuality approach with more elements of the real-world augmenting the virtual reality that assists with focus. That way, users can specify which tools they want to track and/or use while performing sensemaking tasks, occasionally referencing the real world when needed, and remain focused on their dataset and goals.

### 3.4.1 Research Questions

**RQ2.1: Between AR and VR, which level of mixed reality sensemaking offers a higher quality user experience?** Previous studies’ qualitative feedback suggested that affording users the ability to see the real world would increase user experience and satisfaction [76], but this effect has not yet been verified empirically. We hypothesize that seeing the real world will increase user satisfaction, but also distract the user from performing their tasks. We further hypothesize that users will prefer the AR implementation over the VR for sensemaking tasks, as they will feel physically safer and can use various tools that could assist them in their analysis.

**RQ2.2: How do the different styles of mixed reality affect user strategies?** We saw in previous studies [75, 76] that in VR users created different structures using various methods of annotation, with some being anchored on landmarks seen in the environment and others in semicircular or planar patterns. We want to know if this behavior changes in AR. For example, do users leverage real-world landmarks for the placement of documents, or perhaps do they continue to use semicircular patterns in free space? Will users branch out and use everything they can see at their disposal to understand the datasets? This will impact how the approach can integrate the environment into assisting the user. We hypothesize that users will use tools in the AR implementation, as well as use more varied landmarks to sort their data in AR.

**RQ2.3: Between AR and VR, which level of mixed reality affords higher performance?** We want to understand how varying the style of MR affects how well participants can perform sensemaking analysis tasks. Since users can see and interact with physical real-world tools, they have more ability to use methods that are familiar to them, which may increase performance. However, we understand that there are many confounding factors that
3.4. DIFFERENT REALITIES: A COMPARISON OF AUGMENTED AND VIRTUAL REALITY FOR SENSEMAKING TASKS

Figure 3.5: On the left is the physical space that users work in. When performing tasks in VR, they see the virtual environment to the right. The environment has a floor-with-railings setup so users know where they can travel safely when considering physical obstacles and the tether to the XR-3.

can affect performance, such as a lack of sleep or even hunger levels and due to these factors technology choice may have little impact on performance [7]. As this is a multi-session study where participants would come in on consecutive days, we could not control for these factors. Therefore, our hypothesis is that we will not find any significance in overall performance.

3.4.2 Experiment Design

Apparatus & Experimental Setting

In order to accommodate an augmented reality version, we shifted from an untethered HWD to a tethered setup that can display both AR and VR. This allows us to further control for confounding variables such as a lower field of view on a typical AR HWD like the Microsoft HoloLens than most VR HMDs. The Varjo XR-3 is a high resolution HWD that can activate its forward facing cameras in order to present the user with a pass-through AR experience. It includes a five meter tether that affords the user reasonable area that they can use to interact with objects.

The physical environment was a computer lab open office space with a large four by eight meter central area devoted to MR tracking. This area had a SteamVR Lighthouse 2.0 tracking system. For the AR condition, a whiteboard with dry erase markers was placed within the tracked area along one of the four meter edges. The physical environment can be seen in figure 3.5. The XR-3 was running on a desktop PC with an Intel Core i9-9800 processor and an NVIDIA GTX 2080 graphics card. The participants also used a single Valve Index controller to move artifacts around the virtual environment. Lastly, participants
had a Logitech G780 wireless keyboard with number pad on a wheeled standing desk that was tracked with an HTC Vive Tracker 2.0, as seen in figure 3.2.

In VR, the environment had a 3 by 6 meter floor and railing model as seen in figure 3.5. This represented the area that the participant could freely move in and was kept clear of all physical obstacles with the exception of the tracked keyboard desk. The railings represented the general limits of the tether. Outside one of the long edges was a bulletin board where all artifacts initially appear. In AR, the floor and railing model and the bulletin board are not represented, though the artifacts appear in the same location in the space as in the VR condition.

Experimental Tasks

Our goal with this study was to better understand the tradeoffs between AR and VR for sensemaking tasks. To that end, we designed two separate prompts for a historical analysis essay of roughly equal difficulty. These prompts were designed to be answered when paired with open-ended responses to Survey 32, which was a survey conducted by the US Military in 1942 during World War II. Survey 32 covered the topic of racial integration within the military, and had service members answer 77 multiple-choice questions and an open ended short answer. In particular, the responses shown to participants were the short answer question “Do you have any additional comments?” The survey responses are available through The American Soldier Project [43]. The prompts were:

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? 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?

2. How closely did the experience of this cross-section of soldiers reflect Marshall’s view of the army as a democratic institution? If it did not, how do the soldiers you read try to make peace with military regimentation? How, if at all, did the race of the soldiers who wrote these remarks influence their views of the military as an institution and of their wartime experience?

Since we wanted to simulate the sensemaking process with these tasks, these prompts required close reading of the artifacts to connect themes between responses. Furthermore, we selected fifty responses for each prompt based on five keywords that related to their prompt. Each keyword would have five responses from white soldiers and five responses from black soldiers under them. Responses to the survey were displayed as data artifacts like one seen in figure 3.1. The racial identification of the respondent was encoded on the artifact through
3.4. DIFFERENT REALITIES: A COMPARISON OF AUGMENTED AND VIRTUAL REALITY FOR
SENSEMAKING TASKS

its coloration: responses from white soldiers were on white artifacts and responses from black
soldiers were on yellow artifacts. This was done so the participants could quickly understand
a key attribute of each response.

This study was approved by the institution’s institutional review board.

Procedure

This study had four phases: a pre-study phase, two main session phases (performed in
separately scheduled sessions), and a post-experiment phase. The two main session phases
were further divided into a training phase, a study phase, and a post-study phase. These
are described in detail below.

For the pre-study phase participants were welcomed and provided with a physical copy of
the informed consent form to sign. We had the participants answer a brief background
questionnaire in order to gain some insight into their demographics as well as their past
experience with VR/AR devices. Lastly, we introduced the concepts behind IST to them.
This phase took approximately 5-10 minutes.

In each main session phase participants were first given a tutorial for the version of IST
they were using that session, as we alternated which version each participant saw first to
counterbalance the study. First, the tutorial introduced the area, whether it was the virtual
environment or the physical space, to the participants and had them move along the bound-
aries to get comfortable. Additionally, the tutorial taught them how to move artifacts, scroll
text, select text, highlight text, copy text, and annotate the artifacts. It further taught them
how to save their layout, create new labels and free-form notes, and search the document
set for keywords. This phase used an example set of artifacts that were CNN articles taken
from their website. This subphase took 10-15 minutes.

In the main session study phase the participants were given a description of the dataset
and the prompt they were going to answer, and they were asked if they had any questions.
They were then put into the same version of IST as the tutorial with the prompt’s dataset
and informed that, while there were 50 documents displayed, they only needed to use as
many as they thought were needed to answer the prompt fully. As soon as they were ready,
the experimenter started the recordings and observed their actions, giving the participants
time updates every fifteen minutes. Participants were asked to read, sort, and analyze the
documents and write the answer to the prompt in a free-form note to the best of their ability.
This subphase had 60 minutes allotted to it, but participants were allowed up to 90 minutes.

The post-study subphase had participants answer a short semi-structured interview while
still wearing the HMD. They were asked the following questions:

- What was your overall strategy for analyzing the document set? Did this change over
the course of the session?
• Please describe the spatial layout you formed this session, and how it changed over the course of the session.

• What was the transition from your IST artifact spatial layout to the essay, if any?

In the second main session, participants were also asked to describe how their strategies changed from the first session. These were designed to understand how the users utilized space to make sense of these large document sets, answering RQ2.1&2.2. After participants finished the interview, they were then asked to fill out the User Experience Questionnaire (UEQ). This phase took approximately 10 minutes.

The post-experiment phase involved a semi-structured interview asking them various questions to compare and contrast their experiences with the two versions of IST. They were asked the following questions:

• What was your overall impression of each version of the system and their ability to support analysis tasks?
• If you were writing a real paper or essay, would you want to use a system like this? Why or why not?
• What were the most useful features of the system and why?
• Were there features that you wanted that were missing from the system?
• Was there anything confusing, annoying, or difficult about completing the tasks?
• Which level of mixed reality did you feel better supported your focus?
• Which level of mixed reality did you feel more grounded in reality?
• Did you refer to or use any real-world tools in either level of mixed reality?
• Do you have any other comments?

These questions were aimed at extracting qualitative user experience feedback in order to answer RQ2.2&2.3. This interview took 10-15 minutes. All participants were compensated $50 for an expected 3-4 hours of work.

Data Collection & Measures

Data for this mixed methods study was collected in a number of ways. Each participant took a pre-study questionnaire on Google forms to gather background data, including how familiar they were with AR and VR. Log files were generated from the sessions that captured
every listed action above or in the previous study [76] the users made as well as camera, controller, and keyboard location data up to ten times a second. We made video recordings using the built-in Varjo software of exactly what the participants saw during the experiment, and during the interviews an iPhone was used to record the audio. Each participant took the User Experience Questionnaire (UEQ) after each main session using a Google form [69]. Lastly, each participants’ final layouts were saved for each session with artifact position, rotation, highlights, and annotations as well as any free-standing notes or labels. All the semi-structured interviews were recorded using Apple’s Voice Memos App running on an Apple iPhone. All data was backed up on a Google Drive.

Pilot Study

We ran a pilot study (N=19) to evaluate our experimental design. In this initial pilot, we used a 3D model of a keyboard so users could find the keyboard and type after they found the keyboard’s home row, as seen in figure 3.2. Feedback from participants revealed that the poor typing experience in VR overwhelmed other user experience feedback to the point that the interaction had to be redesigned. This led to the “Desk Portal” idea as described in 3.3.7.

Participants

For the main study, we decided to restrict participation to US Citizens. This limited the pool to people who had some background knowledge in US history, as it is required for secondary education. Furthermore, it insured participants would have a higher chance of understanding the racial tensions in US society, which was relevant to the dataset and prompts we were using. Therefore we recruited undergraduate and graduate students who were US Citizens and were currently taking a history or human-computer interaction course. Seventeen participants were recruited, but one did not attend the second session. The remaining sixteen (3 female) had a mean age of 21.1 with a standard deviation of 2.02.

3.4.3 Results & Discussion

To assess user experience (RQ2.1), we analyzed the results of the UEQ as well as responses to the post-experiment semi-structured interview, as seen in section 3.4.3. We identified user strategies (RQ2.2) by analyzing what participants did in both AR and VR and their responses to the post-session semi-structured interviews, which can be seen in section 3.4.3. For measuring performance (RQ2.3), we hired four experts in historical analysis to grade each essay and performed statistical analysis, which can be seen in section 3.4.3. In all plots, significance is represented by asterisks: a single asterisk represents $p < 0.05$, double asterisks represent $p < 0.01$, and triple asterisks represent $p < 0.001$. 

**Historical Analysis Performance**

To address RQ2.3’s goal of measuring performance on the historical analysis essay-writing task, we hired four history and social studies graduate students who had experience grading college-level history papers. Each of them graded all the essays from both the AR and VR sessions based on a rubric with five ratings on a 0-10 scale. These ratings were:

- **R1**: How evident is it that they have read individual documents?
- **R2**: How evident is it that they have read groups of documents?
- **R3**: Is there a clear organizational scheme?
- **R4**: Have they identified ambiguities and contradictions where they show nuance and understanding of the topic? Do they have an argument or thesis?
- **R5**: What is the overall quality of their answer?

We performed an intraclass correlation (ICC) test to determine if the expert ratings were consistent. Using a one-way mixed effects model, our ICC value for the sum of the ratings was $ICC(C, 1) = 0.647$ which indicated moderate agreement between the experts. We also checked individual ratings, and found moderate agreement for R1 ($ICC(C, 1) = 0.591$), R2 ($ICC(C, 1) = 0.569$), R3 ($ICC(C, 1) = 0.583$), and R5 ($ICC(C, 1) = 0.629$). R4 was rated as poor agreement ($ICC(C, 1) = 0.449$).

Using these grades, we ran repeated measure paired t-tests to determine if style of mixed reality had an effect on the scores. We found a significant effect of MR style on R1 score ($t(15) = 2.39, p = 0.0306$) and R2 score ($t(15) = 2.48, p = 0.0257$). In both R1 and R2, participants scored higher when using AR, as can be seen in the boxplots in figure 3.6. These indicate that it was more evident participants carefully read documents when they analyzed and wrote in the AR condition. We investigated the relationship between the use of writing tools and the scores to see if the writing tools were responsible, but our Kruskal-Wallis tests comparing tool use to R1 ($\chi^2(1) = 2.37, p = .124$) and R2 ($\chi^2(1) = 1.68, p = .196$) did not show significance. We further calculated the Cohen’s D to find the effect size of RSum as well as a post-hoc power analysis to determine the number of participants required to find significance. We found an effect size of .167 for RSum and calculated that it would take a study with $N = 284$ to find significant levels of difference between the two levels of mixed reality, which is not feasible when considering the 4 hours it takes to run a single participant.

We further performed correlation tests on many different variables that were related to the level of mixed reality or their strategy using IST. These included what style of mixed reality that was used, number of artifacts moved, mean distance each artifact moved, how far each participant moved during the session, how far the keyboard moved during the session, how many highlights, labels, or notes they used, how many times they copied text to the
3.4. DIFFERENT REALITIES: A COMPARISON OF AUGMENTED AND VIRTUAL REALITY FOR SENSEMAKING TASKS

Figure 3.6: The left boxplot shows R1 grade performance (How evident is it that the participant has read individual documents?) for AR and VR, and the right boxplot shows R2 grade performance (How evident is it that they have read groups of documents?) for AR and VR.

clipboard, if they used non-digital tools (such as the whiteboard or physical paper prompt), the layout scheme they used, whether they preferred AR or VR, whether they found AR or VR more focusing, whether they found AR or VR grounded them in reality, and the amount of time they took to answer the prompt. For correlations between categorical variables and scores, we used Kruskal-Wallis tests, while for correlations between continuous variables and scores used Spearman’s tests. There were no significant findings.

User Experience

To address RQ2.1, we analyzed the UEQ data as well as qualitative feedback from the interviews. We analyzed the effect of MR style on responses for each question in the UEQ and each compiled scale. We found participants found AR more secure (physically safe, $\chi^2(1) = 6.4, p = .0114$ AR: mean = 1.5, SD = 0.894, VR: mean = 2.5, SD = 1.03), as well as more attractive ($\chi^2(1) = 5.4, p = .0201$ AR: mean = 1.72, SD = 0.894, VR: mean = 1.4, SD = 1.06) and Dependable ($\chi^2(1) = 7.14, p = .00753$ AR: mean = 2.08, SD = 0.472, VR: mean = 1.375, SD = 0.742). Boxplots of these can be seen in figure 3.7.

These results suggest that users are more comfortable in situations where they can see their surroundings and feel safe knowing that there isn’t anything that they might hit accidentally. The dependability scale, in particular, suggests that they felt like they knew how their environment would react to their input. Combined with the significantly higher scores in R1 and R2 while in AR, as seen in section 3.4.3, could mean that their comfort and expectations
Figure 3.7: The significant UEQ results from experiment II showed that participants found AR more physically safe (Secure vs. Insecure), Dependable, and Attractive than VR.

in the system afforded them the ability to more closely read the documents and identify key passages and themes.

Qualitative feedback revealed further user experience findings. The tether of the HWD presented an issue, with three participants explicitly mentioning it. B5 stated “I didn’t want to step on the tether. I was aware of it all the time.” but noted that “I thought about the tether a lot less [in the AR condition]. I kind of just forgot about it... Being able to see the tether makes a huge difference.” This backs up the finding for question 17 of the UEQ that participants felt safer knowing where the tether was. Other participants continued the security theme in a more general way. B3 stated that AR “wasn’t like I felt in the VR yesterday, where I felt I was in danger walking around, like, when you can’t see your environment you don’t feel the inclination to walk around.”

Furthermore, two participants grabbed chairs to sit down during the session while still performing the task in AR, increasing their physical comfort. B5 also stated that “I considered grabbing a chair [in AR]. I would never have considered it in VR.”

The post-experiment interview also revealed user preferences between the two MR styles. Nine of the sixteen participants preferred VR over AR. “AR didn’t add much besides making you a little more comfortable that you’re not going to trip over something, whereas the VR had all the same features and just seemed more real,” stated B5, while B8 said “VR felt more natural and you could move around and use your whole space better.” Similarly, B4 stated that VR afforded “more blank space where I had a greater opportunity of where I could lay
3.4. DIFFERENT REALITIES: A COMPARISON OF AUGMENTED AND VIRTUAL REALITY FOR SENSEMAKING TASKS

As we hypothesized, the majority of participants found VR better for focus than AR, with thirteen of sixteen stating that there were too many distractions in AR. “Stuff in the environment ... I wasn’t consciously getting distracted by it, but it wasn’t helping my focus,” stated B3, while B12 said “I feel VR was better for trying to block everything out and focus on the task at hand instead of thinking ‘oh there’s a computer over there, I wonder what that guy is doing.’ ” In addition, observational data showed that participants would get distracted by notifications on personal devices while in AR. Participants B1 and B13 kept referencing their phone during the session in AR, but in VR B1 removed their watch to prevent further notifications as “taking off the headset was time consuming.” However, we should note that this contradicts the performance findings with respect to R1 and R2 seen in section 3.4.3, where those ratings were intended to measure the amount of focus participants placed on reading documents. This contradiction, while typical in UX research, is evidence that there’s not always a correlation between performance and perception for UX.

Twelve of sixteen thought that AR did ground users better in reality, as we hypothesized. This affected B9’s preference towards AR, saying “I can see what’s going on around me and get a sense of time ... in VR you can’t really get a sense of time ... You can be in there for hours and it’s suddenly nighttime.” Similarly, B1 enjoyed being able to use tools: “I was able to use the [whiteboard], for example ... but VR was like wandering around in the dark.”

Overall, the user experience data suggests moving towards the center of the reality-virtuality continuum. This would entail more portals to the real world, like the Desk Portal, allowing users’ to access traditional tools such as whiteboards, or a portal to see the floor of the tracked area to ensure it is free of obstacles while balancing the users’ desire to block out distractions through a virtual environment. Some physical tools could be represented by tracked objects with virtual representation, such as chairs or other furniture that doesn’t need to be seen to be used. Similarly, the UEQ metrics show that AR delivers on attractive and dependable metrics, while participants found focus and more space to organize in VR. As we already moved towards a more augmented virtuality experience for this study’s VR condition, moving further to incorporate more of the positive features of AR could benefit future immersive analytics sensemaking approaches.

**User Strategies**

We observed participants employing many different strategies between the AR and VR conditions. As Lisle et al. found in previous work on IST, users laid out documents in four identifiable ways: semicircular, environmental, planar, and none. In the AR condition, five participants used environmental layouts, eight used semicircular layouts, two used planar layouts, and one participant had a ‘none’ layout. All of the environmental layouts used the whiteboard as the environmental feature around which documents were arranged. In VR, there were seven environmental layouts, five semicircular layouts, 2 planar layouts, and 2
Figure 3.8: Left: 3D scatterplot of where B8 organized documents in AR. The red line emphasizes the semicircular layout, with the observation point being at the center of the arc. Right: B8’s layout in VR. The blue rectangle is the floor; documents are organized around three sides of the floor in an environmental layout.

‘none.’ In VR, however, the environmental layouts used the floor edges with railings and the bulletin board as the key environmental features, which led to artifacts being more spread out. As a result, we believe there were more environmental layouts used in the VR condition because there were more landmarks with blank area to do so than in AR. Luo et al. noticed a similar pattern for data artifact layout in their study, in that users in AR needed furniture and other landmarks as scaffolding for their organization [79]. It’s possible that this applies in both AR and VR, and our virtual environment simply gave participants more opportunity to do so. Some participants changed their strategies from one session to the next. B8, for example, went from using a semicircular layout in the AR session to using an environmental layout in the VR session, using three of the floor edges to organize documents. This can be seen in figure 3.8.
3.4. Different Realities: A Comparison of Augmented and Virtual Reality for Sensemaking Tasks

Figure 3.9: Participant B11 wrote down quotes and thoughts for their analysis on the whiteboard, separating thoughts from white and black soldiers by using different color markers.

In the interviews, participants explained their environmental layouts by describing them as staging areas or places with assigned meaning. “The final six documents, I placed them by the whiteboard area. When I was finally about to construct my response, I put the ones I didn’t want back on the wall,” B4 stated, with B15 more strongly saying “I grabbed [artifacts] off the wall and placed them around the whiteboard because that was my writing space.” B17 continued this idea, and said “Being able to move documents to the [whiteboard] buffer space allowed me to process them more efficiently. I had a space with meaning, and that helped me.” We term this space the “Shoebox,” based on Pirolli & Card’s sensemaking phase where data is stored for processing [94]. None of the participants reported or were observed using this strategy in VR.

Another way participants assigned roles to space was through having trash piles. Interestingly, this strategy was observed more in VR (four participants) than in AR (two participants). This may be because occlusion worked in VR and not as well in AR, as B13’s trash pile strategy was to hide the artifacts: “If I thought a document wasn’t useful I pushed it behind the bulletin board.” B8 had trash piles in both, stating “thought a document was irrelevant I stuck it in a corner.”

In addition to utilizing space, participants would use tools such as the whiteboard in AR or the physical document on which the prompts were written in both AR and VR (it was possible to see the physical document in VR by placing it on the desk and viewing through
the portal. While nine participants used the whiteboard in AR as either their shoebox area or as scaffolding for their layouts, only six participants used the whiteboard to jot down notes or ideas. B11, for example used the whiteboard to offload cognition and organize their thoughts: “I used the whiteboard to write out and separate the main ideas,” continuing “I used different color markers so I could separate the opinions of the black soldiers from the white soldiers.” This can be seen in figure 3.9. Similarly, six participants used the provided physical paper prompt and pen to write down thoughts or underline key parts of the question. This was slightly more observed in AR at three participants versus two in VR with B9 using it in both but stating that “it was a lot easier to write [on the paper] in AR, so I used a pen to jot some things down... to get me started on what to write for my essay.” Even though the participant didn’t offload their cognition onto the virtual environment, the process of writing helped them with their sensemaking.

B15 had their own strategy in the AR condition, where they read artifacts like other participants, but extracted their thoughts from artifacts and wrote those directly on the whiteboard before returning the artifact to near its original position. This resulted in a layout that had only the prompt and their essay near the whiteboard, and all of the remaining artifacts along the original plane they started on, as seen in figure 3.10, right. We defined this as having no-layout, as even though the artifacts form a rough plane it was incidental because B15 put them back as they extracted what they wanted from them. It is evident from the movement line that B15 spent a lot of time working on the whiteboard, where they created a T-chart of their thoughts and observations. A T-chart is an organizational tool used to separate information into two categories to more easily compare and contrast them. They stated that using the white board “felt quicker, I wasn’t bogged down by pressing the buttons, I could go straight into writing on the board which I’m accustomed to.” B15 repeated this strategy
3.4. DIFFERENT REALITIES: A COMPARISON OF AUGMENTED AND VIRTUAL REALITY FOR SENSEMAKING TASKS

Figure 3.11: Participant B15 utilized a strategy for analysis called a T-Chart. Using the whiteboard in the AR condition, they wrote about the varying experiences of white and black soldiers, filling the board with quotes, themes, and ideas, as seen on the left. They attempted to recreate a T-Chart in the VR condition, using the note feature, as seen on the right.

in the VR session, but had to adapt several notes to create the same structure. These T-Charts can be seen in figure 3.11. B15 expressed displeasure with their VR T-chart, saying “I wanted to do the [chart], but it wasn’t supported in VR, it wasn’t streamlined.” These observations show that familiar methods and tools are desirable, and should be supported in an approach for the sensemaking process.

3.4.4 Limitations

While we performed several iterations on the design of this experiment, there are several limitations that need to be discussed. First, while the Varjo XR-3 is certainly a state-of-the-art HWD capable of both VR and AR, the use of video-pass-through AR presents difficulties, such as incorrect ocular placement (about two inches ahead of normal eye placement). This can affect user depth perception, as seen in other studies [4]. Furthermore, as several participants noted, despite the XR-3’s high pixel density it still renders real-world objects at a lower fidelity than their eyes, causing them to be blurry. An optical see-through display, such as the Microsoft Hololens, would be more appropriate for real-world tool usage, but its relatively low field-of-view and lack of VR capabilities made it unsuitable for this study. Limiting the confounding variables allowed for a more fair comparison between the two versions of mixed reality.

The study also used novice analysts to complete the sensemaking tasks, where the pool of
participants lacked expertise in performing historical analysis. However, for the study we
did select US citizens to have a background in both essay writing and US history. While this
allowed us to gauge the effectiveness of AR and VR, recruitment of domain expert analysts
would provide a better look at the overall performance.

Lastly, we should acknowledge that this study focused on the singular experience. There are
likely differences when performing sensemaking in a collaborative setting, such as Luo et al.’s
research or many other collaborative studies [35, 79, 105]. However, we chose to focus on
the individual experience as a good starting point as many sensemaking tasks are performed
in-part or entirely as solo activities, such as literature reviews or historical analysis.

3.4.5 Conclusions

In this study, we examined how the use of different approaches to immersive sensemaking,
drawn from different points on the mixed reality continuum, affect user experience, strategy,
and performance. We found tradeoffs between VR (enhances focus and limits distraction)
and AR (enhances comfort and affords the use of physical tools).

Small but meaningful changes in user experience design can result in different outcomes
when comparing AR to VR. The enhancements we made to improve keyboard usability in
VR through an augmented virtuality portal proved to be effective, but we propose that
future immersive analytics approaches expand this approach further. Environments should
retain most of the virtual setting to keep the focus on the data but open more portals to the
real world to enable the use of more traditional tools or viewing the tracked area for obsta-
cles. The process of using these non-digital tools can enhance the sensemaking process such
that the user’s understanding grows and they can write reports more easily. Furthermore,
the virtual environment can provide scaffolding easily for organization of documents, a phe-
nomenon we observed in our study and in a past study in VR [76]. This is more difficult in
an AR setting, as Luo et al. discussed how different furniture had different patterns of usage
[79]. In an augmented virtuality setting, varied workspaces can be created easily by adding
object-structures that provide the same affordances, such as a big cabinet, whiteboard, or
office chair.

3.5 Chapter Summary

The design of IST evolved over time as we gathered more data from both expert analysts
and experimental feedback. We have contributed a set of necessary user interactions for
non-quantitative datasets that assist and support the sensemaking process. In particular,
users need the ability to move artifacts around the environment, search a given dataset,
and offload cognition onto the environment in various ways. To support that, we designed
text input methods, including an augmented virtuality “Desk Portal” to afford the ability
to interact with real-world keyboards, tools, as well as the ability to put down controllers when needed. We further evaluated AR and VR to better understand the tradeoffs of each version of mixed reality for sensemaking tasks, and found that a blend of the two, augmented virtuality, would best support analysts. This allows users to interact with real-world tools and have enhanced comfort while still being able to focus on their particular tasks and goals.
Chapter 4

The Effect of 3D Immersive Space on Sensemaking Strategies

Immersive space and 3D user interaction techniques offer a distinct way to interact with non-quantitative data. However, we do not know what strategies users will to interact with the source documents and understand them collectively. Will they create visible structures out of documents? Will they use notes and labels as a source of external memory? This is at the heart of RQ2: How does 3D immersive space affect how users perform sensemaking tasks with non-quantitative data?

We performed a mixed-methods exploratory study to better understand how users will leverage the IST approach to understand these datasets. This study was the first main study performed in this dissertation in the Fall of 2019, using a version of IST where users could move and resize the documents. Furthermore, they could offload cognition through highlights or notes and labels created through Wizard-of-Oz dictation. Through this study we found IST was an effective sensemaking tool where users create 2 or 2.5D structures in several distinct ways. In semicircular layouts, the user arrays documents around themselves in an arc, with a single viewing point, whereas in planar layouts the user creates many flat layouts with separate viewpoints. In environmental layouts, participants leverage cues given to them in the environment as scaffolding for the arrangement of their documents in space. Additionally, we found that users had varied combinations of notes, labels, and highlights to offload cognition onto the environment.

4.1 Sensemaking Strategies with Immersive Space to Think

Sensemaking is a cognitively difficult task that involves foraging through large amounts of data to find meaningful items and inferring how those items relate to one another [94]. 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 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 the evidence between artifacts, and synthesize a story from the evidence into a coherent account [126]. 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. Our immersive analytics tool, the 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 figures 4.7 and 4.3. While studies we have already conducted suggest there is merit to our implementation [11, 75], 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 users will navigate the virtual environment and what 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 in, as physical navigation has been shown to improve recall [68]. They were also given tools such as a highlighting tool and a note tool to mark the documents or create summaries to also 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. In semi-circular arrangements, the user created document clusters that all faced a general center point so the user could stand there and observe all the documents at once. In the environment-based arrangement, users would utilize environmental cues such as a virtual floorboard or a virtual surface in order to keep the documents localized to landmarks. Planar arrangements involved creating multiple walls of documents that did not share a single viewpoint. We further analyzed how our new initial 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 of expectation of what we will see in future studies.

Understanding how users make sense of documents in a large dataset and how these documents come together to form a coherent story can help guide the development of new IST features. For example, our layout findings suggest that we might provide explicit support for common layouts, such as allowing users to create their own virtual walls, such that doc-
ments brought near the wall would automatically snap to it. Our results on document selection bias also suggest how we could tweak how the documents are displayed in order to ensure a more thorough analysis. This study establishes some strategies for analysts to make use of as well as pointing to further enhancements we can make within IST to make it a better all-around tool.

4.1.1 Experimental Design

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.

RQ2.1: 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 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.

RQ2.2: 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 we observed in previous studies and if they were effective additions to the implementation overall.

RQ2.3: 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.

RQ2.4: 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. For example, if there were frustrating elements to the design, how could we improve the design to remove these annoyances?
4.1. Sensemaking Strategies with Immersive Space to Think

Figure 4.1: 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.

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.

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. In our previous study with IST, we noticed that participants had trouble connecting with the prompts and caring about actually understanding the underlying themes in the documents [11]. We therefore decided to incorporate a new task and limit our participant pool to people who should, in theory, be more interested in the topic we were presenting them. With these restrictions in mind, we decided to use a historical analysis task paired with students currently taking a history class at our university. We selected a set of 100 responses to a survey of American Soldiers conducted by the US military during World War II. We had access to the survey responses through The American Soldier Project, which is a citizen science project that is transcribing, annotating, and tagging documents that have been digitized for the project [43]. The survey asked soldiers about race relations in the armed forces. In particular, we chose the responses of the last question in the survey, which was an open-ended prompt that allowed soldiers to expand on their answers and provide additional insights and opinions. The theme of the survey and racial identification of the
CHAPTER 4. THE EFFECT OF 3D IMMERSIVE SPACE ON SENSEMAKING STRATEGIES

responses also informed the prompts we would ask 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. We also wanted to limit their time in VR to 40 minute chunks to prevent fatigue.

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 relevant to both the dataset and the prompts. This represents a typical analysis process where the analyst prunes irrelevant documents from closer analysis. 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 3.1. 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. While the responses were anonymous, the race of the respondent was encoded in the base color of the document: we used a base color of white for white soldiers and yellow for black soldiers. In our example from figure 3.1, the keyword was “fight,” the response was from a white soldier (as seen in both the title containing “w_” and its base 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.
Participants

We recruited 24 participants for our study from an undergraduate history course on World War II. However, due to software issues related to both Microsoft XBox Game bar and hardware issues related to the HTC VIVE Pro with wireless attachment, 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.

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, explaining that we are studying how it allows users to extract concepts, themes, or other types of information from large sets of text-based documents. We told them that IST allows them to organize documents into clusters, label those clusters, take notes on the documents, highlight key words or phrases, and more. This phase typically lasted five to ten minutes.

In the training phase we introduced the participants to IST’s environment as seen in figure 4.1 and taught them how to use the controls of IST. We used a set of CNN article transcriptions so as not to reveal the main document dataset. We showed the participants the boundaries of the tracked area which is encoded in the wood flooring panels in the virtual environment. We explained how to grab the documents, move them, and resize them next, so that they can organize the documents easily. Next, we showed them how to scroll the documents and highlight text. This introduced the UI panel as seen in figure 3.1. We further showed them how to copy documents and activate the note feature. After describing the Wizard-of-Oz technique we implemented for notes, we then described the similar technique for creating labels. Lastly, we introduced the search feature on the label UI panel, as seen in figure 3.3. Once all of the UI features were explained, we then allowed 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 4.1.1. 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 rather than one long session using the save and load feature as described in section 3.3. 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:

1. How would you answer the first prompt, and what evidence would you use?
2. How would you answer the second prompt, and what evidence would you use?
3. What clusters of documents did you form, and how were they labeled?
4. What spatial relationships did you use in order to extract meaning? Does the spatial layout reflect the organization of your answer?

The first two questions were designed to reveal their performance on the experimental task through seeing their analysis and what they could infer between the documents they read. Furthermore, we didn’t expect the participants to actually write the essays we were asking them to outline, due to wanting to keep the experiment duration shorter and due to not having a way for them to directly enter text or type in IST. Therefore, these questions were a way to extract their plans for the essay without having them write it. The last two questions were aimed at learning what structures of meaning they created with the documents. For example, if a participant made one document in a cluster noticeably larger, we wanted to know if this was due to its relative importance or due to some other factor. This phase took between ten and fifteen minutes.

Lastly, we wanted some feedback about the general design of IST and how we implemented it. This resulted in us doing a post-experiment interview where we asked the participants about their experiences and what could be improved or what was good about it. We asked the following questions:

1. What was your overall impression of the system and its ability to support historical document analysis?
2. If you were writing a real historical manuscript or paper would you want to use a system like IST?
3. What were the most useful features of the system and why?
4. Did you feel that there were features that were missing and that would be helpful in this task?

5. Was there anything confusing, annoying, or difficult about using the system?

6. Do you have any other comments about using the system, or any questions about the system?

The last question, in particular, aimed at getting a dialogue between the participant and experimenter to discuss further areas of improvement that weren’t addressed by the prior questions. This phase took between five and ten minutes.

**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 mic. 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 whenever the program was launched. These included camera and controller positional data and user actions as described in section 3.3. 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 the save feature we described earlier. This allows us to see the final clusters or structures of meaning that each participant formed during the experiment and see any similarities between participants. Furthermore, this allows 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. This was positioned between the experimenter and the participant during the post-experiment interview so that it could fully capture both voices. These interviews were also transcribed for further analysis.

4.1.2 Results & Discussion

To address RQ2.1 and RQ2.2, we analyzed how our participants moved and interacted with documents in IST (section 4.1.2), how they used the immersive space to organize documents during their analysis process (section 4.1.2), and how they used the new markup and note-taking features (section 4.1.2). For RQ2.3, we analyzed the analytic performance of our participants and attempted to correlate this performance with different IST usage strategies (section 4.1.2). Finally, for RQ2.4 we analyzed the results of our post-experiment interview (section 4.1.2). In all plots, significance is represented by asterisks: a single asterisk represents $p < 0.05$, double asterisks represent $p < 0.01$, and triple asterisks represent $p < 0.001$.

Participant & Document Movement Patterns

We tracked movement data of both the participants and the documents they used through the log files generated by IST, as can be seen in the aggregate plot in figure 4.2. Participants traveled a mean of 475.2 meters (standard deviation of 259.4) while performing the main task, with a minimum of 175.3 meters and a maximum of 1125.6 meters. However, there were only three complaints about fatigue in the post-experiment interview, and all three were related to eye strain. This, combined with a mean main task and in-VR interview completion time of 51.7 minutes (standard deviation of 11.4) was surprising to us, as often people get fatigued or affected by simulator sickness while experiencing a virtual environment.
for that long [96]. Furthermore, the participant movement map seen in figure 4.2 also shows that, while participants did utilize a lot of the tracked area, they tended to stay generally in the center of the area, with a slight bias to the right side. We hypothesize that this is due to it being the starting position that they are put into IST. Therefore, as they get more comfortable being in the virtual environment they spread out more, but initially they tend to stand in the same small region. In addition, just beyond the “floor” were a set of physical tables. While the participants were informed that the virtual floor was kept free of obstacles, they may have recalled the tables and kept their distance from the edges of the floor for that reason. The participants still used a significant portion of tracked area.

We observed a pattern of interaction that most participants worked through during the experiment. Participants would choose a document and bring it towards them and resize it for ease of reading. Then, after a few moments of reading the document, they would either highlight a section or take a note to remember key sections of the document. At this point, they would typically move the document into either a temporary or final position as they went to another document. While this may seem fairly intuitive, we had actually designed the initial size of the documents to be readable from an intermediate distance so that participants could read them without moving them at first.

One surprising interaction was that eight participants put documents back on the bulletin board if they didn’t need them. This also resulted in some amount of frustration, as the bulletin board allowed documents to pass through it, and documents would get lost behind it accidentally. A couple of participants used this as a way to “trash” documents they didn’t need, though documents deemed unnecessary were dealt with in a couple of different ways. Seven participants had specific clusters that they devoted to this kind of document. They would put them there as a signal that they had already read those documents but that they weren’t useful to them. Four of these participants mentioned specifically that they didn’t want to delete them in case they became relevant through more analysis. In contrast, one participant asked for the files to be deleted altogether. In the post-experiment interview, several participants discussed ways of handling trash documents as well, which we cover in section 4.1.2.

Clusters & Final Layouts

In our study, 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 4.4. 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. 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 subclusters. Participant 4’s semicircular structure can be seen in figures 4.7 and 4.6. They had stacked documents layered 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. For 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 4.5. Seven participants in all used these environmental cues as ways of organizing documents. This, combined with the feedback given in the post-experiment interview, suggest that users want some sort of existing structure to build upon. For example,
Figure 4.4: A heatmap 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 bottom of the heatmap) and the edges of the virtual floor (represented by the rectangular outline). Both the environmental and semicircular layouts can be seen in this aggregate data, with a rough semicircle around $(0, 0)$ and higher numbers of documents near the edges.

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 and they further wanted the floor and existing bulletin boards as impenetrable objects so that documents couldn’t pass through them. This is 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 used 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. How-
CHAPTER 4. THE EFFECT OF 3D IMMERSIVE SPACE ON SENSEMAKING STRATEGIES

Figure 4.5: Participant 3’s final layout of documents, with a first-person view on the left and a birds-eye view on the right. The birds-eye view is annotated with arrows to indicate the direction the cluster is facing. This participant used an *environmental layout*, where the documents all conform to an environmental cue. In this case, they are aligned to the edges of the virtual environment’s floor.

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 our previous study using the initial prototype of IST [11], we observed layouts with both similarities and differences compared to those seen in this study. An arrangement similar to the semicircular arrangement was seen, though the radii of those layouts were often much smaller. The environmental layout wasn’t possible with that version, either, as there were no objects in the environment other than the documents. The previous study also observed “wall” layouts, which are similar to the planar layouts but contained to a single plane due to the smaller tracked area.

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; users wanted more than a phrase to describe a theme.

**Notes, Highlights & Labels**

Notes, highlights, and labels saw widespread usage. Among the seventeen participants there were 164 notes, 41 labels, and 200 highlights created, for means of 9.65, 2.41, and 11.8 respectively per participant. However, it seemed that each participant had a preference of which tool to use. Five participants had fewer than five notes, while seven participants
had no labels, and nine participants had fewer than six highlights. Furthermore, only two participants had above average numbers of both notes and labels.

Notes were typically used to summarize documents. These summaries were typically short, such as “unity” or “Fight for freedom,” though participants 2 and 3 had 2-5 sentences for each of their combined 37 notes. One common non-summary technique was to use the note section for organization. Participants 16, 20, and 22, for example, used documents’ note sections as labels for where in their essay they’d use it as evidence. This would be done with an alphanumeric, such as Participant 22’s note “B6” on document 11_0034 to indicate it was the sixth quote from a black soldier they wanted to use. Another unique usage was seen in the note section of the prompt. Seven participants used that note section to write an outline of their essay, with varying levels of structure. This was a behavior we hadn’t
expected, and often their outlines would exceed the length of the visible note object, with the remaining text of their outline floating beneath it.

Labels were used in various ways to organize other documents. One way they were used was to denote themes directly from the prompt. For example, “democracy” was referenced three times in labels, while “principles” was used four times. Another common usage was for the labels to create structure for their essay. Participants 20 and 22 used a combination of physical space and labels to organize their quotes and themes for their answers. This can be seen in figure 4.8. Otherwise, labels were used to identify themes discovered during the sensemaking process, such as “Discrimination vs. Inconvenience” or “Complaints about Race.”

Highlights were used, as we expected, to identify key themes or quotes quickly. Four participants specifically stated that they used highlighting to get quotes for their essay outline. This was also seen in the in-VR interview: when asked for direct evidence, participants with highlights could quickly pull out a quote to read as proof of their claim.

**Historical Analysis Performance**

As we stated for RQ2.3, we wanted to understand whether there were any particularly effective strategies for analyzing documents in IST. To evaluate how well each participant performed, we designed a four-question Likert survey. The questions were:

- **R1.** How closely is the participant reading the documentary historical evidence?
- **R2.** How comprehensive is the historical analysis?
- **R3.** How organized is the historical evidence?
- **R4.** How well did the participant interpret and analyze the historical documents?
4.1. Sensemaking Strategies with Immersive Space to Think

Figure 4.8: Example of Participant 20’s emphasis on paper structure. On the left, they have black respondents’ views on the war. On the right, they have the white respondents’ views on the war. The sides have labels for pragmatism or principle, and many notes are annotated for the order they want the quotes to appear in the paper. In the center of the arc is the prompt, where they could reference it quickly.

Each question was answered on a five-point scale. We recruited nine graduate students without prior knowledge of the IST results to serve as raters. We trained these raters by giving them a written rubric describing characteristics of good historical analysis for each question, and by showing them video examples of good analysis practices for each question. Three raters evaluated each participant by watching video captures of participants’ in-VR interviews and then answering the survey questions.

We then performed an inter-rater reliability test to assess if the evaluations were homogeneous. Unfortunately, we found that our raters did not provide consistent ratings. We performed an intraclass correlation (ICC) test. Using a one-way mixed effects model, our ICC value for the sum of the ratings was $ICC(C, 1) = 0.485$. We also checked the individual ratings, and got $ICC(C, 1) = 0.469, 0.301, 0.228$, and 0.223 for questions R1, R2, R3, and R4 respectively. Therefore, the consistency between judges was poor. Although we report some results below correlating analysis strategy with performance rating, we caution that these results are not fully reliable and should be taken with a grain of salt.

We performed correlation tests on many different variables to see how they impacted the scores from the raters. These included layout type, number of documents that were 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 paper
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 [22].

The strongest correlation we found was between the response to question R1 and the layout style the participant chose (ANOVA, $p = .022$). Total score was also weakly correlated with layout (ANOVA, $p = .051$). In particular, the semicircular layouts tended to get higher rating scores than the environmental layouts. If this correlation is valid, we believe this could be due to the semicircular layout affording the ability to see all documents at a glance, and requiring only head movements to access different clusters of documents arranged by theme.

We found a weak correlation (Pearson correlation, $p = .077$) between total document distance traveled and R1. The trend indicates that the less a participant moved documents, the better their rating for close reading of the documents and recall of document contents. Finally, we found a weak correlation between whether a participant used notes or highlights and R4 (ANOVA, $p = .094$). In this correlation, a participant performed better on question 4 if they used both notes and highlights, compared to just one or the other.

In future studies, we will focus more on the rigorous evaluation of analytic performance. However, we believe there is no one perfect strategy for doing analysis in IST, and that analyst preferences and tasks may lead to many successful strategies for sensemaking in immersive space.

Post-Experiment Feedback

The post-experiment interview asked the participants how they felt about the IST approach, our implementation, and any suggestions they had for its design. Fifteen of the seventeen participants would have liked to use IST to write future papers or essays, though three of those were conditional upon more practice. These participants particularly liked the ability to concentrate through the virtual environment removing distractions. Participant 1 said “I felt like I had my own space there, to really work and just, 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 specifically called out liking the ability to enlarge documents. This could explain why we found relatively little fatigue or eye strain, since we had a readily available tool that could ameliorate some discomfort. 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. I really liked the notes feature.” This points to the importance of being able to offload cognition and memory onto the environment through note-taking.
Four 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, scrolling through, 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.

The participants had quite a bit to suggest regarding possible avenues of improvement. The most popular suggestion (made by four participants) was to add some form of text entry to IST so that users wouldn’t need to rely on an intermediary to type notes or create labels. While this had already been a planned feature, this sentiment confirms the necessity to add text entry to IST. Another feature that was requested was the ability to create their own bulletin boards. Users who requested this feature also said that documents should not be able to pass through bulletin boards or the floor. Participant 16 went so far as to suggest that this would be a way to move entire clusters at once by selecting and moving the bulletin board instead of the individual documents.

Two participants suggested that we have a method of supporting multimedia documents in IST. “If I was working with documents that were also linked to audio, that were also linked to video, and all sorts of things, I could have all of that in the environment and just point to the audio file and listen to it and take notes on it.” The ability to support different document types is something we had already added (although unused in this study) ([75]), but this is further evidence multimedia support is desired.

### 4.1.3 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 tracked area a few times, 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. We can take more care in the future to better calibrate the area and test it before each participant to ensure it works well.

A key limitation of this study is that we are studying novice users. While novices can reveal some flaws in designs quickly, they likely have different methods and results as compared to expert analysts. However, experts are either expensive or hard to find, which is why we
chose the population group that we had. We endeavor to have a smaller-scale study that can look at expert usage in the future.

4.2 Chapter Summary

Immersive space affects how users interact with data artifacts and the environment in terms of the usage of space, the strategies used, and how participants offloaded cognition onto the environment. We found participants traveled the entire tracked area, as well as using large portions of it to organize the data artifacts and their thoughts through notes and labels. This indicates that the approach should afford users as much area as possible. Furthermore, we found participants creating three main structures: semicircular, planar, and environmental layouts. Future implementations should support these layouts. Participants also had different ways of offloading cognition onto the environment, with notes, labels, and highlights used in varying amounts. There was no significant change in performance between what strategy they used for offloading cognition, which would indicate that there’s no correct way to perform the action. We should, therefore, continue to support as many ways as possible in order to support a myriad of different strategies for performing the sensemaking process.
Chapter 5

Comparison of IST to Other Sensemaking Tools

One of the most persistent questions in the design of a tool is “so what?” Does this tool do what its designed to do, better than other comparable tools? IST has been shown to be an effective tool through the previous two main studies, but we need to determine how immersive space and 3D user interaction techniques compare to traditional 2D tools. RQ3 aims at directly answering this question, and therefore, IST’s key aspects need to be compared to other tools that also purport to assist with the sensemaking process. In order to better evaluate RQ3, we performed a preliminary study and a more comprehensive study.

In the preliminary study, we had a single participant perform three sensemaking tasks: one with a traditional single screen desktop computer, one with the VR variant of IST, and one self-exploration sensemaking process with the VR variant of IST. This study, performed in the Fall of 2019, used a version of IST where 100 documents from our Survey 32 dataset were displayed in VR, and the participant could move, resize, and annotate the document (using Wizard-of-Oz style dictation). We found that IST produced a more comprehensive sensemaking deliverable than the desktop tool, but with a single participant we can not come to any definitive conclusions.

In the comprehensive within-subjects study, we had twelve novice participants perform three intelligence analyst tasks over the course of three sessions. This was the third main study, and was performed in the Summer of 2022. Participants used a virtual small monitor, a virtual large monitor, and a seated version of IST. In the seated-IST condition, participants would use a special file browser viewed above their keyboard to preview documents before opening them in the immersive space. In addition, we used both a small and large virtual monitor to understand the comparison between both the traditional desktop experience where most users have a 24-inch two-megapixel display [1], as well as the comparison between Andrews & North’s Space to Think approach and IST [8]. We found that both the large and IST conditions were preferred over the small condition in terms of attractiveness, efficiency, stimulation, and amount of frustration. We also found that participants preferred IST over either other condition in terms of how well it supported intelligence analysis tasks. Furthermore, we found that participants were able to leverage spatial memory to find documents quickly while using IST. Lastly, we found that IST promoted more engagement with the documents over the small and large conditions.


5.1 Initial Comparative Study

Large multimedia datasets are problematic to analyze and understand thoughtfully. It could be that a history researcher wants to know more about the motivation of a small town during the Civil War and has personal letters, farm yields, and pictures of its inhabitants. Perhaps an intelligence analyst has various sources of information on a possible terrorist cell and wants to figure out what their target is through cell phone records, video recordings of meetups and bank records. In both of these cases, it requires a significant effort on the analyst or researcher’s part to understand the story behind the data. Sometimes the problem is that the analysts have too many data artifacts. In this case, powerful searching algorithms have been developed to identify documents with the highest relevance. Sun et al., for example, developed four different algorithms to search a dataset of 1000 text documents with only 34 relevant to the associated problem. Their best algorithm selected 5% of the dataset for further analysis that contained 95% of the actually relevant documents [110]. While this is impressive, the analysis of roughly 50 documents is still a difficult task for an analyst. When an analyst starts trying to make sense of a dataset, other issues arise such as how to organize the data artifacts into categories, corroborate evidence across artifacts, and synthesize a coherent account from the evidence [126]. While single data artifacts can be easily understood, they can only provide a partial understanding. Data analysts must interpret large, complex datasets to tell a complete, coherent story.

As Piroli and Card establish, sensemaking is a cognitively intensive task where an analyst takes a dataset, forages for key evidence to support a claim, and revises their claims as they understand more of the dataset [94]. While sensemaking can cover tasks like writing an essay or coordinating a schedule, a common example of meaningful sensemaking is intelligence analysis. Often, sensemaking tools are evaluated using this kind of dataset [25, 74, 100]. Counter-terrorist datasets are often large, which requires an analyst to develop critical thinking skills in order to discard artifacts with no relevant data or combine several artifacts to create one coherent plot point.

Despite the importance of sensemaking with large multimedia datasets, technology-based tools to support this process have not advanced rapidly. Common practice uses traditional desktop computers, monitors, and input devices for document analysis. While these tools can take advantage of digital features such as search, copy/paste, and expressive markup, they are fundamentally limited by available screen real estate, which only allows viewing of a few documents at a time, and which does not offer opportunities to spatially organize documents. In contrast, traditional approaches using paper documents that can be pinned up on a wall or spread out on a floor support rich spatial organization, but lack the powerful features of digital tools. Andrews et al. combined these two approaches by using large, high-resolution displays in their “Space to Think” work, which showed the potential of externalizing the analysis process into a large, interactive 2D space. We hypothesize that this approach might be even more powerful using immersive 3D technologies such as virtual reality (VR) and augmented reality (AR). VR and AR offer a rich three dimensional space that can
5.1. Initial Comparative Study

place virtual artifacts anywhere around the user. Furthermore, recent advancements in VR displays have resulted in wireless interactions that remove the cables from the HMD to their desktop counterparts. These improvements have afforded better user experiences and should be explored for sensemaking opportunities.

Most existing immersive analytics research is confined to using quantitative datasets, rather than using multimedia data. Immersive analytics enables analysts to see data visualizations in new ways that can (sometimes literally) flip the data on its head to get a different perspective. However, multimedia datasets require a different kind of sensemaking. To explore AR and VR’s potential for supporting multimedia sensemaking, we have developed the “Immersive Space to Think” (IST), which allows analysts to work in VR with text, image, video, and audio-based data artifacts. We hypothesize that IST could improve an analyst’s sensemaking process and outcomes as compared to traditional methods and tools.

The study reported in this section is a first step in comparing IST to traditional desktop sensemaking tools. We asked a single participant to perform historical interpretation tasks using a large dataset of primary source documents using both IST and common desktop tools for viewing, marking up, and note taking. The output of each task was a written essay; the essays were evaluated by three experts in historical interpretation. Results supported our hypothesis that IST would result in a better understanding of the source documents and a more coherent interpretation. While far from conclusive, these findings demonstrate the potential of the IST approach. We also discuss our observations of the sensemaking process in IST and the participant’s feedback on our tool.

5.1.1 Methods

Hypothesized Design Benefits

We designed IST to support sensemaking of complex multimedia datasets in an expressive 3D space, and we hypothesized that it would enable better sensemaking, both in terms of process and outcomes, as compared to a traditional desktop setup.

Modern desktops rely on 2D displays that are extremely limited in the amount of space provided. A single 61 cm (24 inch) diameter screen at the United States Occupational Safety and Health Administration (OSHA) minimum recommended distance of 50 cm provides a 61.7 degree field of view and field of regard. The HTC VIVE Pro, on the other hand, provides a 110 degree field of view with a field of regard of 360 degrees and as much depth as the tracked area allows. This increase in space allows for users to create much more complex document organization structures and leverage distributed cognition more easily. Although desktop monitors currently offer greater pixel density, VR users have the ability to move close to documents for readability and step back to see an overview of many documents.

Users can also leverage distributed cognition to quickly recall their thought process in IST.
Since the window layout on a desktop is limited, users will often only have one to three documents open, implying that their relative layout will likely have no additional meaning. In IST, users can create structures of meaning, where the relative spatial organization of related files denotes more information about their relationship. For example, a group of documents placed together in IST can mean that they share a common theme or support an argument denoted by a nearby label. In addition, VR users can make use of depth and document orientation in their layout, resulting in more expressive structures.

Finally, searching for specific words or phrases can be difficult if the user's dataset is spread amongst a large set of files. With IST, however, we ameliorate this through an emphasis on visual feedback. The title bar can be seen on either side of each document and is highlighted during search with a contrasting color to quickly draw the user's attention. If the desktop has a variant of this multi-file search, it typically displays a list of files with the matching string. In IST, we leverage spatial memory to allow the user to assign meaning to locations in the virtual environment.

**Experimental Design**

The goal of the study was to gather preliminary data about the effectiveness of IST compared to traditional desktop tools for sensemaking of multimedia datasets. Our participant was an undergraduate male student working for honor's credit in history; he had no prior experience with AR or VR. To address our goal, the study was split into four phases: writing the first essay using a traditional desktop method, writing the second essay using the Immersive Space to Think, writing a self-guiding topic exploration essay, and a post-study questionnaire. Since we wanted to use a historical analysis task, we selected a set of 100 responses written by white and black World War II US soldiers on a race-relations survey conducted by the US Army (Survey #32). We then generated two prompts related to this dataset:

1. In many corners of the United States, race relations were tightly regulated by custom and law into and beyond the 1940s. What impact did military service have on black service members’ views on race and race relations in the US?

2. As a whole, men who served in World War II were better educated than soldiers in previous US conflicts. How in World War II did having a higher education impact soldiers’ views on the army in relation to their own personal experience in uniform?

These prompts were designed to elicit 500-700 word essay responses within a time limit of two hours. The time limit was intended to allow for time spent weeding out irrelevant survey responses to the prompts as well as finding particularly compelling responses with supporting evidence for their answer. Both the first and second phases used the same desktop computer.

For the first phase of the study, we created a directory of the 100 survey responses split into five sub-directories that served as categories. The responses were chosen by selected
set of keywords relevant to the survey. Furthermore, these documents were in plain-text format (.txt), and the filenames had “w_” or “b_” preceding the ID number of the file to denote the response was from a white or black soldier, respectively. The participant worked at a desk with one 61 cm (24 inch) monitor with an aspect ratio of 16:9 and a display resolution of 1920x1080. The participant was instructed to use the Microsoft Notepad application to view the files and to use Microsoft Word to type their essay. They were told they could edit the files, create new directories, or move the files to different directories or sub-directories. All interactions were recorded through a Python tracking script that recorded window focus changes, files and directories being open or closed, files or directories being created, clipboard interactions, file or directory hierarchy changes, general key-logging, and window resizing. Furthermore, the screen was being video-recorded for later reference. No training was required for this phase.

After a two-week interlude, the second phase required the participant to answer the other generated prompt using IST. This first required a training period that lasted approximately 20 minutes to acclimate the participant to the features described in section 3.3; this was not considered part of the two-hour time limit. All interactions were logged in a separate text file by the IST software, as well as camera and controller positions every 100ms. Additionally, the Unity development environment was screen-recorded so we could reference actions taken by the experimenter as well as record what the participant was seeing, with audio recorded by the VIVE Pro’s built-in microphone. The document set was the same as in phase one, separated into the same categories and denoted by a blue label above the grouping. Race of the respondent was denoted by the associated artifact being white for a white soldier and light yellow for a black soldier. During this phase, the participant was given the option to have a keyboard to write the essay while sitting at a desk in the center of the tracked area. However, the participant asked if they could instead dictate their response to the experimenter, and their request was granted.

After phase two, the participant was then tasked to choose their own question that they wanted to investigate further. During the two-week break between phases, we asked the participant to find ten to twenty new sources in pdf, mp4, jpg, or txt formats as additional resources to answer their own question.

The third phase involved writing a third essay with the participant-designed prompt: “What were the lives of African American soldiers serving overseas like compared to those serving at home?” The participant was given two hours using IST to analyze sources and outline/write their essay using the original 100 sources plus their additional 12 sources. The sources they provided included four videos, three images, one PDF document, and four text files. All of these except two text files can be seen in figure 3.4.

Lastly, phase four involved asking the participant to give 1-2 paragraph answers to a series of six questions on their experiences during the study. These questions included:

1. What was your overall impression of the Immersive Space to Think and its ability to
support document analysis?

2. If you had to write a long form essay or paper, would you use IST, and why or why not?

3. Was there any tool or feature missing from IST that could have helped you perform your tasks, and could you describe them?

4. Were there any difficulties that you encountered while using IST, and what were they?

5. Did you change your workflow or thought process while using IST? If so, how?

6. Please compare and contrast the two methods of research you performed.

After all the essays were written, they were given to three experts in historical analysis (faculty members in the Department of History at Virginia Tech) for evaluation. The experts were asked to state which of the first two essays was of higher quality and why in three to four sentences. They were then asked to give a one-paragraph evaluation of the third essay. None of the experts were informed what tool was used to write each essay. It should be noted that one of the authors participated as an expert evaluator, but like the others, he was blind to the source of the essays.

5.1.2 Results & Discussion

All three experts concluded that the second essay (written with IST) showed more understanding and analysis of the source materials compared to the first essay (written using the desktop setup). Expert one stated that the essay written on the desktop, contained “too much uninformed opinion,” and that the essay written with IST was “much more grounded in the actual source material, with a number of nuanced observations that are reflective of broader historical trends.” Expert two agreed, saying that essay two was “by far the superior essay, both in terms of prose and context.” Furthermore, expert three stated that the essay written with IST “makes a concerted effort to provide a coherent analysis of the collection set, observing rhetorical patterns that run across multiple documents.” They went on to state that the participant “not only notes rhetorical themes but the relation between these patterns and their social and geographical context.”

The experts all coming to the same conclusion on the outcomes of the analytic process provides some evidence that IST can afford greater understanding of the content as well as a tool for better organization of thoughts than the traditional desktop setup. Of course, results from a single participant cannot be said to be conclusive.

Looking at the details of the analytic process can also be instructive. In IST, the participant created a dome-like structure of documents when they prepared to write the document, as seen in figure 5.1. Their point of view can be seen in figure 5.2. All of the documents they
5.1. Initial Comparative Study

Figure 5.1: The final layout of documents arranged by the participant for writing essay two (top-down view). The layout was a rough dome shape, where the participant sat in a chair in the center of the documents where they could reference documents while writing their essay.

wanted to reference formed the right side of the dome, while the left was formed by the bulletin board. If they needed to reference a document while writing, they would either turn to look at the document or bring it closer to them. Once they were finished referencing, they would return the document to the prior position. The participant kept documents that were on related topics close together in the dome, but with no clear separation of categories. The dome layout affords the ability to quickly look at any document they needed, whether to quote or reacquaint themselves with the context or meaning. Furthermore, it puts the essay writing in a prominent place at the center of all of the documents, while still keeping references visible. It should be noted that this is in stark comparison to their strategy for essay one, where they had, at most, one file open while writing and it was always partially obstructed by the essay window.

Through the log files we can observe a few notable tidbits of information. Essay one, on the desktop, took 89 minutes for the participant to write, while they spent 104 minutes in IST writing essay two. This increase in time may be due to having to dictate their answers while using IST. Interestingly, the participant only had one file open at a time in the desktop scenario, much like the original Space to Think study [8]. However, the source document was always open next to their essay document in a way that they could read either easily. This small difference could be due to our screen being a 24-inch monitor versus the 17-inch
CHAPTER 5. COMPARISON OF IST TO OTHER SENSEMAKING TOOLS

Figure 5.2: The final layout for the participant for writing essay two, as seen by the participant while they were writing the essay. From this angle, the bulletin board is seen on the left side of their view, while the right side contains documents they wanted to reference during writing. The central document in the foreground is the window they used to write the essay.

monitor used in their study. Furthermore, the participant interacted with 60 of the source documents with an average time open and in focus of 49.3 seconds while working on the desktop, while interacting with 31 documents with an average time in the center of view of 41 seconds (during foraging phase) while working in IST. However, this can be accounted for by some familiarity with the dataset, as the participant had already completed essay one using the desktop setup.

During the foraging phase with IST, the participant would form rough temporary clusters of documents that had shared meaning. For example, they created a “trash” pile of documents that didn’t add relevant content or insight. While many of these clusters did not appear in the final dome-like structure, they did indicate meaningful relationships that formed during the sensemaking process. The formation of the final dome structure started at 28 minutes into their VR session and involved seventeen of the source documents. Lastly, the participant did not make use of highlighting text, taking notes, search, or copying (text or documents) in either setup. We do not believe this will be typical behavior.

The feedback on the essay from phase three proved to be a little more mixed. Expert one stated that he liked the prompt that the participant decided to use, and thought the
5.1. Initial Comparative Study

participant did well in addressing the differences the prompt asked about, but also that he would have “liked to see some specific examples of this referenced in the essay.” Overall, however, expert one thought it was solid. Expert two continued this theme in stating that essay three “has a lot more context ... but the attempt makes the overall essay less organized.” While the response for essay three is less enthusiastic than essay two, it seems that the main problem with it, according to the experts, was that it lacked good examples. This may point to a weakness in dataset curation, as examples of black soldiers abroad could only come from the additional sources provided by the participant.

The post-study survey yielded positive feedback from the participant on our IST design. The participant started with “[IST] looks beautiful, especially the homepage. It is nice how you could organize the documents in a 360 environment.” However, they said that they wouldn’t use IST for an assignment unless they became more experienced with it. They also mentioned that the HMD became heavy after during the experiment which impacted their concentration towards the end of essays two and three. The participant also provided a few feature suggestions. “The control feels limited since you are unable to hide or minimize any documents like you could on a PC.” Instead, they “threw not useful files under the floor.” However, the latter quote proved to also be an annoyance, as they complained in the difficulties question that “if a window was moved behind a wall or under the floor, there doesn’t seem to be a way to get it back.” They also stated that the process of typing was hindered as they felt that they didn’t want to “switch between the keyboard and controllers constantly” and that, since they were dictating, it interrupted their normal flow of writing. Therefore, they came to the conclusion that “despite the innovativeness of IST [they] cannot be as efficient with it as [they are] with a computer.” The participant also stated that their process changed depending on the environment. With a laptop, they would “organize different sources into different folders and pick the useful ones,” though they admitted they didn’t do that in phase one. With IST, they “moved useful files to different locations.” This does describe their process in IST, where they formed the dome of data artifacts.

5.1.3 Limitations

Obviously, a clear limitation of this study is that there was a single participant. This resulted from an opportunity to perform a long-term project with this student in the context of other research on IST, so we designed this experiment specifically to take advantage of this opportunity. Our intention was to learn from their process to inform a more complete study in future research. Nevertheless, the insights gained from the study still indicate the potential of the IST approach.

Second, we cannot rule out the possibility that the historical analysis may have been better with IST simply because the participant used IST second, after using the desktop. While the ordering of the questions, as well as which was performed in IST, was kept from the experts, they may have deduced the ordering on their own. As one reviewer stated, “historical
5.2 Comparison of Small, Large, and Immersive Displays

Sensemaking tools have changed considerably over the past decade. We have traditional desktop scenarios involving windows, previews, search bars, and limited screen space (even when including “virtual” desktops that users can switch between), the Space to Think approach involving entire desk-spanning displays, and finally, the immersive space to think approach that allows virtual documents floating in mid-air. Yet these tools have not been directly compared against each other.

The key way our approach differs from desktops or the Space to Think approach is that IST offers 3D immersive space as well as 3D interaction techniques, as compared to the 2D displays and 2D interaction techniques. Immersive space has been shown to provide benefits to both spatial memory [55, 97] and context switching [37], and 3D interaction has been shown to provide benefits through embodied interaction [12, 27, 52]. We should directly explore how well these benefits translate to the sensemaking process.

We therefore designed and performed a 3x1 within-subjects study to evaluate the benefits of immersive space and 3D user interaction techniques on the sensemaking process. Our conditions included a small monitor condition simulating the traditional desktop scenario, a Space to Think style large monitor condition, and a seated version of IST. In particular, Space to Think has been also shown to provide some of the benefits of immersive technologies, such as assisting with spatial memory [8], so the large monitor condition should tease out differences provided by 3D interaction.
We found that users preferred having more space while performing these tasks, as well as users having greater engagement and interaction with documents as space grew. We also found evidence of users employing “working areas” in immersive space, which we define as physically separated areas that users devoted to a particular task or category, such as report writing or topic organization. In addition, we found evidence of increased spatial memory usage as space grew, with participants using spatial memory the most in a document search task in the IST condition.
Table 5.1: This table details the various file system features for the small and large conditions as well as IST.

<table>
<thead>
<tr>
<th>Feature</th>
<th>Small/Large</th>
<th>IST</th>
</tr>
</thead>
<tbody>
<tr>
<td>Display</td>
<td>AR Virtual Monitor</td>
<td>Immersive Space</td>
</tr>
<tr>
<td>Input Method</td>
<td>Mouse/Keyboard</td>
<td>6DOF Controller/Keyboard</td>
</tr>
<tr>
<td>View Files</td>
<td>Microsoft Windows File Browser</td>
<td>IST File Browser</td>
</tr>
<tr>
<td>View Documents</td>
<td>Microsoft WordPad</td>
<td>IST Data Artifact</td>
</tr>
<tr>
<td>Notes</td>
<td>Microsoft Sticky Notes</td>
<td>IST Notes and Labels</td>
</tr>
<tr>
<td>Write Report</td>
<td>Green Sticky Note</td>
<td>IST Report</td>
</tr>
<tr>
<td>Search</td>
<td>Microsoft Windows Search</td>
<td>IST Search Feature</td>
</tr>
</tbody>
</table>

5.2.1 Condition Design

IST Design Adjustments

We redesigned some of the IST interaction methods in order to have similar functionality to Windows to remove as many confounds of interface capabilities between the conditions. These are detailed below.

Seated IST Experience  The primary realization is that a desktop or laptop experience is largely a seated one; most people are at a desk and sitting while doing traditional work with a computer. Therefore, we redesigned the IST experience to utilize a seated rotating chair with a desk attachment. A picture of this desk can be seen in figure 5.3. As such, this impacted several other design choices.

Egocentric Movement of Documents  Since a seated experience is essentially a single viewpoint, we made the design decision to use the raycast and reel movement metaphor where documents move egocentrically towards or away from the user instead of the HOMER technique [17]. In addition, we implemented a maximum distance that documents can travel away from the user at 2.5 meters. This affords the ability for the user to read the documents while still providing ample space to move and organize documents. Lastly, we locked documents to consistently face the same viewpoint that is locked to where the user is at the start of the session. We chose an initial locked position so that the documents don’t visibly rotate when the user is not manipulating them, but also remove the need for the user to rotate documents themselves.

We chose to limit the 3D interactions with these documents in part to control for confounding variables and in part due to the structures and feedback we found participants generating in prior studies. First, we wanted to keep the fatigue between conditions as constant as possible between conditions, and the traditional desktop experience is seated. Furthermore, we
5.2. Comparison of Small, Large, and Immersive Displays

We created a file browser implementation of document selection. This allowed us to mimic Windows in that documents could be previewed (on the right) before opening them. It is opened via the virtual menu bar that’s attached to the desk and can be closed via the ‘X’ button in the top right corner.

Figure 5.4: We created a file browser implementation of document selection. This allowed us to mimic Windows in that documents could be previewed (on the right) before opening them. It is opened via the virtual menu bar that’s attached to the desk and can be closed via the ‘X’ button in the top right corner.

expected our tasks to take an extended amount of time: 60-90 minutes. Therefore, we chose to have participants use a seat with a desk attached to it. Furthermore, in the sensemaking strategies section 4, we found that many users created semi-circular arrangements that faced a singular viewing point, which also would support a singular seated position. Furthermore, we did not find evidence of users manipulating the yaw or roll axes intentionally while performing the sensemaking task. Therefore, we believe that limiting these does not impact the overall use of the IST approach.

**File Browser**  The Windows operating system allows users to open and close documents independently, as well as previewing those documents before they are opened. This affords the ability to close documents that the user deems irrelevant while also allowing them to retrieve them later should they change their mind. We therefore created a file browser type implementation for IST that can be seen in figure 5.4. This implementation shows the filename, an icon for the type of document shown, a preview of the document on the right, and a close ‘X’ button to exit the file browser. The user can access it via a “File Browser” button on the virtual menu bar attached to the keyboard as in prior IST implementations. As soon as the user selects a document to open, the file browser closes itself and the document appears in the same location as the button they pressed to open the document.

**Enhanced Search** Since documents are not guaranteed to be open, search needed to be redesigned to also show hidden documents that contained the search term. This matches Windows’s own file browser search bar. IST’s search function now opens IST’s file browser
CHAPTER 5. COMPARISON OF IST TO OTHER SENSEMAKING TOOLS

Figure 5.5: An example of the virtual monitor conditions shown to users. The left is the small condition, which represented a 24-inch 4:3 monitor with 1600x1200 pixels. The right is the large condition, which represented an array of 4x2 of those monitors. This resulted in an 85-inch monitor with 6400x2400 pixels.

and changes the color of the title bar for documents that contain the given search term to red. This allows the user to easily see and understand which documents they are looking for. Furthermore, the file browser will remain open while the user opens as many of the documents that contain the search term rather than automatically close.

Report Button

Due to the importance of the final report deliverable, we decided to visually encode that into the document to differentiate the final report from any notes the users creates. We therefore created a ”New Report” button on the virtual menu bar attached to the desk. This button creates an editable note that’s gray instead of the typical blue.

Small and Large Monitor Conditions

We wanted to avoid confounding variables such as HMD weight and field of view, so we elected to use AR virtual monitors, as described by Pavanatto et al. for the small and large conditions [91]. Instead of having a physical display, we simulate the display in augmented reality that is spatially registered and viewable through a HMD, but otherwise behaves in a similar manner to a real-world monitor. Examples of these can be seen in figure 5.5. The small condition was a 24-inch 4:3 monitor with 1600x1200 pixels, while the large condition was meant to represent a 4x2 array of these monitors, similar to the original Space to Think [8]. This results in a virtual monitor that is 85-inches in diagonal with 6400x2400 pixels. This size was chosen such that the monitor could display all the documents in our chosen 24-document datasets at once with minimal extra space. This monitors were placed 70 cm
from users to represent a typical seated monitor experience. A list of features given to participants and how they were implemented in each system can be seen in table 5.1.

5.2.2 Experimental Design

Research Questions & Goals

RQ3.1: How does the increase in available space affect how users process large datasets? As we increase the amount of available space, we also afford users the ability to clutter their area with more and more documents. However, that means that fewer documents will also be piled on top of each other as users organize the documents and their thoughts. We hypothesize that as the space increases, users will require less management of how that space is filled and can allocate those resources to solving the sensemaking tasks.

We hypothesize that participants will find the small condition frustrating and report higher levels of effort, mental demand, temporal demand as they have to reorganize what documents are displayed and visible repeatedly. We further hypothesize that the large condition will also have some cognitive overhead for space management, but significantly less than the small condition.

RQ3.2: As available space increases, how do users utilize the additional space? Large datasets can fill a workspace quickly. But how do users manage those spaces when they have different amounts of space? We hypothesize that we will see different strategies that users employ to manage the space in each condition, as the different amounts afford different ways of processing the data. For example, the small condition will require a lot of window management as there simply will not be enough space to look at all the data at once. How users manage to get around the disadvantages of the smaller spaces as well as how they utilize the larger spaces can affect how they perform sensemaking tasks.

We hypothesize that, since the small condition is constrained, participants will develop strategies for specifically dealing with the small amount of space. Similarly, we hypothesize that participants will form organization patterns in both the large condition and IST condition based on document placement as they are less constrained by the space. Furthermore, we hypothesize that users will leverage spatial memory more as space increases, as they can assign meaning to different locations relative to their body.

RQ3.3 How does 3D immersive space and 3D interaction methods impact user sensemaking strategies as compared to traditional 2D displays? One of the key ways IST is a novel approach to the sensemaking task is that it leverages 3D immersive space and 3D interaction methods, but we don’t know the comparative benefits of adding 3D to
the task. We want to explore if and how adding depth to document placement affects users and how the introduction of 3D input methods affects the user experience.

We hypothesize that users will have a particular focus on the depth of both documents and their notes in ways that assist their spatial memory and writing process. We further hypothesize that users will prefer the familiar mouse interface to the 3D controllers, which will affect their mental workload or frustration measures as they recall how to interact with the documents.

RQ3.4: What, if any, is the comparative difference in performance between the differing amounts of space? Naturally, the best indication of a tool’s worth is how well it performs the task it was designed for. We also recognize that there are many factors that go into the process of sensemaking, so this is the most difficult research question to answer. For example, if a participant skipped lunch or didn’t sleep well before one of the sessions, that will likely affect the outcome of the study. This is seen in many studies about learning, which is a similar process to sensemaking [104, 125].

Due to this, we hypothesize we will see no significant differences in scores. However, we also hypothesize that IST will rate higher in user experience than the large condition, which will then, in turn, rate higher than the small condition. Furthermore, we expect that measures of effort will be higher in the small condition due to a cognitive overhead in managing the placement of various windows.

Apparatus & Experimental Setting

To remove confounding variables such as HMD weight and clarity of images, all conditions used the Varjo XR-3 to view the displays. The XR-3 was running on a desktop PC with an Intel Core i9-9800 processor and an NVIDIA GTX 2080 graphics card. To accommodate the immersive nature of IST, participants sat in a rotating chair with a desk attachment such that they could easily access the keyboard at all times. This arrangement also helps with the large condition, as the arrangement of the virtual display and seated position of the participants meant that the display was larger than the XR-3’s field of view. The rotating desk meant that the participant could face any part of that display and still have easy access to typing. Furthermore, for the small and large conditions the participants were given a Logitech K380 TKL Wireless Scissor Keyboard and a Logitech Performance MX Wireless mouse to use with the Windows 10 operating system. The IST condition gave participants a Valve Index wireless controller, and a VIVE Tracker 2.0 to track the desk position. Furthermore, participants were given a Logitech K780 full size wireless scissor keyboard because they didn’t need the space for the mouse and so they could have a text-entry confirmation key that didn’t overload another key they would normally use. In our case, that was the number pad enter key.
5.2. COMPARISON OF SMALL, LARGE, AND IMMERSIVE DISPLAYS

All conditions placed the user in an AR environment so they could see the area around them like a normal open-office environment.

Experimental Tasks

The aim of this study was to better understand how users utilize the space given to them, increasing with each condition. Therefore, our datasets needed to be large enough to take up all the space available in the middle condition, ensuring that they had to make some tradeoff decisions in the two 2D conditions. To these ends, we used three datasets that were made as training sets for intelligence analysts. These were “The Sign of the Crescent” (crescent), “The Case of Wigmore Vs. Al-Qaeda” (manpad), and “Stegosaurus (Excellent Apples)” (stegosaurus). Each dataset was adjusted to be 24 documents on purpose such that the small condition could not display all the documents at once, the large condition would have difficulty displaying all documents, and IST had ample space for all the documents.

In addition to the datasets, participants were given a set of instructions on what to write for the final report. Participants were asked to spend at least 30 minutes with the dataset before submitting a final report to ensure that they were spending a sufficient amount of time making sense of the various connections between the documents. Furthermore, after consultation with professional intelligence analysts, we determined the four essential questions to intelligence analysis are who, what, when, and where. In other words:

1. Who are the actors in the plot?
2. What are they planning on doing?
3. When are they planning on executing the plot?
4. Where is the target(s) or location(s)?

These datasets were intended to be challenging and require close reading of the documents in order to connect people, places, and themes and determine what was happening. This allows us to simulate a real-world difficult sensemaking task that requires effort from our participants.

This study was approved by the Virginia Tech’s instructional review board.

Procedure

This study had four phases: a pre-study phase, three main session phases (performed in separately scheduled sessions), and a post-experiment phase. The main session phases were further divided into a training phase, a study phase, and a post-study phase. These are described in detail below.
Participants were welcomed and provided a copy of the informed consent in the pre-study phase. We further asked them to answer a brief background questionnaire to get an understanding of their past experiences with VR, some basic demographic data, and their personal monitor setup that they most frequently used. We went over the phases of the experiment with them, and, lastly, the hardware they were using in each condition. This phase took approximately 5-10 minutes to complete.

During the main phase, participants were given a tutorial for the interface and condition they were using that session. This was counterbalanced in terms of the order of what condition they saw first. Four participants used the small condition first, four used large first, and four used IST first. The dataset order, however, remained constant. Each participant was given crescent first, then manpad, then stegosaurus to solve. This order was determined by a combination of pilot participants and analysis of the length of each dataset that suggested stegosaurus took longer to solve due to its documents being longer than the other two.

The tutorial for the small and large conditions asked participants to open a folder on the desktop. Once that was opened, they were asked to open one of the text files. These would open in WordPad, which we set as the default application on Windows 10. We then instructed participants how to move and resize the window, as well as how to highlight specific passages in the document. Participants were then introduced to Microsoft Sticky Notes. This included how to create new sticky notes, how to move them and resize them, how to change the sticky note color, and how to search all sticky notes. Participants were asked to use yellow for their notes and a (single) green sticky note for their report. Lastly, participants were instructed on how to search all the text documents for keywords or phrases.

For IST, participants were instructed how to open the file browser and open the text documents. They were further shown how to move documents, scroll documents, select text, highlight text, and copy text. They were then shown how to create notes, labels, and their final report document as well as how to edit each of them with the keyboard. Lastly, the participants were introduced to the search feature and how to find both open documents as well as documents in the file system that contained the search string. Participants were given as much time to explore the features as they needed. Each tutorial subphase used an example dataset of documents of CNN articles taken from their website and took 10-15 minutes.

In the main session study phase participants were given an introduction to the dataset they were going to solve that session and asked if they had any questions. They were also instructed that their final report document had to be a minimum of 2 paragraphs that answered the essential questions listed in section 5.2.2, and to work for at least 30 minutes. This was to ensure a minimum level of effort and quality for each written report answer. In addition, participants were told they had up to 90 minutes to work on each dataset and report, and that they would be informed every 15 minutes of the remaining time left. As soon as they indicated they were ready, they were given the dataset and the experimenter started the recordings. This subphase took 40-100 minutes.
The post-study subphase was split into two parts: a semi-structured interview performed while still in AR, and then answering two questionnaires using Google Forms. The semi-structured interview aimed to understand how the participant worked through the dataset to make sense of all the various documents and how they fit together, as well as to gauge certain aspects of how the condition assisted them or performed. The questions asked were:

1. Please walk me through how you analyzed the dataset in this session.
2. How did the available space impact your analysis?
3. I’m going to ask you to find a particular document in the dataset. You may use any features available in the main phase to find it. Once found, please read the title of the document. Can you please find [description of document in the session’s dataset]?
4. Could you please describe how you located that document?
5. Could you please walk me through how you reported your findings?
6. Do you have any additional comments on the condition you experienced in this session?

In particular, the third and fourth questions were aiming at understanding how the conditions assisted users with memory and/or organization. In particular, we expect the conditions that better support spatial memory to perform better in both time to find the document as well as recollection of the document. The question was also designed in a way such that we could time how long it took for participants to find the document. After the semi-structured interview, participants doffed the XR-3 and were asked to fill out two questionnaires: the user experience question (UEQ) [69] and the NASA Task Load Index (NASA-TLX) [48]. These questionnaires were intended to evaluate the interfaces they used in each condition as well as the perceived difficulty of the task combined with the interface they used. This subphase took 10-15 minutes.

After the completion of all main sessions, participants were then given another semi-structured interview to describe their experiences with each condition in the post-experiment phase. We asked the following questions:

1. Could you please compare and contrast the three conditions you experienced over the course of this study?
2. Which condition, in your estimation, best supported the task and why? Please rank the other two as well.
3. As the amount of available space changed, how did you organize documents?
4. How, in particular, did 3D space in the IST condition impact your process?
5. Was there anything confusing, annoying, or difficult about the interface in each condition?

6. Do you have any additional comments on the study?

These questions were designed to extract user experience feedback relevant to RQ3.1 and RQ3.2. This interview took 10-15 minutes. All participants were compensated with $70 for an expected 5-6 hours of work.

Data Collection & Measures

We collected data for this mixed methods study in a number of ways. Each participant took a pre-study questionnaire on Google forms to gather background data. Log files were generated from each condition logging interactions with windows (such as grabbing or maximizing) in the small and large conditions or IST interaction methods in the IST condition. Furthermore, camera position and rotation is recorded up to ten times a second in all conditions. In the IST condition, the controller and desk positions and rotation is also recorded. Video recordings of all conditions were made using the Varjo Base software, and the semi-structured interviews were recorded using the Apple iPhone Voice Memos app. The UEQ and NASA-TLX data was recorded using Google Forms after each main session. Lastly, all final layouts were recorded - either using the sticky notes app for the small and large conditions or a special save file for the IST condition. All data was backed up on a Google Drive.

Participants

Our participant pool was restricted to Virginia Tech students who had strong English skills and were eighteen years old or older. We recruited 18 people initially through a human-computer interaction listserv. However, two of these participants did not show up to later sessions nor responded to further email queries. An additional four were paid for their time and were removed for not putting in the requisite effort of 30 minutes and/or writing less than a two paragraph report.

The remaining twelve participants (5 female) had a mean age of 23.7 and a standard deviation of 3.33. Two wore glasses, and two more wore contact lenses. All participants had prior experience with AR or VR. Two participants reported using a single laptop monitor as their personal display solution, while the remaining ten reported using two or more monitors. The most common (five) configuration was using a laptop screen and an external monitor.
5.2. COMPARISON OF SMALL, LARGE, AND IMMERSIVE DISPLAYS

5.2.3 Results & Discussion

First, we investigate the user experience data from the questionnaires and the qualitative data from the interviews to investigate RQ3.1 and 3.2. Then we move onto evaluating the task performance for each condition to address RQ3.4. From there, we move onto the various user strategies that were employed in all conditions and some that were employed in certain conditions to further address RQ3.1, 3.2, and 3.3. Lastly, we explore the results of the search task from the post-session interview to address RQ 3.1, 3.2, and 3.3. In all plots, significance is represented by asterisks: a single asterisk represents $p < 0.05$, double asterisks represent $p < 0.01$, and triple asterisks represent $p < 0.001$.

User Experience & Workload

Participant perceptions of user experience assist with our understanding of the differences between the three conditions to help answer RQ3.1. We performed ANOVAs to see the effect of condition on the aggregate category ratings in the UEQ and found significant differences in attractiveness ($F(2, 22) = 12.16, p < 0.001$), efficiency ($F(2, 22) = 10.79, p < 0.001$), stimulation ($F(2, 22) = 11.64, p < 0.001$), and novelty ($F(2, 22) = 23.55, p < 0.001$). Boxplots of these measures can be seen in figure 5.6.

For attractiveness, Tukey’s post-hoc analysis with bonferroni corrections revealed significant differences between the IST and small conditions ($p < 0.001$), and the large and small
CHAPTER 5. COMPARISON OF IST TO OTHER SENSEMAKING TOOLS

Figure 5.7: These boxplots show the NASA-TLX results for self-assessed performance and frustration. Participants found IST significantly less frustrating than the small condition.

conditions \( (p < 0.001) \). We then looked at efficiency, and Tukey’s post-hoc analysis with bonferroni corrections revealed significant differences between the IST and small conditions \( (p < 0.001) \), and the large and small conditions \( (p < 0.001) \). And lastly, for stimulation, Tukey’s post-hoc analysis with bonferroni corrections revealed significant differences between the IST and small conditions \( (p < 0.001) \), and the large and small conditions \( (p = 0.00246) \). All three of these combine to showcase how the small condition (and thus a small and constrained amount of space) is undesirable to users. However, once they have enough space to spread out in the large condition, they are able to freely move documents around and organize them with our dataset size of 24 documents.

Novelty showed more significant differences, with IST being considered more novel than either the large \( (p = 0.00197) \) or small conditions \( (p < 0.001) \) and further large being more novel than the small condition \( (p = 0.00164) \). These results would indicate that the small condition feels normal and ordinary, while both large and IST feel like newer experiences that are uncommon, and IST further separates itself from large. AR and immersive environments are still an uncommon experience, however the large condition separating itself significantly from a small display was a surprise. It would further indicate that this much seamless space without bezels or multiple monitors to users is itself novel; today’s users are used to having smaller spaces.

We also looked at the NASA-TLX measures and ran ANOVAs for how condition affected each of them. We found trends or significance in both self-assessed performance and frustration, and boxplots of these metrics can be seen in figure 5.7. Self-assessed performance was
trending towards performance \((F(2, 22) = 2.584, p = 0.0981)\). Further investigation doing Tukey’s post-hoc analysis with bonferroni corrections revealed that small and IST \((p = 0.147)\) as well as small and large \((p = 0.147)\) were trending towards significantly different. Furthermore, we found significance in the frustration metric \((F(2, 22) = 4.363, p = 0.0254)\). Tukey’s post-hoc analysis with bonferroni corrections found differences in small and IST \((p = 0.0127)\) and a weak trend between small and large \((p = 0.116)\). These results show that participants were frustrated with the lack of space in the small condition. Combining this with the self-assessed performance metric, participants found the lack of space in the small condition to affect both their frustration with solving the task as well as their ability to even solve the task, while both large and IST provided ample enough space for the task.

**Qualitative Feedback**

The interviews we performed both post-session and post-experiment were designed to get qualitative feedback from the participants on their experiences during the study.

In particular, we asked participants to rank the conditions based on how well each of them supported the experimental task; in our case, that was sensemaking for intelligence analysis. Seven of the twelve ranked IST first, followed by the large condition second, then the small condition last. Of the remainder, four ranked large first, IST second, and small last. One participant (P13) ranked them large first, small second, then IST last. Even then, the participants who ranked IST behind Large had positive things to say about IST. “[IST] had enough space to cluster different files, which was not the case for [large],” said P1, and P3 stated [IST] was just more fun. P6 went further, saying that organization was better in IST, I loved using all the space in IST, and you could put things in their own section, and categorize them and you can move the panels around as you see fit. With the monitors it was more limited, you didn’t have the comfort room. P13, despite ranking IST last, said I liked IST the most, and further went on to state that if I had one of [the other] tasks for IST, I think I would have completed it better than in either the small or large [conditions]. Based on our interview with the participant, we believe they ranked IST last because they used it with the hardest dataset of stegosaurus.

In addition to P6, the participants who rated IST over large also praised IST for assisting with their organization. P17 stated I could organize my notes a lot better, all in a grid where I could see them... but I was able to swivel away from them when I didn’t want to look at them and look at documents... that was better than minimizing the notes. P15 stated that it was better for visual learners and assisted with their spatial memory. In IST I had so much space, I’m a visual learner, and I had the space to visualize what I was working with and place the documents the way I wanted to and knew where I could find them compared to the other screens where you couldn’t put 15-20 documents on one screen with your notes. Participants felt that the amount of space and ability to freely organize documents was a particular strength of IST, as we had hoped.
However, there was considerable feedback on how to improve IST as well. Five participants raised an issue with the scrolling interaction in IST. Having a mouse with a scroller which was really sensitive [was better]. In the third dataset, each document was very long and scrolling through them [in IST] was not difficult, but time consuming, stated P1. P2 continued this feedback, stated that [IST] was difficult because the scrolling was imprecise. This feedback would suggest that we should adjust our strategy for longer documents. While scrolling may be practical for 2D based environments, other schemes such as pagination or simply larger documents may be better forms of interaction.

Three participants also raised the issue of context switching from interacting with documents to typing notes. For example, P7 stated that going from a mouse to a keyboard is much faster than going from taking the hand controller off and switching to the keyboard. I even tried typing with the controller on, to [minimize] the interruption. P13 even stated that it hampered their ability to offload cognition, saying in [IST] I felt I took notes the worst because I kept having to pick up and put down the controller in order to write down things. There are many potential ways to solve this context switch. As hand tracking becomes both more common and more precise, we could move to hand-based interaction methods which would allow for an easy transition between interacting and typing. Furthermore, with the single viewpoint offered by a seated experience, we could adopt a mouse-metaphor where the user could move the pointer with the mouse as if it was on the surface of a sphere, moving a document’s depth via the scroll wheel.

The user experience feedback we received are of issues that we can fix to improve IST further, which, for some participants, would have caused them to rank IST higher. P6 stated that I think IST would be above large with just a few minor improvements, and P10 stating the gap between large and IST is very small.

Users also discussed how IST brought the concept of depth to their process. Several participants discussed how it enabled them to focus on particular documents, with P1 stating “Since it was a 3D world, I could bring [documents] very close” and P8 going further “[3d space] helped me focus on specific documents, because I could literally push the other ones away”. P9 felt depth could also imbue meaning into certain documents: “for IST you could have things closer and farther, instead of just 2D spreading things out... Things that weren’t as important could be really small, having it further away made it obvious that it was not significant at the moment.” P13 had been skeptical, but also felt depth could have meaning “I thought moving documents back and forward was going to annoy me, but it helped me put things into reference or put things on the back burner whenever I wanted to look at them later.” Not only did it assist with focusing on particular documents, but some users felt that it simply gave them more space to organize documents. “The z-index was really efficient [in IST] because I could just move [documents] back and forward to bring them more into focus. We could fill up everything and use the z-axis to allow for more documents and opening more and connecting them,” stated P4. P17 even stated that IST changed how they worked “If I was trying to do the same thing on my own computer I would have done something similar to what I did with the large monitor, but with IST I completely changed how I operate and I
5.2. Comparison of Small, Large, and Immersive Displays

Figure 5.8: These boxplots show the results of score on each dataset (left), and score with each condition (right). There was significant differences in scoring for each dataset, but no significant differences in score for each condition.

could have a bunch of things open and it not be cluttered. I don’t like having too many things open because I don’t like clutter, but I was able to contain documents to their own space.”

3D Space provided additional ways of defining documents, both in terms of where they were placed and for when users needed focus.

Participants also discussed how IST assisted with spatial memory. P4 used their spatial memory as a key aspect of their organization “The IST was amazing because I had the 360 degree space to organize windows. I actually made a good arrangement of my windows and I knew where they were. It was really comfortable.” P15 continued this theme, stating “I used most of the 3d space in IST, spinning in the chair, opening screens as I turned around... I could find everything where I needed them to be.” Furthermore, P2 stated that IST’s organization was inherently helpful at remembering key points: “[IST] was the best because you can see all the documents together... Easy to memorize using real space, to free up memory because space is doing work [through] spatial memory.” Participants could leverage the environment itself as a way of understanding the document set as a whole.

Task Performance

Despite our hypothesis that we wouldn’t see significance in performance metrics, we evaluated each participants’ reports for accuracy in answering the 4 questions outlined in the instructions: who, what, where, and when? For the stegosaurus dataset, the when-answer was non-specific, so we adjusted the grading for that dataset to identify additional actors. Otherwise, report grade weighting was the same for each dataset.

We performed an analysis of variance test (ANOVA) to see if the condition affected report score, and found no significant effects ($F(2, 28) = 0.375, p = 0.691$). However, we performed
Figure 5.9: This set of boxplots shows the number of notes created in each condition, the number of documents that were left open at the end of the session for each condition, the time taken in each condition, and the number of grabs in each condition. There was significance found in the time taken between small and large, as well as the number of grabs between IST and large and IST and small.

an additional test to see if the dataset affected report score and found significance ($F(2, 22) = 14.88, p < 0.001$). Tukey’s post-hoc analysis with bonferroni adjustments further found significance between crescent and manpad ($p = 0.0343$), crescent and stegosaurus ($p < 0.001$), and manpad and stegosaurus ($p = 0.0104$). Boxplots of scores versus condition and score versus dataset can be seen in figure 5.8. We ran an additional two-factor ANOVA looking at the effect of dataset and condition on scoring and found no significance of main effects and no significant interaction between the two factors. However, we would like to note that on the dataset that was the most difficult in terms of scoring, stegosaurus, had the highest mean and median scores in IST over the large and small conditions.

User Strategies

Participant interaction with each condition was recorded in a series of logs for each different condition. Several of these interactions could be compared across condition, such as participant grabbing various windows or documents to move around the available space. We looked into how each condition affected the number of grabs and found significance ($F(2, 22) = 14.05, p < 0.001$). Tukey’s post-hoc analysis with bonferonni corrections revealed significance between IST and large ($p = 0.00373$), IST and small ($p < 0.001$) and a trend towards significance between small and large ($p = 0.128$), with IST having the most grabs, followed by large, then small. A boxplot of grabs can be seen in 5.9 We believe this is due to the additional space affording the ability to organize documents more. This, in turn,
5.2. **Comparison of Small, Large, and Immersive Displays**

Figure 5.10: This ridgeline plot shows the number of grabs for all users over time normalized over each session.

results in users moving documents more as they explore the best way to organize documents. Participants in IST grab and move documents more and continue to grab over time more than in either other condition. A ridgeline graph of the grabs over time in each condition can be seen in figure 5.10. This interaction indicates an amount of increased interactivity in the IST condition; participants interact with the documents more than in either other condition.

We also compared how long each participant took in each condition, and found a significant effect ($F(2, 22) = 4.366, p = 0.0253$). Tukey’s post-hoc analysis with bonferroni corrections found a significant difference between small and large ($p = 0.00991$), with participants taking significantly less time in the small condition. Combining this with the increased frustration found for the small condition implies that participants wanted to be done with the small condition as soon as they could be, resulting in shorter times.

One strategy dichotomy that we noticed participants employing was whether to keep documents open or to close them as soon as they read them. So we looked at the final layouts of each participant and counted the number of documents they left open in each condition. While we only found weak significance ($F(2, 22) = 2.677, p = 0.0911$), Tukey’s post-hoc analysis with bonferroni corrections revealed weak effects between IST and large as well as
Figure 5.11: These are various examples of strategies seen in the small condition. P8 (left) used multiple small notes next to the document they were currently reading in order to organize their notes. P9 (middle) utilized the document preview window on the file browser instead of opening files. P2 (right), like most participants, had to layer windows on top of each other in the small condition.

IST and small ($p = 0.135$ for both) and where participants kept more documents open in IST than either small or large. While further investigation is required, this implies that participants felt more comfortable leaving documents open and organized into layouts in the IST condition. This may be due to the ample space that they could work with while not impeding the view of their workspace, or an increase of engagement with documents in the IST condition.

Similar to open documents, we also looked at the number of notes generated in each condition and also found weak significance ($F(2, 22) = 3.416, p = 0.0511$). Tukey’s post-hoc analysis with bonferroni corrections revealed that this was between the large and ist conditions ($p = 0.075$) as well as the large and small conditions ($p = 0.0669$) where they made more notes in large than either other condition. A boxplot of the number of notes created can be seen in figure 5.9. We believe this is caused by a combination of factors. First, the small condition simply didn’t have enough space to support many notes, with P7 stating I didn’t have issues identify documents, it was problems with organizing them. As I had more space, I had more sticky notes. In [small] I had 1 sticky note with all my information on it. Furthermore, the large condition couldn’t support both viewing all documents simultaneously and having multiple notes. Since the sticky notes took up less space, some participants preferred them over the Wordpad windows taking up space, as P17 notes. [In the large condition] I could open multiple documents and see them all at once, but I couldn’t do that and also have my sticky notes open and arranged. Lastly, as noted in the qualitative feedback section, there were minor issues with switching between document interaction and typing that prevented more notes being written. P6 stated If it was just a little easier to make notes or labels in IST, I would prefer it, but as the interface isn’t as clear cut as a virtual desktop, I think I’d prefer the large monitor. We believe all of these factors impacted participants creating more notes in the large condition.
5.2. Comparison of Small, Large, and Immersive Displays

Figure 5.12: These are various examples of strategies seen in the large condition. P4 (left) created a note for each document in their dataset and organized them into clusters. P2 (right) kept all the documents open and arranged around the screen so they could refer to them while writing their report.

We also observed strategies that were only seen in certain conditions. Participants had to create new ways of dealing with the lack of space in the small condition. For example, five of the twelve participants did not open documents in WordPad. Instead, they elected to read the documents in the preview section of the file browser window, which allowed them to not layer as many windows on top of each other. Example of both window layering and the use of the preview window can be seen in figure 5.11. We performed a Kruskal Wallis correlation test but found no significance on if participants employed this strategy and their score.

In the large condition, two participants also used the preview feature rather than opening documents, but both of these participants saw the small condition in the previous session. We believe this was just a continuation of the strategy they learned in the previous session. There was also no significance correlation with scores and this strategy.

Three participants in the large condition kept all of the documents open and organized around the screen. An example of this can be seen in figure 5.12. This allowed participants to identify and organize the documents, with P2 saying “In [large] I would use highlights to identify documents. I also used space on the screen.”. A Kruskal-Wallis correlation test found only a weak correlation between using this strategy and score ($\text{Chisquare} = 1.92, p = 0.1655$).

As noted above, participants created more notes in the large condition than either other condition. Four participants created seven or more notes, and would arrange these notes into clusters, as seen in figure 5.12. P4 stated that “I used clusters of notes. The lower left I had the FBI reports... on the right side I had notes for the people who rented trucks... and the center cluster defines the connection with the website where the code was originally decoded.” The large condition provided enough space to spatially cluster related notes, and these
Figure 5.13: These are various examples of strategies employed in the IST condition. P13 (upper left) used timelines, and you can see they highlighted the dates in each document. P5 (upper right) used a strategy where they had different “working areas” and the area shown here is where the sorted documents, while they had other areas for their notes and a last one for their report writing. P17 (lower left) created a wall of notes that they referred to (instead of the document set) during the report writing synthesis. P1 (lower right) arranged documents in a 360-degree arc above their head.

participants leaned on this strategy. However, a Kruskall-Wallis correlation test revealed only a weak correlation between large numbers of notes and score ($Chi^2 = 1.85, p = 0.174$) in the large condition.

One strategy seen in IST and not the other two conditions was the creation and use of timelines. Five of the twelve participants used this organization in their sensemaking process - they would arrange documents either by the date of the report writing or dates mentioned within the document. For example, P4 stated that I arranged documents by dates at first, I thought that arranging by date would help me. An example of this can be seen in figure 5.13. However, a Kruskal-Wallis test revealed no significant correlation between score and utilization of this strategy.

In another vein, four of the twelve participants mentioned in the interviews that they assigned
5.2. COMPARISON OF SMALL, LARGE, AND IMMERSIVE DISPLAYS

Figure 5.14: Participants used several different strategies in order to find certain documents in the interview stage. On the left is a count of the number of times each strategy is used in each condition. On the right is a boxplot of the strategy type compared to how long it took for each strategy to find documents.

meaning to certain areas which we term “working areas.” The participants would separate in physical space areas to do specific tasks in, such as report writing, document organization, or note taking. P9 stated that this enabled them to group documents to the left and right and turn to face them in IST without having to worry about using them. Kind of like where people have 3 desks and can turn to face different topics. P13 was more explicit, saying that In the IST session, I felt that this method I had employed couldn’t be contained to monitors but having different workstations - like an information station and a chronological one and such... that’s where having AR was really impactful. A Kruskall-Wallis correlation test found weak significance between use of the working area strategy and score ($Chisquare = 3.24, p = 0.072$). This would potentially indicate that separating one’s sensemaking areas might increase the ability for the user to understand the dataset.

As noted above, participants were more likely to keep documents open to the end of their sensemaking process. Eight of the twelve participants kept all documents open and organized while using IST. However, no significance was found with Kruskal-Wallis correlation tests between score and open documents ($Chisquare = 1.44, p = 0.230$). However, participants noted that it was easier to view all the documents at once in IST. P4 pointed out that If the space is smaller, I had to keep smaller windows. In the [small] task, I kept minimizing windows. In the [large] task I used sticky notes and it helped me fit everything to the side of the screen. For the IST session, I used the whole space around me, so there was no need to use sticky notes because I could keep all my windows open at once. The freedom of nearly unlimited space afforded participants the ability to keep the documents open.
Searching Strategies

One aspect of 3D immersive space is that it affords the ability to lean on spatial memory [10, 18]. We wanted to see if participants could use this in order to find certain documents in our post-session interviews where we asked participants to find certain documents through a brief description. Through this task, we were able to categorize five different types of searching strategies each participant employed: manual, search bar, notes, memory, and spatial memory. We defined manual search as when the participant would, either using the preview feature of Windows or IST, go through files one by one to find the document. Search bar, in contrast, was when the participant used the search feature in any condition. The notes category was defined as when participants referred to their notes to find the document. The memory condition was when the participants simply recalled the document without referring to the interface at all. Lastly, spatial memory was defined where the participant remembered where they had organized the document and referenced it. For the small condition, there were three cases of manual search, five cases of using the search bar, two cases of using notes, and two cases of memory. For large, we found two participants using the search bar, six using notes, three using spatial memory, and one using memory. In IST, we found six using the search feature, two using notes, four using spatial memory. A bar chart of counts can be seen in 5.14.

We measured the time to find the documents each time, where the time starts at the end of the question and ends when the participant correctly identifies the document. The timing results for each category of search can be seen in figure 5.14. First, we ran an anova to see if there was an effect of condition on search time and found weak significance \( (F(2, 22) = 3.292, P = 0.0561) \). Tukey’s post-hoc analysis with bonferroni corrections found significant differences between small and large \( (p = 0.0378) \). We believe this is mostly due to the manual search condition, which took the longest amount of time, only appeared in the small condition. The lack of space seemingly affected the organization of participants’ sensemaking process that they resorted to strategies that had lesser performance. To investigate differences in the strategies, we ran an ANOVA to look for significant differences between the five categories of search and their effect on the time it took to find the document and found significance \( (F(4, 7) = 558, p = 0.0242) \). Tukey’s post-hoc analysis with bonferroni corrections revealed significant differences between manual and each other condition \( (p < 0.001 \text{ in each case}) \), and between search bar and spatial memory \( (p = 0.00805) \). This backs up our supposition that small performed worse than IST or large in this task due to the manual search strategy. Furthermore, spatial memory appeared in only the large and IST conditions, with the most appearing in IST. This does back up our hypothesis that IST supports spatial memory.
5.2. COMPARISON OF SMALL, LARGE, AND IMMERSIVE DISPLAYS

Key Takeaways

For each of our research questions we can make some claims based on the data we gathered. For RQ3.1, we found that users can become frustrated with inadequate space for the sensemaking task, as evidenced by the UEQ and NASA-TLX feedback. Participants found the small condition significantly less attractive, efficient, and stimulating than either the large or IST conditions. Combining the frustration and UEQ results, we believe this affects the amount of time users want to spend using the interface, as the participants spent less time overall in the small condition. We further found that as space increases, so too does the amount of interactions users have with each document, seen particularly with how participants moved documents around the space.

For RQ3.2, we found that strategies varied in each condition as participants had more space to work with. In particular, we saw participants develop the strategy of using the preview feature in Windows instead of opening each document so that they could conserve space. This meant that they were willingly sacrificing the ability to use other features, such as highlight or organization techniques in order to maintain their given space. In contrast, we found participants making more notes in large than the other conditions. We believe this is an artifact of them having a lot more space, but not quite enough space to keep all the bulkier WordPad windows open. It saved space to use Sticky Notes instead. Lastly, we saw immersive space provide users with the ability to keep all the documents open at the same time, which they tended to do in that condition. In addition, we found several participants assigning meaning to spaces while using IST. They kept these “working areas” separate from one another so they could have particular activities siloed to certain spaces.

For RQ3.3, we found that 3D interaction methods affected how participants interacted with documents. For example, participants liked bringing documents closer to themselves when interacting with them directly. This allowed them to focus on one document at a time. Furthermore, participants would discuss how the 3D interactions assisted with their spatial memory and aided their recall of concepts or themes.

Lastly, for RQ3.4 we addressed how well participants performed in each condition. While there were no significant findings here, participants did perform the best on the hardest and longest dataset while using IST. Further investigation is required to see if this trend persists and is significant with a higher N.

5.2.4 Limitations

While we did successfully find some differences between conditions in our experiment, there are some limitations in our study design. First, the datasets we used might not have been large enough - participants didn’t feel like they lacked for space in the large display condition. Further studies should use even larger datasets in order to stress the importance of space even further.
Relatedly, our findings suggest that how participants use space impacts how well they perform the sensemaking process. Future studies should try to measure an amount of spatial acuity in order to better understand if a greater understanding of space increases a user's ability to leverage IST’s benefits.

We made small adjustments to each condition in order to balance both ecological validity and mitigate confounding variables; the small and large monitor conditions were simulated using virtual monitors and IST was redesigned to be a seated experience. This meant that for the virtual monitor conditions the participants had the weight of the HMD and tether impacting them, and the virtual monitors themselves are limited in that they were not physical objects that real-world users could move, rotate, or perceive naturally. For IST, this meant that the participants didn’t use physical movement in order to sort the documents, which is a different way of perceiving space as opposed to the rotational movement we had in the experiment. However, our research questions were addressing the impact of immersive space and immersive user interfaces, and the conditions are generalized towards our research questions.

Our population may also impact our results. Despite each participant requiring 5-6 hours of work, an N of 12 is still small and affected our ability to measure significance. Furthermore, novice users may not be the best population to measure the effect of our conditions on sensemaking tasks. Professional analysts could be used to better understand their needs and how they use immersive space.

5.2.5 Conclusions

IST is a promising approach for sensemaking of large non-quantitative datasets, but could still use some minor improvements. The Windows operating and WIMP interface paradigm has been refined for decades, but immersive analytics and its interaction methods have only been recently explored. With time and refinement, we can improve the interfaces to overcome any small gaps in user experience and performance. For example, a seated immersive analytics experience can move away from 6DOF controllers and use bare hand interactions or mouse adaptations to increase user satisfaction and make the context switch of sorting and typing more streamlined.

We found that increasing the amount of space a user has does increase their satisfaction and lower frustration with the environment, as well as creating more varied strategies. The “working area” concept has merit in separating tasks to certain spatial areas, allowing users to leverage their spatial memory in order to better process the data at hand. Users also leveraged more spatial memory to find documents as the available space increased. Furthermore, users that utilized as much space as they were given had a weak correlation with the scores they achieved on the sensemaking task. Overall, we recommend providing users with as much space as available, which would indicate that we should explore 3D immersive space for more analytical tasks.
5.3 Chapter Summary

We performed two separate studies in order to understand the relative benefits of IST as compared to single screen conditions common in desktop scenarios. In a single-participant pilot study, we found that they created higher quality essays in IST as compared to the single monitor condition, with expert graders stating that the essays created in IST showed greater understanding of the source material. In the comprehensive study between a small virtual monitor, a large virtual monitor, and IST, we found that participants showed more engagement with the documents in IST, where participants moved and kept documents open longer in IST than in the other conditions. We also found that users employed different strategies based on the amount of space they were given; participants had to manage the spatial limitations through using the preview feature in small to save space, or using sticky notes in the large condition instead of the documents themselves, to creating timelines and separate working areas in the IST condition. Participants also had mixed reactions to the 3D interactions inherent in immersive space. Some disliked using the controller as it hindered the ability to switch seamlessly between document interaction and typing. Others enjoyed the ability to easily change the depth of the documents to have increased focus on particular documents of interest. Furthermore, we found evidence that participants used spatial memory in the IST condition in a document search task, which performed better than two of the other four search strategies that were employed for the task. Participants also felt that IST was more suited to intelligence analysis tasks than the other conditions. These studies indicate there is potential in the IST approach as compared to 2D monitors.
Chapter 6

Conclusions & Future Work

6.1 Conclusions

This dissertation was split up into three overarching research questions to better understand the how immersive space affects the sensemaking process. This chapter reviews each of those questions and the results of the studies performed to answer them.

6.1.1 RQ1: How can we design analytic tools for 3D immersive space for sensemaking with large non-quantitative datasets?

Over the course of each experiment we gathered more insight from users and expert analysts on how best to refine the design of IST. Essential interactions include the ability to manipulate the documents in the immersive environment, annotate the dataset, and search the dataset. Annotations should include highlighting documents, adding notes to documents or clusters of documents, and adding labels to clusters of documents. Lastly, IST should support synthesis of findings so users can share their insights with others. We designed an augmented virtuality “Desk Portal” to support text input and adding the ability for users to synthesize the notes, labels, and reports. We then performed a study to compare AR and VR and found through user preference, performance, and user experience data that, much like the Desk Portal, we should support an augmented virtuality approach for IST. This affords the ability for the users to focus on the dataset while still allowing for usage of real-world tools that aren’t inherently supported digitally.

6.1.2 RQ2: How does 3D immersive space affect users performing sensemaking tasks with non-quantitative datasets?

Once we initially designed IST, we needed to understand how users leverage 3D immersive space and interaction methods to process large, complex, non-quantitative datasets. We performed an exploratory study where participants performed a historical analysis task. We found participants leveraging the full virtual environment and creating three primary 2.5-dimensional structures with the data artifacts: semicircular, planar, and environmental
layouts. Semicircular layouts laid the artifacts around the user in a semicircle, allowing them to view all of them simultaneously, while planar layouts had several separate viewing points for multiple different clusters of documents. Lastly, environmental layouts used cues embedded in the virtual environment, such as the edges of the floor, as scaffolding for their layouts. Users also employed various different methods of offloading cognition onto the environment through notes, labels, and highlights, and there was no significant difference in performance between each strategy. This would indicate that we need to support each of them, as it is a user preference how they wish to annotate the data.

6.1.3 RQ3: What are the relative benefits for 3D Immersive space as compared to large scale, high-resolution displays and traditional methods of performing sensemaking tasks?

We addressed RQ3 with two separate studies comparing various amounts of space to 3D immersive space and interactions. In a single-participant pilot study, that participant generated essays in both a traditional desktop setup and IST. Expert evaluators stated that the essays generated in IST showed a closer reading of the dataset and a deeper understanding of the issues in the historical analysis tasks. A more comprehensive study was also performed comparing a seated version of IST in AR to a small AR virtual monitor representing a traditional desktop setup and a large AR virtual monitor representing the Space to Think design. Participants showed frustration with the small condition and the lack of space, through shorter task performance times and reduced user experience scores. We also found greater engagement with the document set in IST, as well as increased use of spatial memory as compared to the other conditions. In a document search task, users found documents more quickly when they used a spatial memory strategy over the search tool or manually searching through documents. Strategies also varied based on the amount of space given to the participants. For example, participants tended to use the preview feature in small, sacrificing the ability to annotate documents but saving space for notes and report synthesis. In the large condition, participants used more notes as they didn’t have quite enough space to keep all of the original documents open. In IST, however, participants tended to keep all the documents open and organized around them, and some leveraged the concept of “working areas,” where they separated tasks into physical areas around them. However, participants had mixed reactions to 3D interaction. Many of them leveraged the concept of depth to denote importance, with closer documents receiving more attention than far away documents. In contrast, participants generally disliked the 6DOF controller because it inhibited switching contexts between interacting with documents and typing. Overall, participants felt that IST better supported sensemaking tasks over the other two conditions.
6.2 Future Work

6.2.1 Scalability of Immersive Space to Think

One aspect for the design of IST and research question one that needs further research is how to make accommodations for very large datasets of thousands of discrete data artifacts. The tracked area we had access to is fairly large, but is still larger than many work areas would afford. Users would run out of physical space with these very large datasets. One way this might be mitigated is virtual navigation “moving” around the area. However, this may impact situated cognition and spatial memory as the space is no longer static.

There are many different types of virtual navigation that could be employed to look at how they impact the sensemaking process. For example, we could create a “rooms in a mansion” metaphor where users traveled between them using the revolving bookshelf metaphor as discussed by Yu et al. [123]. Alternatively, we could use more traditional methods such as teleportation or flying metahpors [29]. Additionally, this calls into question whether the user should have the ability to rotate the environment, or how landmarks affect the spatial understanding. These are important factors in how we should design IST to better assist the user in handling very large datasets.

6.2.2 AI-powered Assistance for Sensemaking in IST

Another design consideration for research question one is how we can add more assistive tools to the IST design to enhance the sensemaking process. As stated, this process is cognitively difficult, iterative, and repetitive, and methods to reduce that load could lead to better outcomes. Machine learning techniques can assist the user through automatic clustering in immersive space, as seen in Tahmid et al.’s work [113], which would allow users to move document groupings more easily. These clusters could also afford processing that could automatically create labels of shared themes. Additionally, many 2D tools such as ForceSPIRE have added AI or natural language processing to assist users in understanding datasets and could easily be adapted and enhanced for 3D immersive space [33]. Even further, these techniques can help on an individual document level. For example, natural language processing could summarize longer documents to help distill them into smaller chunks.

With the help of professional analysts, these possibilities could be combined into a single integrated AI-assisted Immersive Space to Think. The resulting approach could then be compared to previous versions, as well as analyzed for user experience and qualitative feedback for how they assist users.
6.2.3 Effect of Real-World Environment

While research question one aimed to understand how to design IST, research question two addresses how users actually make use of immersive space. One way these two questions intersect is in AR, and how the real world affects users strategies. In Chapter 3, we compared AR and VR implementations of IST, but we didn’t vary was the open-office style environment for the AR implementation. Our office was cluttered - the desks were actively being used by graduate students, which raises a question on if this impacted the performance or strategies of participants in that condition.

Several studies have already been performed to understand the strategies employed in AR for the sensemaking process. We analyzed several different strategies in Chapter 3, and Luo et al. looked at the impact of furniture and collaboration on user strategies in AR [79]. However, these studies didn’t vary the environment beyond amount of furniture. Further research should look into clean environments without many small objects as well as cluttered rooms with objects adorning the furniture and environment.

6.2.4 Best Practices for the use of Immersive Space to Think

Research question two focused on how immersive space affected the sensemaking process and what strategies were employed to assist with understanding, and two of the studies suggest that certain strategies work better in IST than others. More investigation should be performed to understand the particular aspects that create those advantages. For example, in the last study that looked at how immersive space and 3D user interaction techniques affected the sensemaking process seen in section 5.2 the concept of “working areas” may indicate that users who have a higher spatial understanding perform better than others. In addition, the Sensemaking Strategies study in section 4 found that participants with semicircular layouts tended to have a closer reading of the source material.

Future studies should gather more information from participants to better understand these performance differences. Various learning styles could potentially impact how well users utilize space, so more background information should be gathered. The Learning Style Inventory [65] or the Learning Style Questionnaire [5] can assess how a participant processes new data and learn from it, while the Paper Folding Test [31] and the Cube Rotation Test [106] can assess a participant’s spatial understanding. Concurrently, now that we have a set of identifiable strategies we can direct our post-session interviews to ask participants why they leverage the various strategies we observe during the study. Using this data, we can create a set of guidelines and best practices for sensemaking in immersive space.
Bibliography


Appendices
Appendix A

Sensemaking Strategies with Immersive Space to Think Appendix

A.1 Informed Consent Form

VIRGINIA POLYTECHNIC INSTITUTE AND STATE UNIVERSITY

Consent to Take Part in a Research Study

Title of research study: Immersive Space to Think a Multi-Session Study (IRB #19-561)

Principal Investigator:
Douglas A. Bowman
Department of Computer Science, Virginia Tech
Telephone: (540)953-5019
Email: dbowman@vt.edu

Other study contact(s):

Lee Lisle
Department of Computer Science
Virginia Tech
Email: llisle@vt.edu

Payel Bandyopadhyay
Department of Computer Science
Virginia Tech
Email: pbandyop@vt.edu

Chris North
Department of Computer Science
Virginia Tech
Email: north@cs.vt.edu

Ed Gitre
Department of History
Virginia Tech
Email: egitre@vt.edu

**Key Information:** The following is a short summary of this study to help you decide whether or not to be a part of this study. More detailed information is listed later on in this form.

**Why am I being invited to take part in a research study?**

We invite you to take part in a research study because you are a Virginia Tech student that is at least 18 years old who has normal or corrected (glasses or contacts) vision with strong English skills and an interest in historical analysis.

**What should I know about being in a research study?**

- Someone will explain this research study to you
- Whether or not you take part is up to you
- You can choose not to take part
- You can agree to take part and later change your mind
- Your decision will not be held against you
- You can ask all the questions you want before you decide

**What should I know about this research study?**

This research project is intended to compare a variety of methods for interacting with virtual reality (VR) and augmented reality (AR) systems. This research will help us understand the best ways to interact with data and information in VR and AR, and produce design guidelines for VR/AR user interface design. The results of this study may appear in future publications and presentations building upon this research, though all results will be presented in summary form.

This study involved up to two 60 minute sessions split into the following phases

- Pre-study Phase (5 minutes): During this section, we will introduce you to the area, the general concept of our research, and the tools we will be using.
- Tutorial Phase (10 minutes): This phase will present the software we will be using as well as how the software will be used during the study.
- Study Phase (45-90 minutes, split into separate sessions if needed): In this phase we will ask you to perform a task using the software shown in the Tutorial Phase.
- Post-study Phase (15 minutes): During this phase we will perform a semi-structured interview on your overall experience in the study phase.
With many VR/AR tasks, there is a chance that you will have “simulator sickness,” which is a kind of motion sickness. Otherwise, there are minimal risks during this study. While there are no personal benefits from this study, the research we perform may improve future immersive analytics software. If at any time during the study you wish to stop your participation, please let us know and we will terminate the study immediately with no repercussions.

**Detailed Information:** The following is more detailed information about this study in addition to the information listed above.

**Who can I talk to?**

If you have questions, concerns, or complaints, or think the research has hurt you, talk to the research team at dbowman@vt.edu

This research has been reviewed and approved by the Virginia Tech Institutional Review Board (IRB). You may communicate with them at 540-231-3732 or irb@vt.edu if:

- You have questions about your rights as a research subject
- Your questions, concerns, or complaints are not being answered by the research team
- You cannot reach the research team
- You want to talk to someone besides the research team to provide feedback about this research

**How many people will be studied?**

We plan to include 20-50 people in this research study.

**What happens if I say yes, I want to be in this research?**

The study will take place in the Sandbox (Room 160) of the Moss Arts Center, and will consist of 1 or 2 60-minute sessions for each participant split into 4 phases.

In the first “Pre-study” phase, you will be shown the Sandbox, the equipment we will be using the experiment (Either VIVE Pro with wireless attachment or MicroSoft Hololens), and given a brief overview of our research. After that, we will go into the “Tutorial” phase, where we introduce our software, called “The Immersive Space to Think (IST).” We will also give you a set of news articles from CNN to manipulate using IST, as well as guide you through all of the capabilities of IST. This tutorial is meant to prepare you for the “Study” phase. During the Study Phase, you will be interacting with interview data from discharge interviews with World War 2 soldiers. Using these primary sources, you will be asked to write an outline of a paper that addresses two prompts focusing on race relations. You are free to use the software in any way you would like to achieve this result, using the tools we describe in the tutorial phase. During this phase we ask that you verbalize all your thoughts while using the software so that we can better understand your motivations and actions.
Once you have completed your outline, we will ask you to walk us through it while still using IST. If needed, we will stop during this phase and schedule a second session for you to come in and complete the study and outline. After this phase, we will conduct a semi-structured interview during the “Post-study” phase. This interview will ask you about your experiences using the software, and its conclusion marks the end of the study.

**What happens if I say yes, but I change my mind later?**

You can leave the research at any time, for any reason, and it will not be held against you. If you decide to leave the research, please tell the researcher that you would like to terminate the study, and they will accommodate you immediately. All of the data collected from you before this point will be destroyed.

**Is there any way being in this study could be bad for me? (Detailed Risks)**

Using VR and AR technology can produce symptoms of sickness or discomfort in some users. These symptoms are usually mild, and may include dizziness, nausea, eye strain, headache, or disorientation. During tasks involving physical movement, there is also some risk that you could collide with obstacles in the physical environment.

You will be given the option to take a break or quit the experiment at any time. To mitigate the risk of sickness and discomfort, we will adjust the display properly, keep task sessions short, provide frequent breaks, and ask you after each set of tasks how you are feeling. To mitigate the risk of physical obstacles, we will keep the tracked area clear, show you where the boundaries of the space are, display a virtual wall when you near a physical boundary, and warn you if you are nearing an obstacle.

Also, this study will include survey responses from WW2 soldiers pertaining to racial relations, which contains racial slurs and other objectionable material and may cause some psychological distress. If this causes you distress and you wish to stop the experiment, please let us know and we can stop at any time.

**Will being in this study help me in any way?**

Participating in this study will provide you with $70.00 in compensation. Payments will be made in the form of cash after each session. The breakdown of the compensation is as follows:

- Session 1 - $20.00
- Session 2 - $20.00
- Session 3 - $30.00

Besides compensation, there are no benefits to you from your taking part in this research. We cannot promise any benefits to others from your taking part in this research. However, possible benefits to others include advances to immersive analytics for researchers in the future.

**What happens to the information collected for the research?**

We will make every effort to limit the use and disclosure of your personal information,
including research study records, only to people who have a need to review this information. We cannot promise complete confidentiality. Organizations that may inspect and copy your information include the IRB, Human Research Protection Program, and other authorized representatives of Virginia Tech.

The results of this research study may be presented in summary form at conferences, in presentations, reports to the sponsor, academic papers, and as part of a thesis/dissertation. If identifiers are removed from your private information that are collected during this research, that information could then be used for future research studies or distributed to another investigator for future research studies without your additional informed consent.

**What else do I need to know?**

This research is being funded by the Office of Naval Research via the University of California, Santa Barbara.

Participating in this study will provide you with no compensation, monetarily or otherwise. All subjects in the study do so of their own free will and a desire to further computing technologies.

I give the 3DI Interaction Group at Virginia Tech permission to record my research participation through audio, screen, and video recordings: _____ (Initial)

Signature of Subject & Date

____________________
Printed Name of Subject

____________________
Signature of Person Obtaining Consent & Date

____________________
Printed Name of Person Obtaining Consent

---

**A.2 Questionnaires & Interviews**

**A.2.1 Background Questionnaire**

1. Gender:
   - Male o
   - Female o
2. Age

_____ 

3. Major of Study

_____ 

4. How are you feeling today?

1 2 3 4 5 6 7
Negative o o o o o o o Positive

5. Do you wear glasses or contact lenses?

o Neither
o Glasses
o Contact Lenses

6. Are you:

o Right-Handed
o Left-Handed
o Ambidextrous

7. Rate your fatigue level:

1 2 3 4 5 6 7
Very Tired o o o o o o o Not Tired At All

8. Rate your expertise with Computers:

1 2 3 4 5 6 7
Novice o o o o o o o Expert

9. How often do you use computer for work?

o Not at all
o Once a month
o Once a week
o Several times per week
o Daily

10. How often do you use computer for fun?

o Not at all
o Once a month
o Once a week
o Several times per week
o Daily
11. How often do you play video games (consoles or on the computer):
   - Not at all
   - Once a month
   - Once a week
   - Several times per week
   - Daily

12. Comments?

13. How many times have you tried virtual or augmented reality?
   - Never used
   - Once or twice
   - Three to ten times
   - More than ten times

14. Comments?

A.2.2 Post-session Semi-structured Interview

1. How would you answer question 1, and what evidence would you use?
2. How would you answer question 2, and what evidence would you use?
3. What clusters of documents did you form, and how were they labeled?
4. What spatial relationships did you particularly use in order to extract meaning?
5. Does the spatial layout reflect the organization of your answer?

A.3 Post-Study Semi-structured Interview

1. What was your overall impression of the system and its ability to support historical document analysis?
2. If you were writing a real historical manuscript or paper would you want to use a system like IST?
3. What were the most useful features of the system and why?
4. Did you feel that there were features that were missing and that would be helpful in this task?
5. Was there anything confusing, annoying, or difficult about using the system?
6. Do you have any other comments about using the system, or any questions about the system?

A.4 IRB Approval Letter
MEMORANDUM

DATE: September 26, 2019

TO: Douglas Andrew Bowman, Christopher L North, Payel Bandyopadhyay, Lee Lisle, Ed Gilre

FROM: Virginia Tech Institutional Review Board (FWA00000572, expires January 29, 2021)

PROTOCOL TITLE: Immersive Space to Think

IRB NUMBER: 19-561

Effective September 26, 2019, the Virginia Tech Institution Review Board (IRB) approved the New Application request for the above-mentioned research protocol. This approval provides permission to begin the human subject activities outlined in the IRB-approved protocol and supporting documents.

Plans to deviate from the approved protocol and/or supporting documents must be submitted to the IRB as an amendment request and approved by the IRB prior to the implementation of any changes, regardless of how minor, except where necessary to eliminate apparent immediate hazards to the subjects. Report within 5 business days to the IRB any injuries or other unanticipated or adverse events involving risks or harms to human research subjects or others.

All investigators (listed above) are required to comply with the researcher requirements outlined at: https://secure.research.vt.edu/external/irb/responsibilities.htm

(Please review responsibilities before beginning your research.)

PROTOCOL INFORMATION:

Approved As: Expedited, under 45 CFR 46.110 category(ies) 6,7
Protocol Approval Date: September 26, 2019
Progress Review Date: September 25, 2020

ASSOCIATED FUNDING:

The table on the following page indicates whether grant proposals are related to this protocol, and which of the listed proposals, if any, have been compared to this protocol, if required.
<table>
<thead>
<tr>
<th>Date*</th>
<th>OSP Number</th>
<th>Sponsor</th>
<th>Grant Comparison Conducted?</th>
</tr>
</thead>
<tbody>
<tr>
<td>07/03/2019</td>
<td>PET47E VN</td>
<td>University of California, Santa Barbara (Title: View management and user interface optimization for wide-area mobile augmented real)</td>
<td>Compared on 09/20/2019</td>
</tr>
</tbody>
</table>

* Date this proposal number was compared, assessed as not requiring comparison, or comparison information was revised.

If this protocol is to cover any other grant proposals, please contact the HRPP office (irb@vt.edu) immediately.
Appendix B

Different Realities: A Comparison of Augmented and Virtual Reality for Sensemaking Tasks Appendix

B.1 User Consent Form

VIRGINIA POLYTECHNIC INSTITUTE AND STATE UNIVERSITY

Consent to Take Part in a Research Study

Title of research study: Immersive Space to Think a Multi-Session Study (IRB-20-762)

Principal Investigator:

Chris North
Department of Computer Science, Virginia Tech
Telephone: (540)231-2458
Email: north@cs.vt.edu

Other study contact(s):

Lee Lisle
Department of Computer Science
Virginia Tech
Email: llisle@vt.edu

Doug Bowman
Department of Computer Science
Virginia Tech
Email: dbowman@vt.edu

Key Information: The following is a short summary of this study to help you decide whether or not to be a part of this study. More detailed information is listed later on in this form.

Why am I being invited to take part in a research study?

We invite you to take part in a research study because you are a Virginia Tech student that
is at least 18 years old who has normal vision or vision corrected by contact lenses (i.e., no glasses), and be a U.S. Citizen.

What should I know about being in a research study?

- Someone will explain this research study to you
- Whether or not you take part is up to you
- You can choose not to take part
- You can agree to take part and later change your mind
- Your decision will not be held against you
- You can ask all the questions you want before you decide

What should I know about this research study?

This research project is intended to compare a variety of methods for interacting with virtual reality (VR) and augmented reality (AR) systems. This research will help us understand the best ways to interact with data and information in VR and AR, and produce design guidelines for VR/AR user interface design. The results of this study may appear in future publications and presentations building upon this research, though all results will be presented in summary form.

This study involves up to two 90 minute sessions split into the following phases:

- Pre-study Phase (5 minutes): During this section, we will introduce you to the area, the general concept of our research, and the tools we will be using.

- Tutorial Phase (10 minutes): This phase will present the software we will be using as well as how the software will be used during the study.

- Study Phase (120 minutes, split into two separate sessions): In this phase we will ask you to perform a task using the software shown in the Tutorial Phase.

- Post-session Phase (20 minutes, split into two separate sessions): During this phase we will answer a user experience questionnaire and perform a semi-structured interview on your experience in the daily study phase.

- Post-study Phase (15 minutes): During this phase we will perform a semi-structured interview on your overall experience in the study phase.
With many VR/AR tasks, there is a chance that you will have “simulator sickness,” which is a kind of motion sickness. Otherwise, there are minimal risks during this study. While there are no personal benefits from this study, the research we perform may improve future immersive analytics software. If at any time during the study you wish to stop your participation, please let us know and we will terminate the study immediately with no repercussions.

**Detailed Information:** The following is more detailed information about this study in addition to the information listed above.

**Who can I talk to?**

If you have questions, concerns, or complaints, or think the research has hurt you, talk to the research team at north@cs.vt.edu

This research has been reviewed and approved by the Virginia Tech Institutional Review Board (IRB). You may communicate with them at 540-231-3732 or irb@vt.edu if:

- You have questions about your rights as a research subject
- Your questions, concerns, or complaints are not being answered by the research team
- You cannot reach the research team
- You want to talk to someone besides the research team to provide feedback about this research

**How many people will be studied?**

We plan to include 30 people in this research study.

**What happens if I say yes, I want to be in this research?**

The study will take place in the Sandbox (Room 160) of the Moss Arts Center, and will consist of two 90-minute sessions for each participant split into 4 phases.

In the first “Pre-study” phase, you will be shown the Sandbox, the equipment we will be using the experiment, and given a brief overview of our research. After that, we will go into the “Tutorial” phase, where we introduce our software, called “The Immersive Space to Think (IST).” We will give you a set of documents from news articles to manipulate using IST, as well as guide you through all the capabilities of IST. This tutorial is meant to prepare you for the “Study” phase.

During the Study Phase, you will be interacting with responses to Survey 32 (a survey given to WW2 soldiers concerning racial relations). Over the two sessions, you will be asked to write answers to two essay prompts using the provided survey responses while using IST. In the “Post-session” phase, there will be a questionnaire on the user experience with IST and a short interview on your experiences with the software. Lastly, in the “Post-study”
phase you will be asked to compare and contrast your experiences with the software in an interview, and its conclusion marks the end of the study.

**What happens if I say yes, but I change my mind later?**

You can leave the research at any time, for any reason, and it will not be held against you. If you decide to leave the research, please tell the researcher that you would like to terminate the study, and they will accommodate you immediately. All of the data collected from you before this point will be destroyed.

**Is there any way being in this study could be bad for me? (Detailed Risks)**

Using VR and AR technology can produce symptoms of sickness or discomfort in some users. These symptoms are usually mild, and may include dizziness, nausea, eye strain, headache, or disorientation. During tasks involving physical movement, there is also some risk that you could collide with obstacles in the physical environment.

You will be given the option to take a break or quit the experiment at any time. To mitigate the risk of sickness and discomfort, we will adjust the display properly, keep task sessions short, provide frequent breaks, and ask you after each set of tasks how you are feeling. To mitigate the risk of physical obstacles, we will keep the tracked area clear, show you where the boundaries of the space are, display a virtual wall when you near a physical boundary, and warn you if you are nearing an obstacle.

Also, this study will include survey responses from WW2 soldiers pertaining to racial relations, which contains racial slurs and other objectionable material and may cause some psychological distress. If this causes you distress and you wish to stop the experiment, please let us know and we can stop at any time.

**Will being in this study help me in any way?**

Participating in this study will provide you with $50.00 in compensation. Payments will be made in the form of cash after each session. The breakdown of the compensation is as follows:

Session 1 - $20.00, Session 2 - $30.00.

Besides compensation, there are no benefits to you from your taking part in this research. We cannot promise any benefits to others from your taking part in this research. However, possible benefits to others include advances to immersive analytics for researchers in the future.

**What happens to the information collected for the research?**

We will make every effort to limit the use and disclosure of your personal information, including research study records, only to people who have a need to review this information. We cannot promise complete confidentiality. Organizations that may inspect and copy your information include the IRB, Human Research Protection Program, and other authorized
representatives of Virginia Tech.

The results of this research study may be presented in summary form at conferences, in presentations, reports to the sponsor, academic papers, and as part of a thesis/dissertation. If identifiers are removed from your private information that are collected during this research, that information could then be used for future research studies or distributed to another investigator for future research studies without your additional informed consent.

**What else do I need to know?**

This research is being sponsored by the U.S. Department of Defense (DoD). DoD representatives are authorized to review anonymized research records. All subjects in the study do so of their own free will and a desire to further computing technologies.

I give the 3DI Interaction Group and Info Vis Lab at Virginia Tech permission to record my research participation through audio, screen, and video recordings: _____ (Initial)

Signature of Subject & Date

____________________
Printed Name of Subject

____________________
Signature of Person Obtaining Consent & Date

____________________
Printed Name of Person Obtaining Consent

**B.1.1 Questionnaires**

**Background Questionnaire**

1. Gender:
   - Male  o
   - Female  o

2. Age

   _____

3. Major of Study

   _____
4. How are you feeling today?
   1 2 3 4 5 6 7
   Negative o o o o o o o Positive

5. Do you wear glasses or contact lenses?
   o Neither
   o Glasses
   o Contact Lenses

6. Are you:
   o Right-Handed
   o Left-Handed
   o Ambidextrous

7. Rate your fatigue level:
   1 2 3 4 5 6 7
   Very Tired o o o o o o o Not Tired At All

8. Rate your expertise with Computers:
   1 2 3 4 5 6 7
   Novice o o o o o o o Expert

9. How often do you use computer for work?
   o Not at all
   o Once a month
   o Once a week
   o Several times per week
   o Daily

10. How often do you use computer for fun?
   o Not at all
    o Once a month
    o Once a week
    o Several times per week
    o Daily

11. How often do you play video games (consoles or on the computer):
    o Not at all
    o Once a month
    o Once a week
    o Several times per week
    o Daily
12. Comments?

13. How many times have you tried virtual or augmented reality?
   o Never used
   o Once or twice
   o Three to ten times
   o More than ten times

14. Comments?

B.1.2 Post-Session Questionnaires

1. I just finished using:
   o Augmented Reality
   o Virtual Reality

   User Experience Questionnaire

2. The experience was:

   1 2 3 4 5 6 7
   Annoying o o o o o o o Enjoyable

3. The experience was:

   1 2 3 4 5 6 7
   Not Understandable o o o o o o o Understandable

4. The experience was:

   1 2 3 4 5 6 7
   Creative o o o o o o o Dull

5. The experience was:

   1 2 3 4 5 6 7
   Easy to Learn o o o o o o o Difficult to Learn

6. The experience was:

   1 2 3 4 5 6 7
   Valuable o o o o o o o Inferior

7. The experience was:

   1 2 3 4 5 6 7
   Boring o o o o o o o Exciting
8. The experience was:
   1 2 3 4 5 6 7
   Not Interesting o o o o o o o Interesting

9. The experience was:
   1 2 3 4 5 6 7
   Unpredictable o o o o o o o Predictable

10. The experience was:
    1 2 3 4 5 6 7
    Fast o o o o o o o Slow

11. The experience was:
    1 2 3 4 5 6 7
    Inventive o o o o o o o Conventional

12. The experience was:
    1 2 3 4 5 6 7
    Obstructive o o o o o o o Supportive

13. The experience was:
    1 2 3 4 5 6 7
    Good o o o o o o o Bad

14. The experience was:
    1 2 3 4 5 6 7
    Complicated o o o o o o o Easy

15. The experience was:
    1 2 3 4 5 6 7
    Unlikeable o o o o o o o Pleasing

16. The experience was:
    1 2 3 4 5 6 7
    Usual o o o o o o o Leading Edge

17. The experience was:
    1 2 3 4 5 6 7
    Unpleasant o o o o o o o Pleasant

18. The experience was:
    1 2 3 4 5 6 7
CHAPTER B. DIFFERENT REALITIES: A COMPARISON OF AUGMENTED AND VIRTUAL REALITY FOR SENSEMAKING TASKS APPENDIX

Secure o o o o o o Not Secure
19. The experience was:
   1 2 3 4 5 6 7
Motivating o o o o o o Demotivating
20. The experience was:
   1 2 3 4 5 6 7
Meets Expectations o o o o o o Does not meet expectations
21. The experience was:
   1 2 3 4 5 6 7
Inefficient o o o o o o Efficient
22. The experience was:
   1 2 3 4 5 6 7
Clear o o o o o o Confusing
23. The experience was:
   1 2 3 4 5 6 7
Impractical o o o o o o Practical
24. The experience was:
   1 2 3 4 5 6 7
Organized o o o o o o Cluttered
25. The experience was:
   1 2 3 4 5 6 7
Attractive o o o o o o Unattractive
26. The experience was:
   1 2 3 4 5 6 7
Friendly o o o o o o Unfriendly
27. The experience was:
   1 2 3 4 5 6 7
Conservative o o o o o o Innovative
B.1.3 Post-Session Semi-structured Interview

- What was your overall strategy for analyzing the document set? Did this change over the course of the session?
- Please describe the spatial layout you formed this session, and how it changed over the course of the session.
- What was the transition from your IST artifact spatial layout to the essay, if any?

Post-Study Semi-structured Interview

- What was your overall impression of each version of the system and their ability to support analysis tasks?
- If you were writing a real paper or essay, would you want to use a system like this? Why or why not?
- What were the most useful features of the system and why?
- Were there features that you wanted that were missing from the system?
- Was there anything confusing, annoying, or difficult about completing the tasks?
- Which level of mixed reality did you feel better supported your focus?
- Which level of mixed reality did you feel more grounded in reality?
- Did you refer to or use any real-world tools in either level of mixed reality?
- Do you have any other comments?

B.2 IRB Approval Letter
MEMORANDUM

DATE: April 15, 2021

TO: Christopher L North, Douglas Andrew Bowman, Kylie Marie Davidson, Lee Lisle

FROM: Virginia Tech Institutional Review Board (FWA00000572)

PROTOCOL TITLE: Immersive Space to Think a Multi-Session Study

IRB NUMBER: 20-762

Effective April 15, 2021, the Virginia Tech Institution Review Board (IRB) approved the Amendment request for the above-mentioned research protocol.

This approval provides permission to begin the human subject activities outlined in the IRB-approved protocol and supporting documents.

Plans to deviate from the approved protocol and/or supporting documents must be submitted to the IRB as an amendment request and approved by the IRB prior to the implementation of any changes, regardless of how minor, except where necessary to eliminate apparent immediate hazards to the subjects. Report within 5 business days to the IRB any injuries or other unanticipated or adverse events involving risks or harms to human research subjects or others.

All investigators (listed above) are required to comply with the researcher requirements outlined at: https://secure.research.vt.edu/external/irb/responsibilities.htm

(Please review responsibilities before beginning your research.)

PROTOCOL INFORMATION:

Approved As: Expedited, under 45 CFR 46.110 category(ies) 4,6,7
Protocol Approval Date: October 16, 2020
Progress Review Date: October 16, 2021

ASSOCIATED FUNDING:

The table on the following page indicates whether grant proposals are related to this protocol, and which of the listed proposals, if any, have been compared to this protocol, if required.
SPECIAL INSTRUCTIONS:
This amendment, submitted April 8, 2021, updates research protocol to update the experimental task from an intelligence task over 3 sessions to a historical analysis task over 2 sessions and to update Experimenter to Lee Lisle (from Kylie Davidson). Recruitment materials and consent forms were updated to reflect experimenter change from Kylie Davidson to Lee Lisle. Data collection instruments were updated to add the User Experience Questionnaire and to update the post-session and post-study interview questions to reflect changes in experimental design. COVID-19 protocols were updated to reflect a change in equipment and updates to CDC guidance on surface disinfection.

<table>
<thead>
<tr>
<th>Date*</th>
<th>OSP Number</th>
<th>Sponsor</th>
<th>Grant Comparison Conducted?</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

* Date this proposal number was compared, assessed as not requiring comparison, or comparison information was revised.

If this protocol is to cover any other grant proposals, please contact the HRPP office (irb@vt.edu) immediately.
Appendix C

Comparison of Small, Large, and Immersive Displays

C.1 User Consent Form

VIRGINIA POLYTECHNIC INSTITUTE AND STATE UNIVERSITY

Consent to Take Part in a Research Study

Title of research study: Immersive Space to Think a Multi-Session Study (IRB-20-762)

Principal Investigator:

Chris North
Department of Computer Science, Virginia Tech
Telephone: (540)231-2458
Email: north@cs.vt.edu

Other study contact(s):

Lee Lisle
Department of Computer Science
Virginia Tech
Email: llisle@vt.edu

Doug Bowman
Department of Computer Science
Virginia Tech
Email: dbowman@vt.edu

Key Information: The following is a short summary of this study to help you decide whether or not to be a part of this study. More detailed information is listed later on in this form.

Why am I being invited to take part in a research study?

We invite you to take part in a research study because you are a Virginia Tech student that is at least 18 years old who has normal vision or vision corrected vision.
What should I know about being in a research study?

- Someone will explain this research study to you
- Whether or not you take part is up to you
- You can choose not to take part
- You can agree to take part and later change your mind
- Your decision will not be held against you
- You can ask all the questions you want before you decide

What should I know about this research study?

This research project is intended to compare differing levels of space and its impact on the sensemaking process. This research will help us understand the best ways to interact with data and information in augmented reality (AR), and produce design guidelines for AR user interface design. The results of this study may appear in future publications and presentations building upon this research, though all results will be presented in summary form.

This study involves up to three 90 minute sessions split into the following phases:

- Pre-study Phase (5 minutes): During this section, we will introduce you to the area, the general concept of our research, and the tools we will be using.
- Tutorial Phase (10 minutes): This phase will present the software we will be using as well as how the software will be used during the study.
- Study Phase (180 minutes, split into three separate sessions): In this phase we will ask you to perform a task using the software shown in the Tutorial Phase.
- Post-session Phase (30 minutes, split into three separate sessions): During this phase we will answer a user experience questionnaire and perform a semi-structured interview on your experience in the daily study phase.
- Post-study Phase (15 minutes): During this phase we will perform a semi-structured interview on your overall experience in the study phase.

With many AR tasks, there is a chance that you will have “simulator sickness,” which is a kind of motion sickness. Otherwise, there are minimal risks during this study. While there are no personal benefits from this study, the research we perform may improve future immersive
analytics software. If at any time during the study you wish to stop your participation, please let us know and we will terminate the study immediately with no repercussions.

Detailed Information: The following is more detailed information about this study in addition to the information listed above.

Who can I talk to?

If you have questions, concerns, or complaints, or think the research has hurt you, talk to the research team at north@cs.vt.edu

This research has been reviewed and approved by the Virginia Tech Institutional Review Board (IRB). You may communicate with them at 540-231-3732 or irb@vt.edu if:

- You have questions about your rights as a research subject
- Your questions, concerns, or complaints are not being answered by the research team
- You cannot reach the research team
- You want to talk to someone besides the research team to provide feedback about this research

How many people will be studied?

We plan to include 50 people in this research study.

What happens if I say yes, I want to be in this research?

In the first “Pre-study” phase, you will be shown the Sandbox, the equipment we will be using the experiment, and given a brief overview of our research. After that, we will go into the “Tutorial” phase, where we introduce the session’s condition. This can be a small 2D virtual monitor (representing a 24” 4:3 display), a large 2D virtual monitor (representing a 4x2 array of 24” 4:3 monitors) or our software called “The Immersive Space to Think (IST).” We will give you a set of documents from news articles to manipulate using the session’s current condition, as well as guide you through all the user interfaces you will be expected to use. This tutorial is meant to prepare you for that session’s “Study” phase.

During the Study Phase, you will be interacting with three intelligence analyst training datasets (one for each session). Over the three sessions, you will be asked to write your three answers to the intelligence analysis puzzles. In the “Post-session” phase, there will be a questionnaire on the user experience with that session’s software, a questionnaire measuring the workload of the session’s task, and a short interview on your experiences with the software. Lastly, in the “Post-study” phase you will be asked to compare and contrast your experiences in an interview, and its conclusion marks the end of the study.

What happens if I say yes, but I change my mind later?
You can leave the research at any time, for any reason, and it will not be held against you. If you decide to leave the research, please tell the researcher that you would like to terminate the study, and they will accommodate you immediately. All of the data collected from you before this point will be destroyed.

**Is there any way being in this study could be bad for me? (Detailed Risks)**

Using AR technology can produce symptoms of sickness or discomfort in some users. These symptoms are usually mild, and may include dizziness, nausea, eye strain, headache, or disorientation. During tasks involving physical movement, there is also some risk that you could collide with obstacles in the physical environment.

You will be given the option to take a break or quit the experiment at any time. To mitigate the risk of sickness and discomfort, we will adjust the display properly, keep task sessions short, provide frequent breaks, and ask you after each set of tasks how you are feeling. To mitigate the risk of physical obstacles, we will keep the tracked area clear, show you where the boundaries of the space are, display a virtual wall when you near a physical boundary, and warn you if you are nearing an obstacle.

**Will being in this study help me in any way?**

Participating in this study will provide you with $70.00 in compensation. Payments will be made in the form of cash after each session. The breakdown of the compensation is as follows:

Session 1 - $20.00, Session 2 - $20.00, Session 3 - $30.00

Besides compensation, there are no benefits to you from your taking part in this research. We cannot promise any benefits to others from your taking part in this research. However, possible benefits to others include advances to immersive analytics for researchers in the future.

**What happens to the information collected for the research?**

We will make every effort to limit the use and disclosure of your personal information, including research study records, only to people who have a need to review this information. We cannot promise complete confidentiality. Organizations that may inspect and copy your information include the IRB, Human Research Protection Program, and other authorized representatives of Virginia Tech.

The results of this research study may be presented in summary form at conferences, in presentations, reports to the sponsor, academic papers, and as part of a thesis/dissertation. If identifiers are removed from your private information that are collected during this research, that information could then be used for future research studies or distributed to another investigator for future research studies without your additional informed consent.

**What else do I need to know?**
This research is being sponsored by the U.S. Department of Defense (DoD). DoD representatives are authorized to review anonymized research records. All subjects in the study do so of their own free will and a desire to further computing technologies.

I give the 3DI Interaction Group and Info Vis Lab at Virginia Tech permission to record my research participation through audio, screen, and video recordings: ____ (Initial)

Signature of Subject & Date

____________________
Printed Name of Subject

____________________
Signature of Person Obtaining Consent & Date

____________________
Printed Name of Person Obtaining Consent

C.2 Questionnaires & Interviews

C.2.1 Background Questionnaire

1. Gender:
   Male o
   Female o

2. Age
   ____

3. Major of Study
   ____

4. How are you feeling today?
   1 2 3 4 5 6 7
   Negative o o o o o o o Positive

5. Do you wear glasses or contact lenses?
   o Neither
   o Glasses
6. Are you:
   o Right-Handed
   o Left-Handed
   o Ambidextrous

7. Rate your fatigue level:
   1 2 3 4 5 6 7
   Very Tired o o o o o o o Not Tired At All

8. Rate your expertise with Computers:
   1 2 3 4 5 6 7
   Novice o o o o o o o Expert

9. How often do you use computer for work?
   o Not at all
   o Once a month
   o Once a week
   o Several times per week
   o Daily

10. How often do you use computer for fun?
    o Not at all
    o Once a month
    o Once a week
    o Several times per week
    o Daily

11. How often do you play video games (consoles or on the computer):
    o Not at all
    o Once a month
    o Once a week
    o Several times per week
    o Daily

12. Comments?
13. How many times have you tried virtual or augmented reality?
   o Never used
   o Once or twice
   o Three to ten times
   o More than ten times
14. Comments?

15. How much experience do you have with the Windows operating system?
   1 2 3 4 5 6 7
   No Experience o o o o o o Advanced Knowledge

16. Comments?

17. How much experience do you have with the WordPad?
   1 2 3 4 5 6 7
   No Experience o o o o o o Advanced Knowledge

18. Please briefly describe your personal monitor setup. (E.g., number of monitors, monitor size, monitor resolution, etc.)

Post-Session Questionnaires

1. I just finished using:
   o Small Virtual Monitor
   o Large Virtual Monitor
   o Seated IST

User Experience Questionnaire

2. The experience was:
   1 2 3 4 5 6 7
   Annoying o o o o o o Enjoyable

3. The experience was:
   1 2 3 4 5 6 7
   Not Understandable o o o o o o Understandable

4. The experience was:
   1 2 3 4 5 6 7
   Creative o o o o o o Dull

5. The experience was:
   1 2 3 4 5 6 7
   Easy to Learn o o o o o o Difficult to Learn

6. The experience was:
   1 2 3 4 5 6 7
C.2. Questionnaires & Interviews

Valuable o o o o o o o Inferior
7. The experience was:
   1 2 3 4 5 6 7
Boring o o o o o o o Exciting
8. The experience was:
   1 2 3 4 5 6 7
Not Interesting o o o o o o o Interesting
9. The experience was:
   1 2 3 4 5 6 7
Unpredictable o o o o o o o Predictable
10. The experience was:
    1 2 3 4 5 6 7
Fast o o o o o o o Slow
11. The experience was:
    1 2 3 4 5 6 7
Inventive o o o o o o o Conventional
12. The experience was:
    1 2 3 4 5 6 7
Obstructive o o o o o o o Supportive
13. The experience was:
    1 2 3 4 5 6 7
Good o o o o o o o Bad
14. The experience was:
    1 2 3 4 5 6 7
Complicated o o o o o o o Easy
15. The experience was:
    1 2 3 4 5 6 7
Unlikeable o o o o o o o Pleasing
16. The experience was:
    1 2 3 4 5 6 7
Usual o o o o o o o Leading Edge
17. The experience was:
18. The experience was:

1 2 3 4 5 6 7
Unpleasant o o o o o o o Pleasant

19. The experience was:

1 2 3 4 5 6 7
Secure o o o o o o o Not Secure

20. The experience was:

1 2 3 4 5 6 7
Motivating o o o o o o o Demotivating

21. The experience was:

1 2 3 4 5 6 7
Meets Expectations o o o o o o o Does not meet expectations

22. The experience was:

1 2 3 4 5 6 7
Inefficient o o o o o o o Efficient

23. The experience was:

1 2 3 4 5 6 7
Clear o o o o o o o Confusing

24. The experience was:

1 2 3 4 5 6 7
Impractical o o o o o o o Practical

25. The experience was:

1 2 3 4 5 6 7
Organized o o o o o o o Cluttered

26. The experience was:

1 2 3 4 5 6 7
Attractive o o o o o o o Unattractive

27. The experience was:

1 2 3 4 5 6 7
Friendly o o o o o o o Unfriendly

28. The experience was:

1 2 3 4 5 6 7
Conservative o o o o o o o Innovative
C.2. QUESTIONNAIRES & INTERVIEWS

NASA-TLX

Please assess your experience on a scale of 1 to 7 where 1 agrees completely with the descriptor on the left and 7 agrees completely with the descriptor on the right.

28. Mental Demand - How mentally demanding was the task?

1 2 3 4 5 6 7
Very Low o o o o o o o Very High

29. Physical Demand - How physically demanding was the task?

1 2 3 4 5 6 7
Very Low o o o o o o o Very High

30. Temporal Demand - How hurried or rushed was the pace of the task?

1 2 3 4 5 6 7
Very Low o o o o o o o Very High

31. Performance - How successful were you in accomplishing what you were asked to do?

1 2 3 4 5 6 7
Perfect o o o o o o o Failure

32. Effort - How hard did you have to work to accomplish your level of performance?

1 2 3 4 5 6 7
Very Low o o o o o o o Very High

33. Frustration - How insecure, discouraged, irritated, stressed, and annoyed were you?

1 2 3 4 5 6 7
Very Low o o o o o o o Very High

C.2.2 Post-session Semi-structured Interview

1. Please walk me through how you analyzed the dataset in this session.

2. How did the available space impact your analysis?

3. I’m going to ask you to find a particular document in the dataset. You may use any features available in the main phase to find it. Once found, please read the title of the document. Can you please find [description of document in the session’s dataset]?

4. Could you please describe how you located that document?

5. Could you please walk me through how you reported your findings?

6. Do you have any additional comments on the condition you experienced in this session?
CHAPTER C. COMPARISON OF SMALL, LARGE, AND IMMERSIVE DISPLAYS

C.2.3 Post-Study Semi-structured Interview

1. Could you please compare and contrast the three conditions you experienced over the course of this study?

2. Which condition, in your estimation, best supported the task and why? Please rank the other two as well.

3. As the amount of available space changed, how did you organize documents?

4. How, in particular, did 3D space in the IST condition impact your process?

5. Was there anything confusing, annoying, or difficult about the interface in each condition?

6. Do you have any additional comments on the study?

C.3 IRB Approval Letter
MEMORANDUM

DATE: July 5, 2022

TO: Christopher L. North, Douglas Andrew Bowman, Kylie Marie Davidson, Lee Lisle

FROM: Virginia Tech Institutional Review Board (FWA00000572)

PROTOCOL TITLE: Immersive Space to Think a Multi-Session Study

IRB NUMBER: 20-762

Effective July 5, 2022, the Virginia Tech Institutional Review Board (IRB) approved the Amendment request for the above-mentioned research protocol.

This approval provides permission to begin the human subject activities outlined in the IRB-approved protocol and supporting documents.

Plans to deviate from the approved protocol and/or supporting documents must be submitted to the IRB as an amendment request and approved by the IRB prior to the implementation of any changes, regardless of how minor, except where necessary to eliminate apparent immediate hazards to the subjects. Report within 5 business days to the IRB any injuries or other unanticipated or adverse events involving risks or harms to human research subjects or others.

All investigators (listed above) are required to comply with the researcher requirements outlined at: https://secure.research.vt.edu/external/irb/responsibilities.htm

(Please review responsibilities before beginning your research.)

PROTOCOL INFORMATION:

Approved As: Expedited, under 45 CFR 46.110 category(ies) 4,6,7
Protocol Approval Date: October 16, 2020
Progress Review Date: October 16, 2022

ASSOCIATED FUNDING:

The table on the following page indicates whether grant proposals are related to this protocol, and which of the listed proposals, if any, have been compared to this protocol, if required.
**SPECIAL INSTRUCTIONS:**
This amendment approved July 5, 2022 updates the following:

Consent Form(s) — Adding a statement of funding from DoD/ONR

<table>
<thead>
<tr>
<th>Date*</th>
<th>OSP Number</th>
<th>Sponsor</th>
<th>Grant Comparison Conducted?</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

* Date this proposal number was compared, assessed as not requiring comparison, or comparison information was revised.

If this protocol is to cover any other grant proposals, please contact the HRPP office (irb@vt.edu) immediately.